Information availability and data breaches

Data breach notification laws and their effects

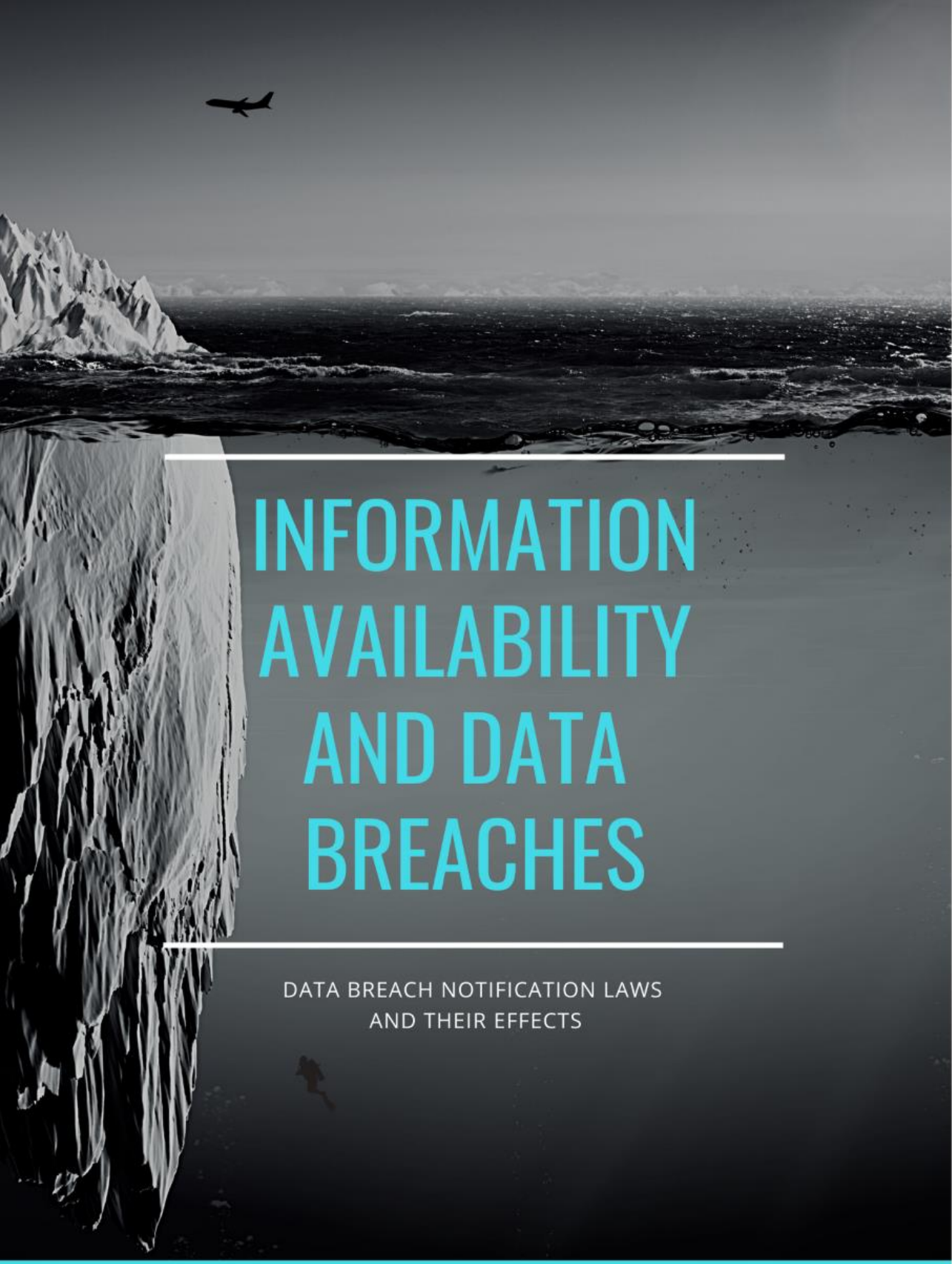
Bisogni, F.

DOI
10.4233/uuid:852a8fb0-5dd6-4759-9510-fa18a0708a29

Publication date
2020

Document Version
Final published version

Citation (APA)
Bisogni, F. (2020). *Information availability and data breaches: Data breach notification laws and their effects*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:852a8fb0-5dd6-4759-9510-fa18a0708a29



INFORMATION AVAILABILITY AND DATA BREACHES

DATA BREACH NOTIFICATION LAWS AND THEIR EFFECTS

FABIO BISOGNI

Information availability and data breaches

Data breach notification laws and their effects

DISSERTATION

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on Tuesday 25 August 2020 at 15:00 o'clock
by

Fabio BISOGNI

Master di secondo livello in governo dei sistemi informativi, Università degli studi Roma Tre, Italy
born in Rome, Italy

This dissertation has been approved by the promotor.

Composition of the doctoral committee:

Rector Magnificus,	Chairman
Prof.dr. M. J.G. van Eeten	Delft University of Technology, promotor

Independent members:

Prof.mr. A.R. Lodder	VU. Amsterdam
Prof.dr. G. Jacobs	Erasmus U. Rotterdam
Prof.dr. M.G. Faure	Erasmus U. Rotterdam
Prof.dr.ir. G.L.L.M.E. Reniers	Delft University of Technology
Prof.mr.dr. J.A. de Bruijn	Delft University of Technology

ISBN 978-94-6402-415-9

Printed in the Netherlands by Gildeprint.

Distributed by Delft University of Technology, Faculty of Technology, Policy and Management, Jaffalaan 5, 2628BX Delft, the Netherlands.

Keywords*:* cybersecurity, economics, governance, privacy disclosure policy, data breaches, identity thefts, data breach notification laws, security breaches.

Experiment escorts us last -

His pungent company

Will not allow an Axiom

An Opportunity

Emily Dickinson

Acknowledgements

I believe that research is not only about science but also about management: managing people (yourself included), time, and, more generally, resources. While the first capacity increases typically year after year, the time that you can dedicate to what you love tends to shrink. This notion gives even higher value to the incredible opportunity I had in writing this dissertation.

Apart from my stubbornness (or the will to challenge myself constantly) inherited from my father, I am especially thankful to the people that have made this journey possible. First of all, my family never ceased to be a constant source of motivation and support. Secondly, Simona helped me, more than ten years ago, to discover the true meaning of research through spending days (and nights) together working on European research projects. Thirdly, Hadi, who I met on my first day at TU Delft, became a very good friend and a great 'complementary' co-author. Last but not least, my promoter Michel believed in the project of an Italian stranger, but more importantly, became a great source of inspiration. Michel allowed me to become a peculiar member of his team: a mix of motivated, talented, extravagant, serious, and funny scientists who always made me feel welcome when in Delft. I owe them many thanks for that.

It has been a long journey; during this time, I built a family (Massimiliano and Nevia made my writing nights full of intervals), I met many interesting people, and I confronted different cultures, approaches, and working dynamics. It has been a special time, special first to have the curiosity, and then the opportunity to research new paths and reasons why things happen: in science, and especially in society.

Fabio Bisogni, July 2020

Table of Contents

List of Tables

Summary

In response to evolving cybersecurity challenges, global spending on information security has grown steadily, and could eventually reach a level that is inefficient and unaffordable. A better understanding of new socio-technical-economic complexities around information security is urgently needed, which requires both reconsideration of traditional cybersecurity issues and investigation of new and unexplored research directions.

In recent times, interdisciplinary research has elucidated the many economic and behavioural dimensions of security. This research is rooted in the field of *Information Security Economics*, and primarily addresses disclosure policy and specifically, data breach notification laws.

Data breach notification laws require any business that suffers a data breach, or believes that it suffered a data breach, to notify customers about the incident that entails the unauthorised acquisition of unencrypted and computerised personal information. Such laws offer incentives to the party who owes the notification duty to minimise the number of triggering events and also enable the affected third parties to diminish the consequences, namely identity theft, and to make prudent choices in the future.

Public policy that seeks to improve the effects of data breach notification legislation must be informed by a comprehensive understanding of the behaviour and incentives of the organisations and individuals involved in the notification flow. Thus, this dissertation poses the following research question:

> *What are the effects of the provisions of data breach notification laws on (1) communications issued by breached organisations to their customers; (2) the timing of breach detection and reaction; (3) the number of data breaches reported; and (4) the volume of identity theft stemming from data breaches?*

As we live in the era of big data, it was possible to access and utilise data on the number of breaches and the number of notifications sent. However, it was also necessary to examine further the types of breaches that occurred as well as the types of communication sent and how individuals perceived them. This analysis allows to develop specific metrics, activating critical thinking about the measurement and the underlying phenomenon.

This dissertation examines these notions and answers the research question through one theoretical peer-reviewed paper and four peer-reviewed empirical studies, each addressing a separate aspect related to the implementation of notification mechanisms, specifically data breach notification laws. Chapter one studies the role of information availability in the cybersecurity landscape and describes a theoretical model for evaluating data breach notification laws as a solution to tackle information asymmetries in the digital arena. Chapter two focuses on the tangible tools needed to implement such laws, specifically the notification process itself, and analyses the extent to which each organisation has leeway to ensure compliance with the law. Drawing on the variation in time for data breach detection and notification and letter content analysis, chapter four discusses the necessity to implement superseding law in order to bring coherence to the diverse approaches used in different geographical areas. Chapter five then addresses underreporting of data breaches. Finally, chapter six explores the relationship between data breaches and identity theft.

The dissertation concludes by reflecting on the shared elements across the studies. The conclusion reflects on the role of disclosure policies in the information security arena and on the implications, given the results of these studies, for European data breach notification policies.

Samenvatting (Dutch Summary)

Als reactie op de zich ontwikkelende uitdagingen op het gebied van cybersecurity zijn de wereldwijde uitgaven voor informatiebeveiliging gestaag toegenomen en zouden ze uiteindelijk een niveau kunnen bereiken dat inefficiënt en onbetaalbaar is. Er is dringend behoefte aan een beter begrip van de nieuwe sociaal-technische economische complexiteiten rond informatiebeveiliging, wat zowel een heroverweging van de traditionele cyberbeveiligingsproblematiek als een onderzoek naar nieuwe en onontgonnen onderzoeksrichtingen vereist.

Recent heeft interdisciplinair onderzoek de vele economische en gedragsdimensies van beveiliging aan het licht gebracht. Dit onderzoek is geworteld in het domein van *informatiebeveiligingseconomie*, en richt zich voornamelijk op de wettelijke meldingsplichten rond datalekken.

De wetten voor het melden van datalekken vereisen dat elk bedrijf dat een datalek ondergaat, of denkt dat het een datalek heeft gehad, klanten op de hoogte stelt van het incident. Dergelijke wetten bieden een stimulans aan de partij die de meldingsplicht heeft om het aantal triggering events te minimaliseren. Daarnaast zijn deze wetten bedoeld om betrokken derde partijen in staat te stellen de gevolgen, namelijk identiteitsdiefstal, te verminderen en in de toekomst veiligere keuzes te maken.

Overheidsbeleid dat gericht is op het verbeteren van de meldingsplicht van datalekken moet worden gevoed door middel van een empirisch inzicht in het gedrag van de organisaties en personen die betrokken zijn bij de meldingenstroom. Dit proefschrift stelt dan ook de volgende onderzoeksvraag:

> *Wat zijn de effecten van de wettelijke bepalingen inzake de melding van datalekken op (1) de communicatie van overtredende organisaties aan hun klanten; (2) de timing van de opsporing van en*

reactie op datalekken; (3) het aantal gemelde datalekken; en (4) de omvang van de identiteitsdiefstal als gevolg van datalekken?

Via online bronnen is het mogelijk om toegang te krijgen tot gegevens over het aantal inbreuken en het aantal verstuurde meldingen en deze te gebruiken. Het was echter ook nodig om verder onderzoek te doen naar de soorten inbreuken die zich voordeden, alsook naar de soorten communicatie die werden verstuurd en de manier waarop individuen deze waarnamen.

Dit proefschrift beantwoordt de onderzoeksvraag door middel van één peer-reviewed theoretisch paper en vier peer-reviewed empirische studies, die elk een apart aspect behandelen dat verband houdt met de implementatie van meldingsmechanismen. Hoofdstuk één bestudeert de rol van de beschikbaarheid van informatie in het cyberbeveiligingslandschap en beschrijft een theoretisch model voor de evaluatie van wetten voor de melding van datalekken als oplossing voor de aanpak van informatieasymmetrieën in de digitale arena. Hoofdstuk twee richt zich op de concrete instrumenten die nodig zijn om dergelijke wetten te implementeren, in het bijzonder het meldingsproces zelf, en analyseert de mate waarin elke organisatie ruimte heeft om de naleving van de wet te waarborgen. Aan de hand van de verschillen in de tijd voor het opsporen en melden van datalekken en de analyse van de inhoud van brieven, gaat hoofdstuk vier in op de noodzaak van de invoering van vervangende wetgeving om samenhang te brengen in de verschillende benaderingen die in de verschillende geografische gebieden worden gebruikt. In hoofdstuk vijf wordt vervolgens ingegaan op de onderrapportering van datalekken. Tot slot gaat hoofdstuk zes in op de relatie tussen datalekken en identiteitsdiefstal.

Het proefschrift sluit af met een reflectie op de gemeenschappelijke elementen in de studies. De conclusie geeft een beeld van de rol van het openbaarmakingsbeleid op het gebied van informatiebeveiliging en van de gevolgen, gegeven de resultaten van deze studies, voor het Europese beleid inzake de melding van datalekken.

Chapter 1 Introduction

1.1 Problem Definition

New developments substantially increase dependency on technologies and services and give rise to new risks and security challenges, which require further investigation of the social, technical and economic complexities of the cyberworld. On the one hand, both consumers and providers benefit from effective and efficient technologies. On the other hand, new and severe security threats and vulnerabilities can expose consumers and providers to unwanted consequences and substantial losses.

In response to rising and evolving cybersecurity challenges, spending on information security has grown steadily and could, at the current rate, eventually exceed a reasonable level of sustainability. Furthermore, both governments and market-oriented organisations must carefully balance the trade-offs between security and privacy. We urgently need a better understanding of new socio-technical-economic complexities, which requires both reconsideration of traditional cybersecurity issues and investigation of new and unexplored research directions. Indeed, it is impossible to eliminate all vulnerabilities and to obtain complete security via technology alone.

Recent interdisciplinary studies have examined and clarified many economic and behavioural dimensions of security. One of the key aspects in the field of *information security economics* on which researchers have concentrated is disclosure policies, specifically **data breach notification laws** (DBNL). Data breach notification laws require any business that has suffered a data breach, or believes it might have suffered a data breach, to notify customers about the incident that entails an unauthorised acquisition of unencrypted and computerised personal information. DBNLs provide incentives to the parties who owe the notification duty to minimise the number of triggering events[1] and also enable the affected

[1] Faulkner 2007 defining events requiring notice.

third parties to mitigate the consequences of the breach and to make prudent choices in the future.

What exactly is a ***data breach***? The best way to explain a data breach is to refer to the 'Collection #1' which occurred in January 2019. It is one of the largest data breaches of all time, comprising 772.904.991 unique emails and 21.222.975 unique passwords. Troy Hunt initially reported the breach that appears to derive from many different sources, not a single corporate entity. The large volume of the data was contained in 12.000 separate files, 87 GB of data accessible on hacking forums.[2] The files contained 'dehashed' passwords. Consequently, the hackers were able to circumvent methods used to scramble the passwords into unreadable strings and to expose them. In terms of the number of individuals involved, this violation is far below the scale of Yahoo's breach, where 3 billion user accounts were compromised. That being said, its reach was significantly broader than the Marriott/Starwood Hotel breach of 2018, when information on 383 million guests was accessed, or the LinkedIn case of 2012, when information on 117 million users was stolen.

'Collection #1' is a clear example of a data breach and the potential scale of breaches. It more specifically represents 'a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed'.[3] Data breaches are the core issue discussed in this dissertation, which focuses in particular on the instrument used to mitigate breaches: data breach notification laws. This study analyses the effectiveness of such laws and the areas of potential improvement as part of the effort to fight a phenomenon that is and will grow as part of our lives. Data breaches are a phenomenon that affects all sectors, including companies offering security services to governments (Walsh 2014). Moreover, it is a phenomenon that generates costs at the individual, business and societal level. In fact, both public and private costs result from attacks on firms' information technology networks. Successful attacks result in data breaches which incur private damages from business interruption, reputational harm and investigation forensics, among other

[2] https://finance.yahoo.com/news/apos-email-one-773-million-163936123.html?guccounter=1

[3] As defined in the General Data Protection Regulation (GDPR) Article 4 (12).

damages. Social losses also result from the exposure of individuals' personal information, motivating state, national and international policymakers to enact legislation to manage these costs.

The Identity Theft Resource Center (2018) reported a total of 1.244 breaches in 2018 in the United States alone, with over 446 million exposed records. There are no comparable statistics from European countries, as reporting of data breaches only recently became mandatory in all sectors in Europe. Additionally, for those sectors where it was already mandatory, public communication of data breaches was not assured. However, following the botnet, phishing and ransomware trends, which we can consider as a proxy for data breaches, the situation is no better in Europe (e.g., see Microsoft 2018, 8, 31, 48).

Despite the exponentially growing economic impacts of cyberspace, e-commerce and social networking, and although security and privacy are ranked high on the agendas of organisations and governments, the existing economic modelling that analyses data breaches is limited. The phenomenon derives firstly from cyberattacks (e.g., hacking and malware). However, it is also facilitated from a mix of other factors that generate vulnerabilities, such as poor business processes, human error and lack of awareness. It is not feasible to eliminate all such vulnerabilities, nor the human mistakes generating security failures. It is clear that some security systems succeed more than others. Several theories from economics and other social sciences help explain why this is the case. Data breaches can ultimately be contained by implementing proper measures which require understanding the underlining dynamics (including economic dynamics) of the phenomenon and the solutions chosen to mitigate its effects (e.g., DBNL).

Information disclosure policies are increasingly used at all levels of authority, and there is every reason to think that public demand for information about corporate and government actions will continue apace (Kraft, Stephan & Abel, 2011). Disclosure policies and related laws require private companies to disclose certain information for the benefit of consumers. For example, long-standing food labelling requirements, such as calorie counts and fat and protein contents, provide consumers at least some of the information they need to make smart choices about their food purchases. Hospitals must publicise performance results for

certain medical procedures (Twerski and Cohen 1999). Manufacturers of household appliances must label their products with energy-efficiency ratings. Factories must disclose information about toxic releases and workplace injuries. Sunstein (1999) termed this trend 'regulation through disclosure' and characterised it as 'one of the most striking developments in the last generation of American law'.

Regulation through disclosure has two effects. Firstly, it provides incentives to the party who owes the duty to minimise the number of triggering events and to avoid public shaming through better procedures. Secondly, it enables affected third parties and the wider community to make prudent choices, take protective actions and bring market pressures to bear on repeat offenders (Shafer 2016). DBNLs perform both tasks and appear to be the solution adopted by most countries in order to tackle the issues related to data breaches and their consequences, namely identity theft.

The U.S. was among the first to act; its first data breach notification law was enacted in 2003, obliging breached organisations to notify breached consumers. Europe recently followed with the implementation of the NIS Directive in 2016 (with breach notification duties for operators of essential services and for digital service providers) and of the general data protection regulation (GDPR) affecting all sectors in EU[4] from 2018, which also requires notification to consumers in case of a breach. Besides, Australia and Canada are among the other countries that have only recently implemented notification duties, both in 2018.

Data breach notification laws are promulgated under the theory that customers have the *right to know* when their personal information has been stolen or compromised. Furthermore, data breach notification laws provide organisations with incentives to take adequate steps to secure personal information held by them (*sunlight as disinfectant principle*). In order to properly evaluate data breach notification laws and the extent of their achievement of these two objectives, an interdisciplinary approach is required, which is the essence of this research.

[4] Single European countries adopted regulations with data breach notification provisions even before 2016 (e.g., Ireland in 2010 and the Netherlands in 2012 with respectively the Personal Data Security Breach Code of Practice and the Telecommunications Act.

1.2 Research Questions

The goal of this dissertation is to identify opportunities for improving cybersecurity through disclosure policies. This objective requires understanding the security behaviour and incentives of breached organisations and assessing how they might be affected by specific DBNLs. Therefore, this dissertation poses the following primary research question:

> *What are the effects of the provisions of data breach notification laws on (1) communications issued by breached organisations to their customers; (2) the timing of breach detection and reaction; (3) the number of data breaches reported; and (4) the volume of identity theft stemming from data breaches?*

The United States is the area of investigation for this research, chosen based on data availability. The empirical studies under examination were conducted in the U.S., where there are more than 15 years of related legislation and various databases which collect information on data breaches and identity theft. However, the potential implications of the findings are relevant not only within the area of analysis (U.S.), but also for Europe, where the GDPR is now in force.

The main research question is divided into several areas of inquiry or sub-questions. The sub-questions are explored through five chapters in this dissertation. The following paragraphs outline the contents of each chapter and their corresponding sub-questions.

Study 1. Name and shame: Introducing data breach notification laws and a proposal for the evaluation framework (chapter 2)

After analysing the importance of information availability and how disclosure policies contribute to increase information availability in the DBNL introductory chapter, chapter two focuses specifically on data breach notification laws. The chapter first presents a theoretical model with a reference framework for the analysis of data breach notification laws and their effects. It then provides a brief literature review as an introduction to the subsequent empirical studies. In short, the chapter aims to answer the following research question:

(1) What are the different elements to consider in evaluating data breach notification laws?

Study 2. Let's not sugarcoat it: An investigation into communication styles used for breach notifications (chapter 3)

The first empirical study examines the dilemmas a company faces regarding how to notify their customers in case of a data breach. The choices that a company makes for the missive contents and style are decisive in achieving a prompt customer reaction against identity theft and ultimately in shaping the relations between customers and the organisation. Beginning from the various regulations in place in the U.S., chapter three proposes a specific evaluation framework which takes into consideration the intersection between business communication and information security created by DBNLs. The methodology enables analysis of the contents of breach notification letters sent in the U.S. in 2014. The letters are classified based on elements that can be isolated and analysed (e.g., clarity of incident description, communication tone on possible consequences). As a result, six message types are identified to assess how companies in each case fulfilled the notification duty. This analysis provides insights on how companies, given the loose provisions on notification contents, have the option to sugarcoat breach notifications, which results in negative consequences for consumers. Specifically, the performed analysis answers the following research question:

(2) What are the core elements of consumer notification letters and how do company decisions on what to include and how to express the message define specific letter types?

Study 3. Slow and (not) safe: An investigation into company reactions to breaches (chapter 4)

The second empirical study investigates the sufficiency of data breach notification laws by examining company reactions to breaches. Based on the analysis of the notifications issued in 2014, three observations for the development of law are presented. First, the study raises the question of underreporting and suggests a possible option for reducing its prevalence. Second, the study identifies notification of the breach and of the date on which it occurred as essential to foster consumers' timely reactions. Finally, the study proposes stricter regulation of the contents of notifications in order to prevent companies from minimising, in the eyes of

the customer, the actual risk deriving from a breach, and to incentivise them to act faster and safer for the benefit of their customers. This study answers, therefore, the following research question:

(3) How different are the choices organisations make in terms of if they notify, what they notify, and when they notify?

Study 4. Estimating the size of the iceberg from its tip: An investigation into unreported data breach notifications (chapter 5)

By leveraging the findings of the previous chapter, the third empirical study aims to tackle the issue of the number of breaches that have not become public knowledge. The chapter examines the available breach statistics to model the impacts of DBNL provisions on the number of known data breaches and breach notification times while controlling for sector and state differences. Moreover, the study estimates the number of breaches with notifications issued by breached organisations, but not publicly reported. If we consider the set of data breaches as an iceberg, the dimensions of what is visible (i.e., data breaches that are first detected, then notified and finally publicly reported) and what is hidden below the water surface have proven to be dependent on how DBNLs are designed. The focus is on the following research questions:

(4) What effects do specific DBNL provisions have on reported data breaches?

(5) How large is the portion of data breaches we are unaware of?

Study 5. More than a suspect: An investigation into the connection between data breaches, identity theft and data breach notification laws (chapter 6)

Given that limited research has been conducted on the relationship between data breaches and identity theft and on the impact of DBNLs on the prevalence of identity theft over time, the purpose of the fourth study is (i) to further investigate the correlation between data breaches and identity theft and (ii) to understand if and how DBNLs in practice have an effect on data breaches and identity theft by modelling the impact of DBNL enactment (and revisions) on the number of data breaches and incidences of identity theft. We collected data on data breaches and identity theft over a 13-year time span (2005-2017) in order to arrive at robust

findings on the possible correlation and impact. The following research question is addressed:

> *(6) What effects do DBNL enactment and revisions have on incidences of data breaches and identity theft?*

Finally, the conclusions reflect on the findings of the individual studies, and investigate the trends revealed in terms of the dissertation's main research question. The chapter broadens the discussion to the GDPR, focusing on the key elements of DBNL that make the laws effective in view of the implementation of the European regulation on security and data breaches. The related research sub-question is, therefore, as follows:

> *(7) What are useful lessons learned for EU regulators on managing and monitoring the implementation of the GDPR?*

1.2.1 Contributions

The below paragraphs list this dissertation's practical and academic contributions, which are discussed in detail in the concluding chapter.

First, this research contributes substantively to better understanding the challenges of data breaches and their consequences. The findings on information availability, communication styles in data breach notifications, timing of breach identification and notification, hidden breaches and identity theft offer significant insights for further policy discussions and development.

Second, the research contributes to the current literature on the economics of cybersecurity through applying methodological innovations in analysing and interpreting data breach notifications and the overarching regulations.

Third, it further contributes to the economics of information security literature by advancing discussions on the role of disclosure policies in information security governance. The dissertation concludes by reflecting on how the effect of data breach notification laws in the U.S., and the GDPR in EU, can be improved through specific actions.

1.3 Dissertation Outline

The remainder of this dissertation is organised in seven chapters, which are listed in **Table 1.1,** along with the corresponding publications.

Chapter 2 introduces the central role of information availability for cybersecurity, highlighting the need for disclosure policies. It then illustrates the theoretical model employed to analyse one of these policies, namely, data breach notification laws. The chapter subsequently reviews the literature related to disclosure regulations as the overarching context for this research.

Chapters 3 to 6 address the four empirical studies. These four chapters have been published partially or fully in journals or peer-reviewed conferences before the culmination of this thesis.

Finally, chapter 7 concludes the dissertation by drawing broader conclusions from the studies to answer the main research question and to illustrate the implications for European data breach notification policies.

I was fortunate enough to conduct two of the empirical studies and to write the theoretical paper in collaboration with excellent researchers, who are also listed in **Table 1.1**. I gratefully acknowledge their contributions in the section *authorship contribution*, located at the end of this dissertation.

Table 1.1. Dissertation overview

Chapter	Publications
2	Bisogni F., Cavallini S., & Trocchio S. (2011). Cybersecurity at European Level: The Role of Information Availability. Communications & Strategies, Number 81, The Economics of Cybersecurity, pp.105-123, March 2011. Bisogni F (2013). Evaluating Data Breach Notification Laws - What Do the Numbers Tell Us? Conference TPRC 2013, Virginia, September 2013.
3[5]	Bisogni F. (2015). Data Breaches and the Dilemmas in Notifying Customers. WEIS 2015 - 14th annual Workshop on the Economics of Information Security, At Delft University of Technology, the Netherland, June 2015. Bisogni F. (2016). Proving Limits of State Data Breach Notification Laws: Is a Federal Law the Most Adequate Solution? Journal of Information Policy Vol. 6, pp. 154-205 - Penn State University PressConference.
4	Bisogni F. (2016). Proving Limits of State Data Breach Notification Laws: Is a Federal Law the Most Adequate Solution? Journal of Information Policy Vol. 6, pp. 154-205 - Penn State University PressConference
5	Bisogni F., Asghari H., & Van Eeten M. (2017). Estimating the size of the iceberg from its tip. An investigation into unreported data breach notifications. WEIS 2017 - 16th annual Workshop on the Economics of Information Security, At La Jolla, US, June 2017
6	Bisogni F., & Asghari H. (2020). More than a suspect. An investigation into the connection between data breaches, identity thefts and data breach notification laws. Journal of Information Policy Vol. 10, pp.45-82 - Penn State University PressConference.

[5] Part of the JIP paper (2016) used also for chapter 3.

Chapter 2 Name and shame: Introducing data breach notification laws and a proposal for the evaluation framework

Starting from an analysis of the cybersecurity investment behaviour of organisations, we focus on disclosure policies and examine data breach notification laws used to reduce data breaches and their effects, namely reducing identity theft. The research sub-question addressed in this chapter is, 'what are the different elements to consider in evaluating data breach notification laws?'

2.1 Introduction

Before introducing data breach notification laws, it is essential to reflect on the issue of information availability, its relevance for cybersecurity in general and its connection to disclosure policies. This chapter explores the relationship between security investments and costs suffered as a consequence of cyber-attacks through the lens of information availability. The proposed model concerns the lack of information that characterises businesses and information and communication technology (ICT) operators' investments in cybersecurity and suggests policy actions that may improve the security level for all relevant actors. We then focus on one of the policy actions proposed, namely disclosure policies, and broaden the discussion to data breaches and the related regulation (i.e., data breach notification laws). We subsequently propose an evaluation framework for DBNLs and illustrate the state of current research. Finally, the concluding remarks anticipate the areas of investigation to be addressed in the following chapters.

2.2 The importance of information availability

With the spread of information and communication services and the emergence of related threats and vulnerabilities in recent years, cybersecurity has evolved from a valuable economic good to a societal need.

Business users, public authorities and citizens demand secure information systems, and ICT operators have established investment strategies in order to provide ICT services at a suitable level of security. For an organisation, the optimal level of investment in cybersecurity is the level which provides protection that minimises its expected costs in case of a cyberattack. This optimal solution occurs when marginal security investments equal the expected marginal costs that the operator would incur. Nevertheless, market failures may impede the achievement of the optimal level of investments and the consequent optimal level of security (Bruck, Karaisi & Schneider, 2006).

Gordon and Loeb (2002) constructed a model to determine the optimal amount of investment needed to protect a given set of information. Considering the vulnerability of information systems, their main finding was a biased behaviour on the part of the operator: they found that a firm spends only a small fraction (approximately 37%) of the potential loss that would result in case of a breach occurrence. According to this model, the level of cybersecurity investment can be defined on the basis of the expected loss (E(L)) associated with its available information set, with L representing the incurred loss in case of a cyberattack. The expected loss is the product of the probability of the threat occurrence, t, multiplied by the vulnerability of the system, v (which is the probability of threat effectiveness) and the potential loss due to the threat realisation, λ.[6] In order to avoid significant unexpected losses, the organisation establishes a level of security, S, as a function of the implemented security investments, Is, and of the level of vulnerability of the system, v.

In order to illustrate the investment choice in Gordon and Loeb's model, the relationship between the optimal investment choice of the organisa-

[6] The random effect of the exogenous factors affecting the model structure could be addressed inserting an uncertainty variable into the model. The most likely uncertainty factors would be the probability of threat realization t and the potential loss λ. Both of them affect the expected value of loss due to the lack of information randomly affecting the ICT network actors. Assuming that the uncertainty variable would be inserted in the form of white noise, with zero average value independent and identically distributed, (which implies no autocorrelation), expected value of this uncertainty would not affect the final outcome of the model. For reference, see Greene (2007).

tion and expected loss can be plotted (Figure 2.1), with the level of investment in security, I_s, on the x-axis and the expected loss E(L) on the y-axis. As common sense suggests, a lower level of investment corresponds to a higher expected loss in case of cyberattack and vice-versa. The firm chooses a level of cybersecurity investment according to its risk attitude and risk assessment. The level of investment chosen depends on the operator's risk propensity. If the operator is risk-averse, the firm will prefer a lower level of expected loss and thereby increase its current costs; if the firm is risk-loving, it will accept a high-risk situation in order to increase its current benefits (e.g., reduced security costs).

We assume the risk neutrality of the operator.[7] Risk neutrality implies that the value of the level of cybersecurity investment is equal to the value of the expected loss, such that the optimal investment level chosen by the operator is represented by the intersection between the optimal choice curve and the tangent line representing the risk attitude of the agents. In Figure 2.1, the intersection point is O^*, the optimal cybersecurity choice, with a level of implemented investment I^*_s and consequent expected costs $E(L)^*$ for cyberattacks.

[7] An agent is risk neutral when he/she is indifferent to sustaining current expenses in order to implement cybersecurity provisions or to bear the same expected expenses in the future to recover the losses caused by a cyber-attack. The idea of the risk aversion/propensity could be linked to inter-temporal choice, but it is crucial to consider the presence of a choice between certain and uncertain choice and not only between current options and future option. For reference, see Kreps (1991) and Mas-Colell et al. (1995).

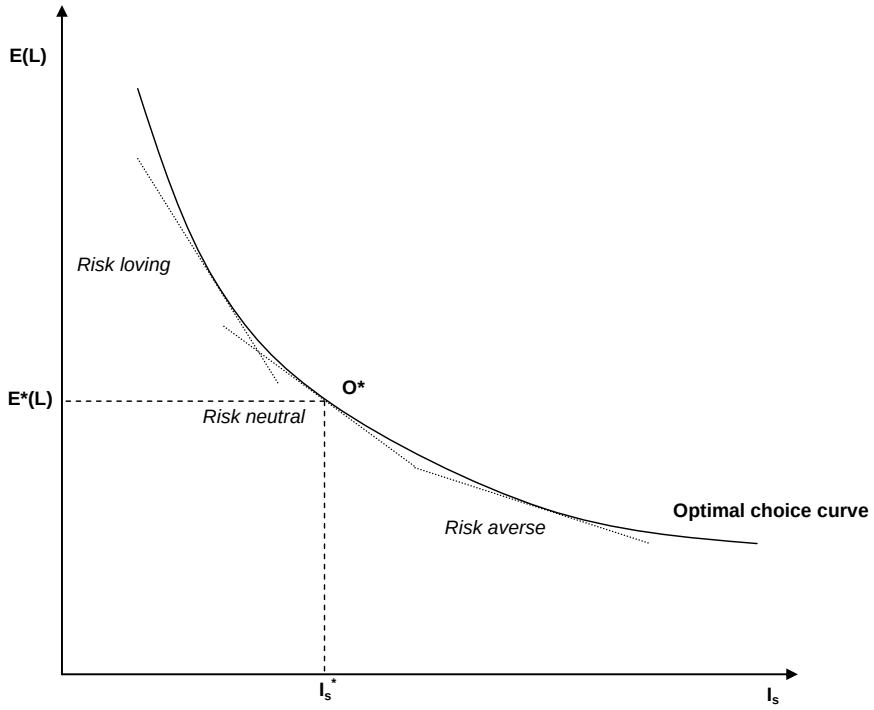


Figure 2.1 The optimal level of investment in cybersecurity.

O: optimal cybersecurity choice with a level of implemented investment (I^*_s) and the consequent expected costs for cyberattacks ($E(L)^*$).*

The optimal choice in this cybersecurity framework relies on the assumption that the organisation possesses complete information on cyber-crime effects and makes a proper assessment of cyber-attack risk. The expected loss and the subsequent investment choice are determined as the result of estimation of the probability of threat occurrence (t), effectiveness in breaching the information system (v) and the economic consequences of a breach (λ).

In the real world, complete information on cyberattacks and related risk is not available, in large part because cyberattack techniques evolve rapidly and are becoming increasingly sophisticated. Besides, firms targeted by cyberattacks are often reluctant to publicly communicate and report to the authorities any disruption in services as well as the causes, frequencies and costs. This behaviour can be ascribed to concerns over suffering reputational damages, breaking confidentiality obligations and

being sued on the grounds of liability. Moreover, the sensitivity of information on cybersecurity incidents makes information sharing a particularly risky issue, hindering the development of a confident and fruitful environment. From the perspective of a single operator, there are no immediate advantages to sharing information on past attacks.[8] However, the community members would gain from better information on the cyberattack framework.

The reluctance to share information about experienced cyberattacks results in incomplete knowledge on cyber-risk, which leads to under-estimation of the probability and impacts of cyberattack. These circumstances influence the extent of implemented security provisions and the realised security investment. Because firms are not adequately aware of the real extent of cyber-risk, the chosen level of investment is lower than the level which would be desired by the operator with complete information.

These assumptions are supported by the results of a leading study on information sharing by Galor and Ghose (2004). Their analysis, featured in the article 'The Economic Consequences of Sharing Security Information', examined the competitive implications of information sharing on breaches and the level of investment dedicated to security. The main conclusion was that market characteristics affect incentives for information sharing among competing firms, but information sharing encourages additional security investments.

In cybersecurity management, the availability of information guarantees proper risk assessment, which is essential for an efficient protection strategy. For a single firm, any security investment choice depends on its evaluation of the balance between the potential costs of disruptions and the benefits of protection. A proper risk evaluation is the result of assessments of threat probability, vulnerability and potential threat damage. Thus, limited information on the potential damages of a cyberattack may lead a firm to underestimate the effective risk, thereby lowering desired investment. In addition, a large body of literature, starting from the

[8] In the perspective of the operator, the immediate advantages of sharing information are not enough to overcome the potential risk of reputation loss coming from breaches or improper disclosure.

seminal contribution of Dixit and Pindyck (1994), regards the uncertainty of market conditions (for example, the probability of occurrence of threats) as a costly condition in case of investments. An organisation investing in security at a specific moment loses the possibility to wait for better market conditions, thus bearing higher costs. Several empirical studies have highlighted situations in which such costs are very high and particularly affected by the degree of market uncertainty, leading to significant security underinvestment compared to the theoretically optimal level.[9]

The effect of the operator's lack of adequate awareness of cyber-risk is represented in Figure 2.2, with the perceived optimal choice curve under the optimal choice curve. The threat probability and the cyberattack impact, which contribute to the shape of the optimal choice curve, are biased by the absence of a proper level of information and are perceived by the operator as equal to $t^p<t^*$ and $\lambda^p<\lambda^*$.[10] Assuming that the firm is risk-neutral, the resulting optimal level of investment (I^{**}_s) is lower than the previous level (I^*_s) due to an expected cost $E^{**}(L)$ according to the organisation's perception. Considering the real level of threat probability (t^*) and the real cyberattack impact (λ^*) for a level of investment I^{**}_s, the expected loss that the firm would sustain is $E^*(L)'$, which is higher than estimated.

This analysis demonstrates that the lack of complete information on cyberattacks may lead to insufficient awareness of the related risk (represented in the position of the perceived optimal choice curve). As a result, organisations invest in cybersecurity in a suboptimal way, with a level of implemented security provision insufficient not only for social demand, but also for the firm's preferences. In this context, cyberattacks

[9] On this topic interesting articles have been written by Caballero (1991) and Abel & Eberly (1999).

[10] The vulnerability variable, v, composing the expected loss, is considered constant at least in the short term. In fact, it is assumed that the vulnerability of the ICT operator is a technological concern linked to the variability of the cybersecurity environment, where dangerousness and frequency of cyber-attacks change only in the long-term. In this study, vulnerability is considered constant as "protective capacity" and can be effectively modified in the mid-term only through current security investments implemented by the ICT operator.

cause greater economic damages than expected by operators, with amplified consequences on public authorities/bodies, businesses and citizens

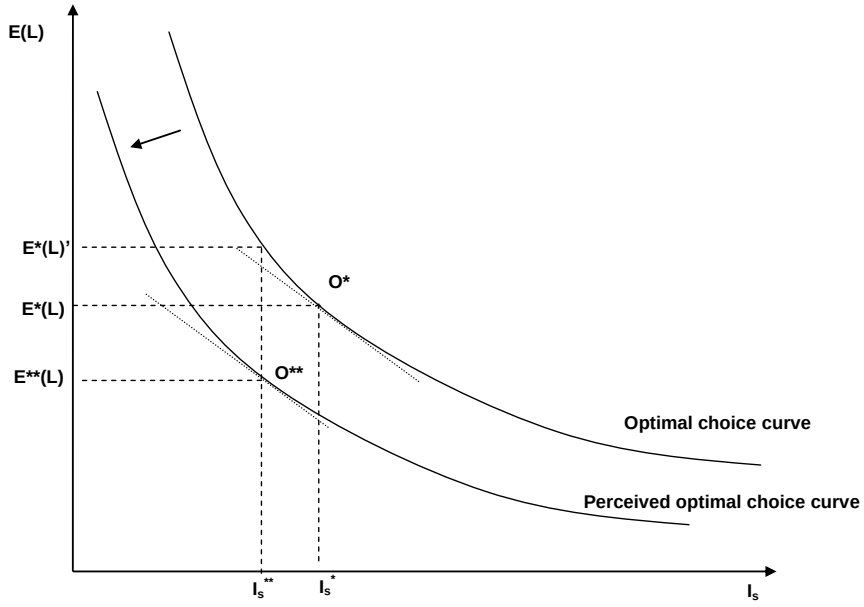


Figure 2.2 Effect of lack of information on cyber-security on the optimal investment level.

O: optimal cybersecurity choice with a level of implemented investment (I^*_s) and the consequent expected costs for cyberattacks ($E(L)^*$).*

2.3 The improvement of information availability through disclosure policies

Regulatory initiatives against cyberattacks undertaken at the institutional level are focused on the critical role of information for mitigating cybercrime, particularly given the network nature of information systems and its consequence on security. Most proposed measures aim to increase social awareness of the effects of cyberattacks and to reduce the biased optimal choice behaviour of operators. Many policies also target other stakeholders who can directly impact the security provisions.

In order to improve cybersecurity, policymakers can establish an incentive framework through disclosure policies. Such policies and related laws require companies to disclose certain information for the benefit of the community and consumers. Information disclosure and the indirect promotion of information sharing are the primary goals of disclosure policies.

Reporting may reduce the effects of cyberattacks. The increase in shared information from impacted organisations on threats, vulnerabilities and incidents may improve the risk evaluation process on which security and resilience investment relies. Among the several ways to address the challenge of lack of information, one solution is the implementation of homogenous practices for disruption reporting, which allows authorities to have a complete overview of the emerging threats and related vulnerabilities and to collect consistent data for social risk evaluation.

A key element for overcoming the lack of information is, therefore, a universal strategy for collecting it and the widening of sources of incidents. Despite the effort made globally by central institutions and bodies to adopt appropriate technical measures to harmonise incident reporting practices, existing practices unfortunately remain extremely heterogeneous, reducing the effectiveness of the information sharing.[11] Thus, appropriate reporting schemes should be established in line with the general reporting objective and shared data collection. Information collection from organisations by public authorities/bodies may have the immediate effect of reducing the distance between the perceived probability of cyber-threats (t^p) and the effective probability of threats (t^*), leading to new actions against cyber-attacks (e.g., imposition of security standards, cooperation at international level). Society as a whole would benefit from the increase in cyber-security sustained by additional investments by firms (towards $I_S^*$for the reduced $E(L)$).

[11] According to the ENISA 2009 report "Good practices on reporting security incidents" differences in incident reporting exist between Countries especially in terms of objectives: emergency or incident response, incident prevention and legal rectification.

2.4 Data breach notification law and a proposed theoretical evaluation model

As highlighted in the previous section, the rapid and ongoing evolution of the ICT sector makes it difficult for stakeholders to maintain a satisfactory understanding of cybersecurity risks. Comprehensive understanding of cyber-risks represents an important driver for policymakers to identify and manage the measures that are currently causing market actors to protect systems at a suboptimal level and, eventually, for enhancing security quality and the capabilities of software and communications systems through adequate regulations. In this context, an increasing regulation effort has been launched to impose mandatory disclosure policies for security breaches in most affected economic sectors. On the basis of the analysis performed on information availability, it is clear that disclosure policies, which generate disruption reporting on security incidents, can truly support a more secure cyber environment.

The starting point of this dissertation is to propose a framework for the evaluation of data breach notification schemes in order to offer a basis for debate in additional geographical areas where data breach notification regimes are more recent. This research aims to set the basis for a comprehensive investigation of information disclosure as a policy strategy for data protection and continuation of operational services. This section proposes a conceptual model to study the effectiveness of data breach notification laws. The model captures the primary causal relations around laws and regulations and the relevant actors (e.g., government, sectors, community, law enforcement, media). A proper evaluation of the effectiveness of laws and regulation is made possible not only by analysing the number of notified data breaches over time, but more specifically by enabling the assessment of effects directly related to the behaviour of single actors and their interdependencies. This assessment includes the evaluation of economic, legal, crime and response effects.

Data breach notification serves cybersecurity purposes by encouraging business entities that hold personally identifiable information to protect that information better. It also supports protection against identity theft, a crime that can easily destabilise cybersecurity. Data breach notifications are a consumer privacy phenomenon and are addressed internationally with data breach notification laws that vary in strictness and the

rigour of enforcement. In Europe, the new General Data Protection Regulation (GDPR) includes, for the first time, a broad breach notification requirement, following the U.S. data breach notification laws. The requirement applies not only to organisations based in the EU, but to all organisations that process information and intend to the offer goods or services to people in the EU or monitor their behaviour in the EU.

Data breach notification has become a significant reform topic in the European Union. For example, the introduction of data breach notification requirements for the electronic communication sector introduced in the review of the ePrivacy Directive (2002/58/EC as amended in Directive 2009/140) is an important development with the potential to increase the level of data security in Europe and to foster reassurance amongst citizens regarding how their personal data is being secured and protected by electronic communication sector operators (ENISA 2011). The GDPR reinforces and extends to additional actors the notification requirement in cases of a personal data breach. In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, report the breach to the competent supervisory authority, unless the breach is unlikely to result in a risk to the rights and freedoms of natural persons. When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the breach to the data subject without undue delay. Businesses could be fined up to €10 Million or 2% of their global revenues for not being compliant with the required procedures and mechanisms.

While the EU regulation has been in place since May 2018, the Australian Senate voted in 2016 to pass the Privacy Amendment Bill (Notifiable Data Breaches), which amends the Privacy Act of 1988 (Cth) (Federal Register of Legislation), 'to introduce mandatory data breach notification provisions for agencies, organisations and certain other entities that are regulated by the Privacy Act'. Under the amendments, organisations must report an 'eligible data breach'. Other countries such as Canada and New Zealand have also followed, with law reform proposals being put forward in their jurisdictions.

The first data breach legislation was enacted in 2003 in the United States (California Civil Code § 1729.98) and became the basis for further legislative developments throughout the U.S.. All but three U.S. state legislatures have now enacted DBNLs, which vary in many elements from state to state. For example, the data breach notification law in California requires any business that suffers a data breach, or believes that it has suffered a breach, entailing an unauthorised acquisition of unencrypted and computerised personal information to notify California residents about the incident.[12] The Attorney General must also be notified if more than 500 residents' data are involved in the security breach. Law enforcement agencies can request a delay if the notification would impede a criminal investigation. Furthermore, individuals are to be notified within a timeframe that is expedient and without reasonable delay. Notifications can take different forms including by letter, electronic notification or substitute notice, which entails 'conspicuous posting' on the organisation website or via state media sources. However, some data breaches are exempt from notification. These include breaches of encrypted personal information or 'good faith acquisitions' of personal information by an employee or agent of the breached entity.

Determining whether the current data breach notification laws are too weak or too strong for pursuing their goals is not easy. It is difficult (and perhaps impossible) to assess the aggregated costs and benefits for both consumers and firms of different privacy regimes in purely monetary terms (Romanosky and Acquisti, 2009). Even just understanding the landscape is a challenge (Romanosky, Hoffman & Acquisti, 2014). This section establishes a framework in order to thoroughly investigate the landscape, enabling vertical analysis of DBNLs with the possibility to extend to broader security incidents.

[12] California Civil Code § 1729.98

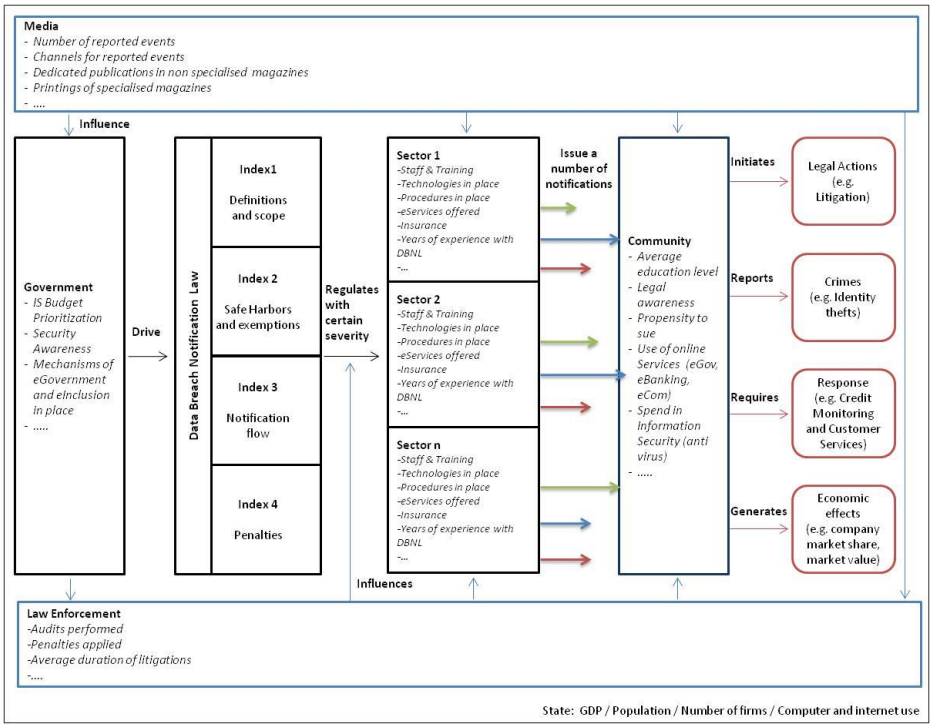


Figure 2.3 DBNL Theoretical Model

Figure 2.3 above represents a model of the factors that influence notifications as required by DBNLs: apart from the state context variables, the key actors are the government, the economic sectors including the relevant organisations, the community, the media and law enforcement.

These actors' behaviours shape the consequences of the enactment of DBNLs, and therefore may mitigate or strengthen the effects of the notifications issued on the basis of the law provisions. Data protection is not the work of any individual company or government agency. It is the work of many interconnected, interdependent stakeholders (Wolff 2018). The same applies for data breach notification.

The model's starting point is the **government** decision to adopt a specific DBNL. The expertise, resources, and beliefs of governmental actors can be expected to affect the resulting regulation. To better understand the motivation behind the passage of each specific DBNL, it is prudent to analyse the extent to which the state government relied on (or promoted) mechanisms that are vulnerable to data breaches, such as eGovernment and eInclusion, and to consider the degree of security awareness at the

institutional level. The government budget prioritisation of information security, for example, can provide relevant information to determine the degree of importance given to security. More specifically, to perform this form of analysis on states, which are politically bounded units that shape policy choices (Kraft et al. 2011), the principal-agent theory can be helpful, as it is the dominant approach to understanding the extent to which politics influences bureaucratic behaviour and the policy choices of state agencies (Gerber and Teske 2000). In these ways, the government drives the decision regarding the characteristics of **data breach notification laws,** as summarised below.

The scope of laws in terms of the definition of personal information may vary. The law firm Baker Hostetler provides a standard definition of personal information based on the definition commonly adopted by most states.[13] However, 30 states have a broader definition for personal information than this standard one, which also broadens the definition of data breach.

Moreover, in some states the trigger for notification is not only data acquisition, as in California, but also data access. In addition, in six states the breach of security is not only limited to electronic records, but also includes paper records. In terms of coverage, in all 47 states with DBNLs the notification requirements describe the categories of entities to which the law is applicable. There are two broad categories: entities that own or license computerised data and entities that maintain computerised data. Whereas all the state laws apply to entities that own or license personal information, one-fourth of the state laws also apply to entities that maintain personal data. Almost all states provide notification exemptions, for example, for encrypted data (30 states) and publicly available government records (all). Some states also provide exemptions for investigation purposes as determined by law enforcement and for breaches that are deemed either immaterial or not 'reasonably likely to subject the customers to unauthorized disclosure of personal information' after a required proper risk-of-harm analysis. Furthermore, exemptions are also provided through legislation regarding specific sectors, such as the Gramm-Leach-Bliley Act for financial institutions or the

[13] Baker Hostetler. State Data Breach Statute Form (2014).

Health Insurance Portability and Accountability Act for healthcare providers, or through compliance with rules, regulations, procedures or guidelines established by the primary regulator.

The level and the limit of penalties may also vary. There are two possible limits enacted by DBNLs: those related to a single security breach or limits related to the number of records accessed/acquired as a result of the breach. The majority of states left the maximum measure of a penalty undefined, while 26 states have included a limit either for a single breached record, for a single breach or both. The penalties limit can be linked to the duration of the missing notification, linked to the extent of the caused damage or expressed as a flat value, ranging from $10.000 (Arizona) to $750.000 (Michigan). The penalties, and therefore the financial burden for companies, can become more severe in case of a private cause of action, which may result in civil and penal consequences for the involved organisations. Only in 16 states do residents have the right to take private action against companies that disclose their information; in the remaining states this activity must be performed by the Attorney General.

A further element that takes into consideration the reputational risk of companies is the compulsory nature of notifications to be delivered to authorities in addition to those delivered to residents whose data have been subject to access or acquisition. Few states decided to include such notifications to third parties, specifically to the Attorney General and/or consumer reporting agencies (22 and 31 states respectively).

The mandatory content of the notice to be sent to residents, as specified in the law provisions of 15 states, also plays a role in evaluating the potential reputational effects of a breach for a firm. Finally, all states require that the notice is provided in the most expedient time and manner possible and without unreasonable delay, consistent with the legitimate needs of law enforcement. Only seven states add to this statement a specific maximum timeline of 30 or 45 days after the breach discovery.

Table 2.1 Key DBNL provisions

Law element	Explanation	Present in n. states
Broader definition of personal information	It indicates whether the statute covers more information beyond the standard definition of personal information (PI.)[14] An expanded definition of PI includes other pieces of data, most notably health and medical information.	30
Notification by access	A breach of security is defined by data acquisition. However, in some cases the definition of breach is extended to unauthorised access.	3
Limited coverage	In some cases, the laws do not apply to organisations that own, license or maintain data that includes PI, regulating only those case where the data including PI are owned and/or licensed.	35
Type of data	Acquisition includes acquisition by photocopying, facsimile, or other paper-based methods.	8
Encryption	This provision describes the requirements for receiving an exemption from a state notification law. States in which this exemption is easiest to attain have laws exempting notification if breached data were encrypted or redacted.	30
Risk-of-harm analysis	It refers to whether a statute requires a breached organisation to notify only if the organisation determines that the breach constitutes a reasonable likelihood of harm to the customer.	39
Other applicable laws	When compliant with other laws, such as the Gramm-Leach-Bliley Act or HIPAA, or with the primary regulator, organisations are exempted from data breach notification law provisions.	43
Own notification policy	Such exemptions exist when a state allows an organisation that maintains its own notification procedures as part of its information security policy to be deemed in compliance with the state notification law, so long as the organisation does disclose breaches.	14

[14] An individual first name or initial in combination with a last name and a social security number, driver's license number, state ID card number, or financial account number (see Baker 2014).

Notification to credit reporting agency	If an entity provides notice to more than a certain number of persons (it varies from 1.000 to 10.000 according to the state) at one time, pursuant to the general security breach section, the entity shall notify, without unreasonable delay, all consumer reporting agencies that compile and maintain files on consumers on a nationwide basis.	31
Notification to Attorney General	In the event a business provides notice to an affected person pursuant to this section, the business shall notify without unreasonable delay the state AG's office of the nature of the breach, the number of consumers affected by the breach, steps taken to investigate the breach, steps taken to prevent a similar breach in the future and information regarding the timing, distribution and content of the notice. The AG's website contains a form to be used for notification.	22
Notice requirement	In some cases, certain elements of the notice are mandatory and identified in the law. Such elements include the type of personal information subject to an unauthorised access or acquisition, the specification of the reporting entity's name and contact information so that affected individuals can obtain additional information and specific information on what has happened (a general description of the breach incident).	15
Specific time frame	This provision specifies that notification must occur within a given number of days (usually 30 or 45). Notification laws without a specific time limit require notification as quickly as possible and without unreasonable delay.	7
Private right of action	This provision gives customers the ability to sue organisations for failure to comply with the data breach notification statute.	16
Penalties limit	It defines a limit to the financial civil penalty imposed on an organisation found in violation of the statute	26

It is possible to classify those characteristics described in the previous section into four distinctive categories, which include elements that vary from state to state according to governmental decision:

1. ***Definitions and scope***: The category includes the definitions used for personal information and security breach, the coverage in terms of data

owned or maintained and in terms of individuals and organisations addressed by the law, and the notification triggers generated by data acquisition and in some cases also by data access.
2. ***Safe harbours and exemptions***: DBNLs provide different exemptions. The no-risk safe harbour applies if an entity's risk assessment concludes that there is no reasonable risk that a security breach has resulted or will result in harm to individuals whose sensitive personally identifiable information was subject to the security breach. Additionally, an organisation may be exempted from notification because of its technology (encryption). Finally, the exemption can be due to the implementation of an already existing company notification policy or to compliance with other laws such as the Gramm-Leach-Bliley Act or the HIPAA in the U.S..
3. ***Notification flow:*** Apart from individuals whose data has been acquired (or in some case accessed), a DBNL may mandate notification to one or more central authorities. In the U.S., these include the Attorney General and Credit Reporting Agency, and in Europe, it is a central authority to be appointed at the country level. Two additional components are also considered in this category: if the law specifies mandatory requirements for the content of the notice to be sent to residents, and the timing of notification. The former contributes to improved assessment of reputational effects in line with the amount and type of information to be provided.
4. ***Penalties:*** A limit for penalties may be established in terms of single violations or for a single breach. Under certain state breach notification laws, affected individuals may also bring action against a person or entity that violates the law to recover any actual damages suffered as a result of a violation of the law.[15]

DB notifications and their effects are shaped first by the resulting severity of the DBNL itself, according to the decisions taken in terms of the law elements mentioned above, but also by organisational factors in each sector addressed by the notification law. These two sets of factors are interrelated in many ways. The resulting incentive structure under which a company operates consists of a mix of contradictory forces, with some enhancing efforts to mitigate data breaches and others weakening them. For example, if a missed notification increases the risk of being audited and sanctioned, this constitutes a positive incentive, that is, an incentive

[15] 815 ILCS (Illinois Compiled Statutes)505/10(a) (2010)

to improve security and to mitigate data breaches. In contrast, the cost of acting to prevent data breaches is a negative incentive, as higher prevention costs discourage data breach mitigation efforts. The level of effort that companies exert on data breach mitigation depends on the combination and relative strength of positive and negative incentives.

At the level of a single organisation, relevant factors for data security include characteristics directly related to the company's human resources (such as the number of information security staff and the training performed on security aspects), to the channels used to offer products and services and therefore the accessibility of relevant databases (presence of eServices), to specific internal company elements (technologies and security procedures in place; access to information-sharing circles; managerial, technical and operational controls) and finally to the company's risk propensity (presence of specific insurance). These elements can determine the propensity for the firm to invest in the prevention of security breaches. They could be partially regulated by the information security policy, based upon the size and type of business and the type of information involved.

In this context, breach notification laws can significantly contribute to heightening awareness of the importance of information security across all organisational levels, as DBNLs may empower information security personnel to implement new access controls, auditing measures and encryption. Apart from the organisation's own efforts to comply with notification laws, reports of breaches from other organisations also help information officers to maintain this sense of awareness.[16]

The incentives that are perceived as relevant to an organisation when deciding its level of security and related investments are influenced by its business model, but also by the sector to which it belongs. Different behaviours are expected for commercial organisations (trade and retail), financial and insurance services companies, educational institutions, healthcare - medical providers or government agencies. Commercial organisations primarily respond to incentives that have direct and

[16] Security Breach Notification Laws: Views from Chief Security Officers, A Study Conducted for the Samuelson Law, Technology & Public Policy Clinic, University of California-Berkeley School of Law, December 2007.

indirect implications for their profit. Meanwhile, financial and insurance services companies follow the same rationale, but are more accustomed to strict regulations, as these sectors are already actively controlled. Therefore, sectoral elements are decisive, such as the legal and regulatory framework in which specific companies operate, the market structure and the associated competitive pressures and the conditions in related markets (e.g., for security technology).

Organisations belonging to different sectors issue notifications which may be initiated by a data breach due to different causes: among others unintended disclosure, hacking or malware, payment card fraud, insider and physical loss. The sector and the practices within that sector can experience different vulnerabilities to different causes, thereby requiring different security approaches. In the process of a notification issuing a central role is given to **law enforcement** and its effectiveness in surveilling to dissuade and discover illicit activity. Its control is dependent on the perceived risk of failure to comply, and the perceived risk requires a minimal level of the real risk. Both can be enhanced by the emergence of a threat, the degree of uncertainty and extensive, continuous publicity, preferably in the form of news coverage (Shinar and Mcknight, 1986) which is driven by **media**.

Notifications are addressed to individuals, who react to a received notification according to their behavioural preferences. The *Ponemon Institute 2012 Consumer Study on Data Breach Notification* offers a relevant overview of individuals' reactions to notifications, highlighting their opinions on privacy and the security of their personal information and their expected reactions to a data breach notification. Those opinions include to immediately discontinue the relationship with the organisation that had the breach, to consider discontinuing the relationship or to continue the relationship only as long as another breach does not occur. Consumer behaviours should also be studied with consideration to factors such as their computer and internet use, use of online services (eGov, eBanking, eCom), average education level, and propensity to sue.

In order to fully evaluate the consequences of DBNL implementation, the effects generated by a notification to individual members of the community should be taken into account. These effects can be divided into four main categories:

- ***Initiated Legal actions:*** In recent years, a large number of data breaches have resulted in lawsuits in which individuals seek redress for alleged harm resulting from an organisation losing or compromising their personal information. Data breaches have led to countless, intricate and expensive litigations, many in the form of punitive class actions. The remedies sought in these actions vary, but generally include costs for credit monitoring, costs for closing and opening financial accounts and damages for emotional distress.[17]
- ***Reported Crimes:*** After being notified of a breach of their personal information, consumers can then make informed decisions and activate appropriate actions to prevent or mitigate the impact of identity theft, one of the fastest-growing crimes today. They can report to credit reporting bureaus to flag the account with a fraud alert, banks or financial institutions and finally the police.
- ***Required Response:*** The response effects are largely generated consequently to the consumers' loss of confidence in firms that suffer breaches. Laws pressure firms to greater internalise the cost of a breach through notification letters, customer support call centres and mitigating actions such as marketing campaigns and free credit monitoring (Romanosky, Telang & Acquisti 2011). As companies calculate the consumers' responses in taking precautions, the incidences of data breach should be further reduced.
- ***Generated economic effects:*** Notifications can put customer loyalty at risk. The 2012 Ponemon Institute study found that in response to being notified by an organisation, 15% of respondents said that they would terminate their relationship and 39% said that they would consider ending the relationship. Thirty-five per cent reported that their relationship and loyalty is dependent upon the organisation not suffering another data breach. Apart from the customer loss, an additional direct economic effect is linked to the possible decrease of the firm's market value after a security breach generating data access or acquisition. In this respect, Garg, Curtis and Halper (2003) found that firms victimised

[17] Amburgy v. Express Scripts, Inc., 2009 wl 4067218, AT *1 (E.D. Mo. Nov. 23, 2009)

by a security breach involving theft of credit card information suffered a stock market loss of 9,3% on the first day the breach was announced, increasing to 14,9% over three days. Likewise, Campbell et al. (2003) found that there was no significant effect of breaches that did not involve data security, but that breaches associated with violations 'such as customer databases' did lead to significant losses in stock value.

These effects also produce feedback to the actors themselves, influencing the dynamics of the whole process in the medium-to-long term. Regulators, firms and citizens are likely to take proper countermeasures or at least behave differently in the face of increasing cases of identity theft or data breach litigations. The model elements are linked in multiple feedback loops and can become weaker or stronger through the role of enforcement and media, which co-evolve over time.

Media play an important role in driving the consequences of a notification and in influencing the possible effects. For example, tighter DBNLs may reduce data breach notifications, but can result in more vigorous efforts by affected individuals to pursue litigation, especially if media coverage on a specific data breach event makes them aware of compensation opportunities. News media more generally have the capacity to act as powerful influencers, influencing decision making in particular cases and influencing the system more generally through affecting the decision making of various participants in the system (Greene 1990, Hans & Dee 1991). In this context, media can act as an amplifier of a specific data breach event, pushing individuals to initiate litigation, influencing the effect on the firm's market value and increasing the effectiveness of enforcement, for example by highlighting the importance of investigations on related cases of identity theft.

Based on the elements described above, the model enables analysis at single or aggregate levels, pursuing the ultimate objective of clearly identifying the benefits of notification requirements, namely the reduction of associated costs given the cost of the notification itself (Lenard & Rubin 2005).

2.5 Current state of research on data breaches

In order to locate the issue represented in this dissertation in the broader context of the economics of the security field, we reference Figure 2.4 below (ENISA, 2012).[18]

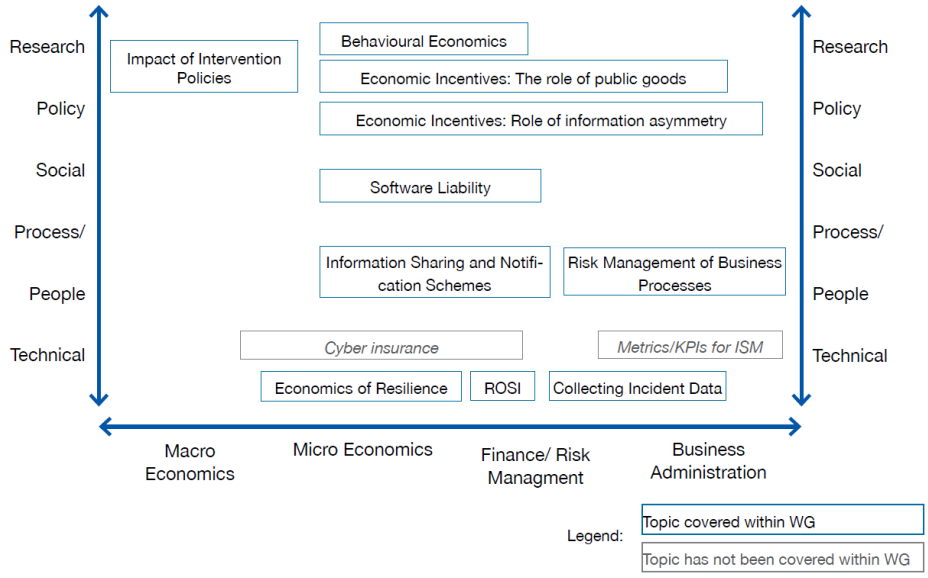


Figure 2.4 Overview of identified topics in economics of security

The horizontal axis illustrates various sub-areas of economics, from 'macro-economics' to 'business administration'. The closer a topic is to macro-economics, the stronger its relationship is to theoretical economic issues. The closer a topic is to 'business administration', the stronger its relationship is with specialised applicability (e.g., within business practices).

The vertical axis reveals the context of the various topics. The closer a topic is to policy, the higher its relevance is to political decision-making and political action. The closer a topic is to technical issues, the more it can be addressed by using existing technical solutions.

[18] This report is the outcome of a collective effort by all the participants (including me) in the working group Economics of Security that was established by ENISA with a view to study the priority topics relating to the Economics of Security.

A further dimension that becomes apparent in Figure 2.4 is applicability and scope: topics closer to the corner 'technical – business administration' are easier to implement and find solutions to, because their applicability is more specialised and the scope is reduced. In contrast, topics close to the other corner (i.e., 'micro-/macroeconomics – policy') will be more difficult for developing and applying solutions, as they have an extensive applicability and scope, and involve macro- and micro-economic relevance and national/international scope.

Data breach notification laws cover at least three of the highlighted boxes in Figure 2.4, namely collecting incident data, information sharing and notification schemes and the role of information asymmetry. DBNLs are typically justified with two objectives which are interwoven with these three elements.

The first objective is to preserve customers' right to know when their personal information has been stolen or compromised. As Schwartz and Janger (2007) described, informing customers allows them to protect themselves – for example, by changing their passwords or by monitoring their credit card statements for signs of abuse.

Prior research has found little evidence that this objective is being realised. By the time the consumer is informed, the attackers have had plenty of time to do damage. Romanosky, Telang and Acquisti (2011) suggested that the adoption of state-level data breach disclosure laws may have reduced identity theft resulting from these breaches by 6,1% on average.

A second objective is to create incentives for organisations to take adequate steps to secure the personal information they store. The reputational damage resulting from a reported breach activates 'the sunlight as disinfectant' principle, leading companies to invest more in cybersecurity and disinfect organisations of shoddy security practices (Ranger 2007).

Researchers have assessed reputational damage primarily through the effects of breaches on stock market prices. For example, Acquisti, Friedman and Telang (2006) investigated the impact of a privacy breach on stock market prices. They found a reduction of 0,6% on the day of the breach disclosure. Campbell et al. (2003) similarly identified a signifi-

cant and negative effect on stock price for data breaches caused by 'unauthorized access to confidential information'. Cavusoglu, Mishra and Raghunathan (2004) reported that the disclosure of a security breach results in the loss of 2,1% of the breached company's market value within two days of the announcement. Ko and Dorantes (2006) reported a mixed effect: although a breached firm's overall performance decreased (relative to firms that incurred no breach), their sales increased significantly. Sinanaj and Zafar (2016) presented another mixed result, finding that breaches had a negative and immediate impact on social media and corporate reputation, but no significant effect on stock market valuations. Kwon and Johnson (2015) used a propensity score matching technique to investigate how data breaches affect subsequent outpatient visits and admissions in the United States. They found that the cumulative effect of breach events (and also of the number of breached records) over three years significantly decreases the number of outpatient visits and admissions. This finding suggests that the effect of a data breach has a significant impact on subsequent consumer decisions.

Gordon, Loeb and Lucyshyn (2003) examined the incentives which foster investments in internal security. They found that expenditures to prevent information security breaches have been proliferating in recent years. The empirical evidence provided in their paper supports the argument that one key driver of actual expenditures on information security activities is the occurrence of security breaches. This finding is also confirmed by Moore, Dynes and Chang (2016), who found that most firms indicated that cybersecurity was becoming a major focus, either as a result of their own data breach experience or the experiences of other firms. The prevalence of breaches undoubtedly changed thinking in most firms' senior management about cyber-risk management.

The ideal final result of pursuing these two objectives is summarised in the Federal S.177 - Data Security and Breach Notification Act of 2015: 'to protect consumers by requiring reasonable security policies and procedures to protect data containing personal information, and to provide for nationwide notice in the event of a breach of security'. DBNLs also contribute to improving the security of the overall Internet ecosystem by increasing transparency for the security community, policymakers and citizens. In this respect, ENISA (2011) believes that the introduction of data

breach notification requirements is an important development with the potential to increase the level of data security and foster reassurance among citizens regarding how their personal data is being secured and protected. The introduction of DBNLs acts in an environment with two opposing trends: increased breach risk due, among other factors, to greater digitalisation and increased investment in security due to better awareness of the risk. Edwards, Hofmeyr and Forrest (2015) developed Bayesian generalised linear models applied to a public dataset to investigate trends in data breaches in the United States, demonstrating that neither the size nor frequency of data breaches has increased over the past decade. This finding may indicate that the competing forces offset each other.

In terms of the degree to which DBNLs increase the public visibility of data breaches, previous research has studied specific sectors (e.g., the medical sector investigated by Kwon and Johnson 2015) or state-level differences in the number of reported breaches (e.g., Faulkner 2007). However, as far as we know, no study to date has examined both simultaneously. This combined focus is important, as there are critical differences among the sectors in terms of the use of information technology and the presence of specific laws for managing personal data, as is the case for the finance and health sectors.

The European context for the law and economics of data breach notification obligations, which fall under the General Data Protection Regulation, was recently studied by Nieuwesteeg and Faure (2018). However, additional contributions to this area are essential to reach a better understanding of the phenomenon and for a better calibration of actions to be taken at the individual, company and institutional level.

2.6 Conclusions

From the potential for increased investment in security highlighted by the lack of information present in the market, to the current state of research, it is clear that there is a need to further investigate the many aspects within the proposed DBNL framework. This dissertation focuses on four aspects: communication styles in data breach notifications, the tim-

ing of breach identification and notification, hidden breaches and resulting identity theft. This scope enables us to understand the degree of achievement of DBNL goals.
As previously discussed, data breach notification laws are promulgated under the theory that customers have the right to know when their personal information has been stolen or compromised. In addition, data breach notification laws provide an incentive for organisations to take adequate steps to secure personal information they hold (sunlight as disinfectant).[19] The notification itself represents the core element of these laws. Issuing data breach notification letters is just one of the challenging tasks an organisation must accomplish after a leak of secure information to an untrusted environment has been discovered. A company that identifies a data breach faces a series of challenges in order to ensure compliance with the law.

Firstly, the company must identify customers whose data may be breached. This identification may be particularly difficult, as certain organisations, such as merchants, that have breached credit card numbers do not always themselves possess the mailing addresses associated with those numbers (GAO-07-737, 2007). Secondly, the company must manage compliance with multiple state laws. In fact, the applicability of the U.S. notification laws relates not to the residence of the breached organisation, but to the residence of the affected customers. This means that a company dealing with customers residing in different states must follow various state laws. Such laws differ in many elements, including who must be notified apart from the customer, the level of risk that triggers a notice, the nature of the notification and exceptions to the requirement. Therefore, a company must perform an analysis of all applicable state regulations in order to be sure that each resident's state law has been thoroughly followed in all of its provisions. Finally, the company must prepare and send the data breach notification letters to customers, supporting them in initiating the necessary countermeasures against possible consequences of the breach, such as identity theft. The following chapters address all three aspects of compliance, highlighting the lack and delay of reporting and related effects on identity theft.

[19] Phrase attributable to Justice Louis Brandeis, 1933.

Chapter 3 Let's not sugarcoat it: An investigation into communication styles used to notify breaches

On the basis of the presented framework, we analyse in this chapter the content of notifications sent in 2014 by breached organisations. We discuss how the notification of data breaches can be weakened or reinforced by means of the communication style. We define the core elements of consumer notification letters and identify how company decisions on what to include and the form of expression can distinguish specific letter types. Six typologies of notifications are identified.

3.1 Introduction

Today, data breaches are a complex phenomenon that must be handled with multifaceted competencies, not merely technical solutions. A company's identification of a breach that generates access or acquisition of personal customer information by third parties triggers a decision-making process, one crucial aspect of which is the communication to customers. This communication occurs through data breach notification letters, which are covered by the data breach notification laws enacted in the U.S..

The choice of the content of such letters provides an opportunity for companies to communicate the importance for the organisation of values not only to customers, but to all stakeholders. These values include security, compliance with the law and cooperation with law enforcement. Such communication, therefore, has an essential impact on the organisation's reputation. Moreover, if duly analysed, notification letters can support

identification of the organisation's risk propensity towards potential losses related to customer churn,[20] fines and class actions.[21]

While the discussion around a federal law on data breach notifications is ongoing and a series of large, costly data breaches have galvanised public interest in the issue, this chapter investigates the phenomenon of data breach notification letters and their content. The first section highlights the different regulations currently in place in the U.S., stressing which elements are mandatory by law in the various states. This overview enables us to clearly understand the starting point for organisations when drafting the notice according to their location and to the residency of their consumers and the related laws. The chapter then proposes an evaluation framework for DB notification according to a classification of letter types. This framework allows us to interpret better decisions taken by organisations when communicating a breach.

Through this process, we address the following research sub-question: '*What are the core elements of consumer notification letters and how do company decisions on what to include and how to express the message define specific letter types?*' This research question is answered while also taking into account the type of event that generated the breach.

This investigation is possible if we consider each notification letter as a set of elements that can be isolated and analysed. Each of these elements poses the communication organisation with a dilemma of how to inform consumers about a breach. This research can be useful to companies in order to inform more conscious decisions when choosing among the options at stake. It can also be helpful to policymakers through contributing to the ongoing discussion on federal law on data breach notifications, highlighting the limitations and effects of the current state laws.

[20] Phenomenon where customers of a business no longer purchase or interact with the business.

[21] A legal action that is organized by a group of people who all have the same legal problem.

The primary sources of information used for this investigation are (1) the data breach notification laws of 47[22] states and selected extensive reports issued by law firms and available online,[23] which identify mandatory elements of the notification letters; (2) 445 data breach notification letters sent in 2014, downloaded from the Attorney General websites of four different states, used to establish the evaluation framework, to identify the different dilemmas and to verify the choices made by the affected companies; and (3) the Ponemon study,[24] used to examine the letters against consumers' perceptions about the importance of receiving a notification when their sensitive personal information has been lost or stolen.

3.2 The legislative framework

Notifications are issued in the 47 U.S. states that have enacted data breach notification laws requiring businesses and other entities to notify affected individuals when a data breach involving their personally identifiable information (also referred to as PII or personal information) occurs. As previously noted, the first U.S. DBNL was enacted in California. Other U.S. states may diverge from the Californian model according to local decisions taken in regard to different legislative elements; however, the implementation of DBNLs is always seen as a potential remedy to address the multifaceted problems of personal information protection,

[22] As of 31.12.2016 Alabama, New Mexico and South Dakota were the only U.S. states that had not yet enacted a data breach notification law. Since 2018 all states have a DBNL in place.

[23] Data Breach Notification Laws by State' (CLLA, 2012) http://www.clla.org/documents/breach.xls
State Data Security Breach Notification Laws' (Mintz Levin, 2012)
http://www.mintz.com/newsletter/2007/PrivSec-DataBreachLaws-02-07/state_data_breach_matrix.pdf
State Data Breach Stature Form' (Baker Hostetler, 2013)
http://www.bakerlaw.com/files/Uploads/Documents/Data%20Breach%20documents/State_Data_Breach_Statute_Form.pdf
Security Breach Notification Chart (Perkins, 2013)
http://www.perkinscoie.com/files/upload/LIT_09_07_SecurityBreachExhibits2.pdf

[24] 2012 Consumer Study on Data Breach Notification, Ponemon Institute LLC, June 2012.

inadequate corporate information security measures and the rapid increase of identity theft crimes (Faulkner, 2007).

As noted in the previous chapter, the requirements of the laws in other states differ from the California model and also vary from one state to another. While most state notification statutes have similar elements, there are substantial differences, such as in the definition of personal identifiable information, in the notification trigger, in the penalties for non-compliance and in the notification flow. In many cases, a one-size-fits-all approach to notifications will not suffice, particularly given that even single states amend their laws over time and the landscape continues to shift. As such, companies that do business in multiple jurisdictions are at significant risk of failing to comply with one or more state notification statutes should a breach occur. The complexity confronted by organisations dealing with customers in multiple states is significant. Unfortunately, there is no single form letter that guarantees compliance with all laws, and most state breach notification laws do not provide specific requirements for the notice's content.[25]

However, an assessment can be performed based on the state breach notification statutes that do establish minimum requirements for the letter content. Fifteen state statutes determine such minimum requirements. The required elements vary from the type of personal information subject to unauthorised access or acquisition to the entity's name and contact information so that affected individuals can obtain additional information. A general description of the breach incident may also be required. Other required elements mandated by a limited number of states include general advice on actions that affected individuals should take, a statement indicating that individuals can obtain information from specific sources such as the Federal Trade Commission (FTC) and consumer reporting agencies and a reminder notice of the need to remain vigilant for incidents of fraud and identity theft.

Predefined letter elements, if effective, should make the public notices useful and easy to understand, and should contribute to mitigating the risks of unauthorised and uncontrolled access to personal information.

[25] Some organizations opt for filling the gap with an annex which fulfils case by case each state legislation.

The fact that many state statutes do not mandate minimum information to be included in the letter is counterproductive to mitigating the risks of breaches, as it increases the number of consumers who receive a notification letter and find it not easy to understand (52% according to Ponemon study), and generates potential confusion with other mail solicitations that may resemble notification letters.[26]

In the few cases where content is specified by law, some of the mandatory elements cannot be modulated, as they are objective details such as the date or contact information. However, the majority of the components can be manipulated, which has a direct impact on the resulting message tone – whether alarming or reassuring, clarifying or confusing – about the event and its consequences.

These laws operate at the intersection of business communication and information security (Veltsos 2012). We investigate the laws through the use of an ad hoc evaluation framework. We observe if and how companies leverage on consumer inaction resulting from their behavioural decision biases. Consumers can perceive the risk of suffering identity theft to be very low (optimism bias), or they can believe that the benefits of taking precautions outweighs the benefits they may obtain (rational ignorance). Finally, consumers may follow their inertia, inhibiting them from anticipating the consequences of identity theft and responding (status quo bias) (Loewenstein, John & Volpp 2012).

3.3 Constructing the Notification Letter evaluation framework through descriptive analysis

We previously discussed how data breach notification laws require that organisations contact customers after the discovery of a breach affecting PII, but offer poor indications of the style and content of the notification. The following sections now investigate how companies use this leeway in delivering bad news related to a breach.

[26] For example, officials at one large national bank noted that marketing solicitations for credit monitoring services often are made to resemble breach notification letters, potentially desensitizing or confusing consumers when a true notification letter arrives. (GAO-07-737, 2007)

To perform this investigation, we constructed an evaluation framework starting with a review of the existing research in the field of risk communication and crisis communication. These two fields are respectively associated with (1) environmental management, public health and emergency management, and (2) crisis management and public relations (Williams and Olaniran 1998). We also looked at the expertise of business books on the communication of negative messages. We enrich the review by posing concrete examples of each of the topics highlighted in the relevant literature. The examples comprise sentences and paragraphs extracted from the sample of letters which was analysed for the construction of this evaluation framework. The sample includes authentic notification letters made public by the Attorneys General of four states.[27]

Risk communication aims to inform people about a potential future harm and the associated dangers so that they can act to lessen the risk (Seeger, Sellnow & Ulmer 2003, Seeger 2006). The goals of risk communication can include building trust in the communicator, raising awareness, educating, reaching agreement and motivating action (Rowan 1991). In consideration of these different aims, it is decisive for organisations to clearly define their objectives when approaching risk communication. The same is valid for companies' data breach notifications and the need for clarity around the goals they decide to pursue.

Risk communication aims at preventing harm, while crisis communication at communicating during an event. Event centred and incident-specific communication focuses on the message and how it is delivered during the event, with an emphasis on the need to distribute accurate, timely, and useful information (Seeger 2006). The message often updates on the current state of affairs or conditions, what is known or not known and the status of the message deliverer (Seeger et al. 2003). In this way, data breach notification messages may not fall under crisis communication if the crisis is defined as the breach that took place. However, if we consider the crisis as the situation generated by the consequences of the breach within and outside the organisation, data breach notifications fall under this field. Indeed, several authors have recently started enlarging the scope of crises beyond the actual event to include pre- and post-crisis phases (Reynolds and Seeger 2005, Heath et al. 2009). As such, crisis

[27] California, Maryland, New Hampshire and Vermont.

communication research has begun to overlap with the risk communication literature (Steelman and McCaffrey 2013) and is consequently suitable to support the study of data breach notification.

The objective of communication during a crisis is to influence the public's perception of the associated organisation and to maintain a positive image or restore a damaged image among stakeholders (Ray 1999). Coombs (1995) developed a five-category model of message strategies used in response to crises. The model is articulated as follows:

- ***Nonexistence strategies***. Nonexistence strategies seek to eliminate the crisis by denying its existence, explaining why there is no crisis, or even attacking those who wrongly report.
- ***Distance strategies***. Distance strategies aim at creating public acceptance of the crisis and at weakening the linkage between the crisis and the organisation. Consequently, they make excuses or justify the crisis.
- ***Ingratiation strategies***. Ingratiation strategies aim at gaining public approval for the organisation. It focuses on reinforcing the organisation image by reminding of its existing positive aspects.
- ***Mortification strategies***. Mortification strategies attempt to win forgiveness and create acceptance for the crisis. Such strategies include remediation to offer some sort of compensation to the victims, repentance to ask for forgiveness and rectification to take action to prevent a similar crisis from occurring again.
- ***Suffering strategy***. The goal of a suffering strategy is to portray the organisation as a victim and draw sympathy from the public.

In addition to these strategies, Coombs (1999) later defined the ***silence strategy***. In essence, it suggests uncertainty on the part of the organisation in crisis, and relies on the 'endorsement of an outside expert' (p. 132) to help increase the credibility of the organisation.

Stephens, Malone and Bailey (2005) found that in a sample of 10 different cases representing six types of crises, the top strategy chosen was mortification. Section 3.6 later analyses how data breach notification types can be linked to Coombs's classification.

Huang and Su (2009) provided three favourable ways in which messages can be communicated and responses formed during a crisis event: activeness, consistency and timeliness. These three elements can also apply to data breach notifications.

An active response concerns taking the initiative in a crisis, such as a security breach of consumers' personal information, to actively issue responses. This action may portray the organisation as honest and forthright. Conversely, the lack of an active response may make the organisation appear unresponsive and eager to conceal information. All notifying organisations satisfy this requisite, whereas those aware of the breach that do not issue notifications would classify as not active.

Meanwhile, a consistent response means communicating messages uniformly to establish legitimacy. This response enhances accountability and credibility by offering clarity about the facts and corporate responsibility. If we apply this concept to data breach notifications, a consistent response will result in a clear description of the event causing the breach, highlighting the possible lack of security in the organisational processes.

Finally, a timely response refers to releasing information at the appropriate time. A lack of timely response during a crisis can sour an organisation's relationship with its stakeholders. In cases of data breaches, a timely response would mean notification in due time to enable consumers to take the necessary countermeasures to protect themselves.[28]

In addition to the literature on risk and crisis communication, a rich source of information is represented by business communication textbooks. However, such textbooks present a limitation. They primarily provide advice for low-risk and routine situations, such as denial of credit, collection requests, rejections for employment, inability to meet deadlines and similar occurrences that have occupied attention in business communication classrooms since the 1930s (DeKay 2012). Even if growing in number, data breach notifications should be seen as high-risk and non-routine situations: 'specific unexpected and non-routine events or series of events that create a high level of uncertainty and threaten an

[28] We report on this element in the next chapter.

organisation's high priority goals' (Seeger, Sellnow & Ulmer 1998, p. 233).

In the field of bad-news, research inquiry and points of contention primarily centre on three key aspects of composing and disseminating negative news messages: (a) arrangements (b) components, and (c) pedagogical techniques (Creelman 2012). We focus on (b) components for our evaluation framework.[29]

3.4 Taxonomy and proposed variations to evaluate letter components

Many researchers have questioned the use and effectiveness of the conventional components of bad-news messages prescribed by business communication textbooks as a means of presenting adverse events. There are generally three primary components to bad-news messages. Firstly, textbook authors mostly agree that explanation is a central aspect of negative messages. An explanation should describe the problem clearly and unemotionally while not placing blame (Carter 2012), as well as protect the organisation's reputation in order to reduce the need for follow-up correspondence (Bovée & Thill 2012). In the analysed breach notifications, we can identify the explanation component in both the incident description and the reaction of the organisation.

Secondly, the bad news itself is a component which contains information resulting in a perceived loss by the receiver and creates cognitive, emotional or behavioural deficits in the receiver after receiving the news (Bies 2013). In cases of data breach notifications, the bad news is that PII has been accessed/acquired, and this access/acquisition may generate possible negative consequences. When possible, bad news is followed by an alternative solution or action, in line with traditional advice in the bad-news literature to 'offer an alternative or a compromise if one exists' (Locker 1999, p.31). In the analysed notifications, we can identify the alternative element in the suggestions for customers to be vigilant, check credit reports, file a complaint with the FTC or activate security freezes.

[29] We discuss arrangements in the next chapter.

Thirdly, the components of a conventional bad-news message also include prefatory and closing buffers that provide background information, good news, thanks and compliments or generally accepted truths or that express empathy with the audience (Shwom & Snyder 2012). In the investigated notifications, buffers are mostly represented by statements on the importance of security within the organisation and on the reassurance of an enhanced level of protection. Closing buffers usually offer support for clarifications by providing company contact information.

After a careful analysis of the notifications, we propose a new approach to classify and evaluate the decisions taken by companies when composing a data breach notification. The three 'conventional' components are embedded in the proposed framework, which describes the dilemmas faced by organisations when composing data breach notification letters. Given their frequency in the letters, six elements, in particular, are worth an isolated analysis:

1. **Clarity** of the incident description and of the PII involved. (Explanation and bad news)
2. **Communication tone** around the possible consequences considering the organisation reaction to the breach (Explanation and bad news)
3. **Approach to actions** to be taken by the affected customers (Alternative)
4. **Interaction** with affected customers (Closing buffer)
5. **Stated relevance of security** to the organisation and **stated steps to reinforce security** (Prefatory and closing buffers)
6. **Style** in addressing customers

The following paragraphs describe each element, provide related extracts from the collected letters, offer comments on the different styles, and finally tie these observations to the results recorded in the Ponemon study. Conducted by the Ponemon Institute and sponsored by Experian, this study aimed to understand consumers' perceptions about the value of receiving notification when their sensitive personal information has been lost or stolen. The study surveyed 2.832 consumers 18 years and

older.[30] The following were the key topics addressed: consumers' reactions to data breach notifications; the importance of notification following a data breach; information considered essential to include in a data breach notification; consumers' expectations when receiving a data breach notification; recommendations on how to improve communication following a data breach; organisations' efforts to help consumers and steps consumers take to protect themselves following a notification.

1) Clarity of the incident description and of breached PII involved (opaque vs. transparent). The decision of how detailed the event description should be and if to acknowledge the organisational or procedural weaknesses of the company depends on the management's evaluation of the legal framework, customer relationships and potential additional harm for the affected customers and/or the company. Organisations may withhold information out of fear or to save face. While this may be a natural reaction, withholding information can cause misdiagnosis of the actual problem or an underestimation of its extent. Moreover, when the hidden facts become public, organisations are viewed in a worse light than if all the facts had initially been disclosed (Bies 2012). This notion is confirmed by the 2012 Ponemon study on data breach notifications. Customers in the study said that they were dissatisfied with the communication and often felt the need for more information. In particular, 61% of customers believed notifications were not easy to understand (mostly because of an overly long and poorly written letter and too much legal language). Many customers did not believe that notifications increased their understanding of the event. In particular, 37% of the customers said that they had no idea what the data breach was about. Besides, 45% of the customers suggested disclosing all the facts in order to improve communication of the breach.

In order to determine the level of clarity, there are three possible options for transparency in the event description: transparent, transparent with no dates and opaque. A notification is classified as transparent when it meets at least two out of the following three requirements: the type of

[30] However, only 25 percent of these consumers (708) were able to recall if they received a data breach notification and could answer survey questions about their experience.

event is specified, the generating causes are described, and the organisation reaction is indicated. It is classified as opaque if it meets only one of the requirements listed above. Full transparency includes the presence of the breach discovery date and breach date, transparent with no dates refers to cases in which neither of the dates is indicated. Below are the texts of three data breach notifications from the sample which highlight the three possible levels of clarity in representing the data breach generating event (i.e., opaque, transparent, transparent no dates).

The letter sent by Experian on July 21st reporting unauthorised access of consumer information reflects an **opaque** description of the event:

This letter is to inform you that your personal information may have been accessed without proper authorization. This unauthorized access took place sometime between April 15, 2014 and June 27, 2014.
Experian, one of the nationwide credit reporting agencies, identified that its client, NRG Assets LLC, had certain Experian consumer information accessed without proper authorization. The consumer information consists of information typically found in a consumer report. Such information includes your name and address and one or more of the following: Social Security number, date of birth, or account number. Experian is actively working with NRG Assets LLC to investigate this matter. [238][31]

A **transparent** approach was used by SIMMS in their letter dated November 25, 2014:

I am writing to inform you of an incident discovered November 6, 2014, involving the theft of personal information from our online store. An unknown criminal installed malware in our online check out system that appears to have intercepted customer purchase information for purchases between September 1 and November 6, 2014. Your name, address, and credit card information, including the credit card number, expiration date, and CVV2 code (Card Verification Value on the back of the card), may have been among the information accessed.
Our website hosting and support vendor has taken the necessary steps to remove the malware and prevent it from being reinstalled. We have reported the incident to and are cooperating with law enforcement. We have also informed the credit reporting agencies and payment card

[31] In brackets the number indicating the letter as per list in reference.

networks about this incident so that they may take appropriate action regarding your credit card account. [398]

Finally, it is possible to be transparent while not specifying the relevant dates (discovery of the breach and start date of potential harm), as Ameriprise Financial did in September 2014:

I am writing to make you aware of an incident that occurred involving your personal information. Recently, my office was broken into and the building set on fire. Many client files were damaged due to smoke and water, and the room where kept client files was accessed. It is not known if your information was taken, but your client file would contain your name, address, date of birth, Social Security and account numbers. Due to the sensitive nature of the information, I wanted to notify you of this incident.
We have taken steps to protect your accounts from unauthorized activity, which includes instructing our services associates to use extra caution when verifying caller and to confirm the signature on written requests related to you accounts. [304]

In order to reduce the analytical complexity, we did not to take into consideration the PII input, analysing only the event description through the details provided in the event explanation. If the description did not indicate the type of the event that generated the breach or of the circumstances related to the presumed cause of the event, we classified the clarity as opaque.

The analysis reveals that most of the organisations decided to describe the events in a very transparent manner. However, it is worth noting that none of the analysed letters provided the number of breached records, which would directly reveal the extent of the breach and therefore the extent of the company's failure in ensuring data security.

***2)* Communication tone in depicting the possible consequences of the data breach (reassuring/neutral/alarming).** Organisations are torn between a range of possibilities when deciding the tone of the notification message. Downplaying the effects of the data breach may mollify readers' anxiety, but may also discourage them from taking action to protect themselves (Veltsos 2012). Some organisations tend to adopt a reassuring tone about the consequences of the data breach in order to mitigate the short-term reputational effects on customers, particularly on

those who ignore the existence of the data breach regulation in place. With this choice, the risk of legal action can be higher if the data breach results in tangible severe consequences, such as identity theft. The reassuring communication tone is established by expressions that stress the absence of actual harm for customers, for example, through phrases such as we have no reason *to believe, we have no indication* or *we have no evidence*. The objective of this tone in almost all cases is to emphasise that there is no current damage and to belittle the potential for future harm. The letter sent by Thomson Reuter on July 7th regarding a security incident involving the misuse of credit card information by an independent contractor reflects a **reassuring tone:**

Although we have no reason to believe that your personal information was misused by this independent contractor or that any fraudulent activity occurred on your credit card account, your EndNote order was one that this temporary contract processed. Nevertheless, as a precautionary measure, we have arranged to have AllClear ID, an identity theft and credit monitoring company, help protect your identity for 12 months at no cost to you. AllClear maintains an A+ rating at the Better Business Bureau. [215]

The opposite tone is to **alarm** the customers in order to provoke them to take all necessary steps to avoid additional negative consequences. The customer will bear part of the cost of the mitigation, but will perceive the company as trustworthy. In the study conducted by the Ponemon Institute, 56% of customers suggested improving notifications by explaining the risk of harm that will most likely be experienced as a result of the breach. One example of such an approach is the letter sent by UPS on August 20th informing customers of malware intrusion:

Based on the investigation, we feel it is critical to notify our customers of the potential data compromise. [279]

Other organisations adopt a ***neutral*** tone, stressing the uncertainty of current damage (*'we are uncertain', 'we do not know')* while explaining the steps to mitigate any potential consequences. The notification sent to consumers on September 5th by Cedar-Sinai in response to a data breach of health information reflects this tone.

Cedar-Sinai is unaware of any attempted or actual unauthorized access to or misuse of your health information, but has provided information in this letter on additional steps you can take to protect your identity should you feel it appropriate to do so [305]

The decision on tone is of course dependent on the event itself, but also on the relevant legal framework. In states without mandatory content for notifications, companies can more easily opt for a reassuring tone instead of alarming customers about the event compared to fully regulated states. The choice of a reassuring tone can be a consequence of the greater leeway when deciding which elements to include in the notification. For example, California regulation does not make it possible to belittle the event given that almost all of the elements must be included in the letter.[32]

3) Approach to actions to be taken by the affected customers (neutral vs. encouraging). Another decision node for the organisation is the choice between listing all of the possible actions a customer could perform or taking a position and recommending selected actions to customers. In the latter case, the letter can act as an alarm bell for customers, contributing to making them take the content of the message seriously. The actions that are usually suggested are to report to credit reporting agencies that one may have been a victim of an identity theft, to ask the credit reporting agencies to put a fraud alert on the credit file (or, more rarely, to put a credit freeze on the credit file), to check credit activity regularly with each credit issuer or to activate a credit monitoring service at no cost to the individual. In some cases, the notification also specifies why the organisation is not performing those actions itself (e.g., '*credit agencies will not permit our firm to act on your behalf regarding your credit data'*).

When adopting a **neutral** tone, messages highlight that the company is not in a position (or does not want) to advise on what to do, or clearly encourage customers to evaluate the situation themselves. Allianz used this approach in December 2014:

[32] See Figure 4.2 for more details.

At this time, we have no reason to believe that your personal information has been or will be misused. However, for your own peace of mind, you may wish to monitor your financial accounts, such as banking, brokerage and insurance statements, for any unusual activity. [439]

The opposite approach is to **encourage** the customer to act to reduce risks with determined expressions such as '*we would like to urge you to…', 'we believe you should…' or 'we encourage you to…*' Such expressions were used by Home Depot after its data breach suffered in May 2014:

We encourage you to review your account to check for any transactions that might reflect improper use of your information. You should immediately report any indication of inappropriate use of your information to your credit card company. Even if you do not see signs of misuse, to be cautious you may want to ask your credit card company to cancel your current card and issue you a new one. [167]

4) Interaction with affected customers (neutral/available/fostering). Activating and managing communication channels increases company costs, through the costs of support services such as call centres and for the higher rate of activated credit monitoring. However, fostering such contact may limit reputational effects, as it demonstrates a strong willingness to cooperate to avoid negative consequences. While almost all notification letters provided the contact information of the breached companies in order to provide additional information or help, the style used in offering this opportunity differs from case to case.

In classifying the tone of the notifications around interaction, we used the following requirements. In essence, a fostering tone refers to a strong invitation for action, supported with expressions such as '*we are eager to help'* or with contact details in bold letters. An availability tone was identified with a standard sentence such as '*please do not hesitate to contact us'.* Finally, neutral interaction describes cases in which no contact number was explicitly provided. Below are examples of a fostered interaction, availability and neutral communication of a contact number.

State Industrial Product Corp. **fostered** interaction in their communication sent on January 27th through the use of capital letters:

We take this matter very seriously. We set up a dedicated call center if you have any questions, or you need further assistance. ***Please call the dedicated (not the HR department) at 1.877.218.2561 and enter this reference number: 2702012514.*** *The call center will be open Monday through Friday, 9:00 AM until 7:00 PM, Eastern Time [32]*

Catamaran highlighted **availability** for interaction in their communication dated February 7th, 2014:

If you notice activity that may be of concern, or if you have any questions or need additional information, please do not hesitate to contact us toll-free at 855-577-6522, 24 hours per day, seven days per week. [47]

Finally, Tinyprints decided to be **neutral** toward interaction in their data breach notification sent in November 2014:

For more information and updates, please go to http://www.tinyprints.com/security.htm by typing this address into your browser. [380]

5) Stated relevance of security for the affected organisation and stated steps to reinforce security[33]

Highlighting the relevance of security for the organisation can be reassuring for the customer. However, it can also generate the thought that even though security is a top priority for the organisation, it has failed in protecting key information. Moreover, pompous statements on the high level of security in the organisation can also be perceived as intending to minimise the event. According to the Ponemon study, 28% of customers who received notification letters in the past suggested that organisations not 'sugar coat' the message in order to improve such communication.

Messages about the importance of security to the organisation and its steps taken are typically either included in the letter introduction or at

[33] In particular, stated actions taken or planned to contain the breach and protect data from further unauthorized access or use.

the conclusion, and they often refer to data protection, data confidentiality and security and privacy as critical priorities in the organisation (see the examples below).

Protecting the confidentiality of this information – and all of our clients' information – has long been a top priority for us. However, ...

The confidentiality and security of our business partners' and former and current customers' personal information is very important to us. We maintain physical, electronic and procedural safeguards that meet state and federal regulations and we limit access to our customers' information.

Your security and privacy are very important to us.

We pride ourselves on creating a positive environment for all of our customers. We wanted to be proactive in bringing a recent incident at our Sacrament division office to your attention and we hope to address any concern you may have.

This last example demonstrates how legal compliance can be communicated as proactivity. The sentence once again proves how companies can make use of customers' lack of information regarding the legal framework in place, which enables them to present a particular action as proactive, when in most cases it is legally mandatory.

As for the actions taken by the businesses to contain the breach and protect data from further unauthorised access or use, more than 50% of the organisations in the sample stated that additional steps had been taken in order to reinforce security and prevent similar events. This is a critical point considering that 35% of the Ponemon study respondents reported that their relationship and loyalty is dependent upon the organisation not having another data breach.

We have implemented additional measures that will help prevent a similar occurrence.

We are taking immediate steps to minimize the likelihood of similar events in the future, including a top-to-bottom review of the company's information security policies, limiting the amount of personally identifiable information stored on devices, and increasing the use of encryption and other protective technologies.

In addition to terminating the unauthorized access, we revalidated our information security infrastructure to confirm that we maintain industry standard protections for customer data.

We have implemented additional control to avoid a similar future incident. These controls include enhanced security measures which limit use to select authorized personnel.

Finally, it is interesting to note that some organisations anticipate the risk of an additional notification related to a new data breach, using expressions such as the one below.

We have also taken additional proactive security measure to help prevent a similar incident from occurring in the future; however due to the nature of cybersecurity attacks, it is virtually impossible to entirely prevent these types of event from ever occurring.

6) ***Style in addressing customers (form/personal).*** Communication style also plays a vital role in influencing customers' perceptions in terms of the seriousness of the news received. For example, maintaining a cold profile through not addressing the customer by name and surname can be an option if the organisation's strategy is to not alarm the customer or to encourage the customer not to take the letter seriously and perhaps confuse it with junk mail. However, if negative consequences ensue and the customer can link the consequences to the data breach, greater negative impacts can be expected for an organisation which adopts this approach.

The personal addressing approach always uses the following style:

Dear ‹‹Title›› ‹‹Last Name››, or
Dear ‹‹First Name›› ‹‹Last Name››

In contrast, the form option uses expressions as follows:

Dear Applicant,
Dear Cardholder,

In other cases, there is no salutation at all.

The Ponemon study found that 62% of notifications used a form letter, while only 19% used a personal letter.[34] The choice of a form letter contributed to generate the perception that the notifications were junk mail or spam (49% of the respondents), and discouraged identification of the notifications as important communications (34%). Furthermore, the same indicators measured in 2005 were 23% and 51%, respectively, revealing a definite growing trend in misunderstanding the real goal of the breach notifications.

In order to limit reputational effects, organisations may also apply solutions often used in cases of product complaints, such as coupons or inexpensive 'goodies'. Such compensation may advance the organisation's symbolic goals, such as demonstrating the importance of the customer to the company and the sincerity of the remorse (Conlon and Murray, 1996).

As a token of your appreciation for your continued patronage, we are also enclosing a 20% discount code that you may use on your next purchase from us at www... [7FL]

For a limited time, we are offering a Preferred Customer Rate discount program for our customers who may have been impacted by this incident. You will receive a 20% discount... [10FL]

3.5 Implementing the framework and defining letter types

Based on the identified content and characteristics of the letters, we created a database to code each letter characteristic, both at the paragraph level to understand the order of the letter contents (arrangements), and

[34] The remaining cases (19%) refer to other options to communicate the breach, including telephone call and Posting in major newspaper.

at the sentence level to identify the content and purpose (use of apologies and developed components). Two coders performed this task. For any variable for which no agreement could be reached, random selection was used.[35] The database provides information on the following elements for each notification in the sample:

Reference Variable

Type of Event: Definition of the event according to privacyrights.org, which classifies events that generate notifications as follows: unintended disclosure (sensitive information posted publicly on a website, mishandled or sent to the wrong party via email, fax or mail), physical loss (lost, discarded or stolen non-electronic records or portable or stationary device), insider breach (someone with legitimate access intentionally breaches information, such as an employee or contractor), hacking and malware (electronic entry by an outside party; malware and spyware), payment card fraud (fraud involving debit and credit cards that is not accomplished via hacking), unknown or other.

Traditional Elements[36]

1) ***Use of apology:*** Choice among apologies, regrets or neither.
2) ***Arrangement***: Choice between direct and indirect patterns, also indicating the use of buffers.

Developed Components: Identification of each of the proposed components for evaluation:

a) ***Clarity of the incident description and of the PII involved***: **transparent**, **opaque**, or **transparent no dates** approach regarding the description of the facts and the accessed PII. **'**Transparent no dates' in case of transparent incident description but no specification of the date of the incident and the date of the breach discovery.
b) ***Communication tone regarding the possible consequences:*** **alarming**, **neutral** or **reassuring** based on the sentences coding.

[35] Results of interceding reliability varies from 96,24% to 100% (percentage of agreement) and from 74,43% to 100% (Scott's Pi).

[36] Useful for the analysis of next chapter.

c) ***Approach to actions to be taken by the affected customers: encouraging*** or not (**neutral**) of customers' actions to minimise their own harm, and subsequently the company's harm.
d) ***Interaction with affected customer:*** encouraging contact with a contact person in the breached organisation (**fostering**), presenting availability for contact (**available**) or **neutral.**

The above four elements define the prerequisites of the letter typologies, and their various combinations by letter type are illustrated below. Information about further characteristics was also collected to offer a clearer picture.

e) ***Stated relevance of security for the affected organisation***: mention of the importance of security for the organisation
f) ***Stated steps to reinforce security***: mention of the steps taken or planned to reinforce security and to avoid future breaches.
g) ***Style of addressing:*** use of the name and surname for a **personal** letter or initiating the notification with a general 'dear customer' or no salutation at all for a **form** letter.

The descriptive statistics from the data analysis are presented in the following section. The analysis aimed to identify relevant patterns in the notification sent given the type of event generating the breach.

3.6 Descriptive Statistics

A total of 213 notifications for the first semester of 2014 were analysed across all framework elements. In essence, each letter was classified in terms of the type of event, use of apologies, arrangement and options for components. As described above, the single notification elements were recorded using an inductive content analysis.

Type of event

Figure 3.1 illustrates the distribution of notifications based on the types of events that generated the data breach. As expected, hacking and malware is ranked first. The second type of event, unintended disclosure, accounted for ¼ of the total data breaches. Insiders and physical loss came in third and fourth place respectively, with the same frequency. Finally, payment card fraud not accomplished via hacking represents 2% of the sample. We use this variable to understand better the dynamics of the elements and components of the data breach notifications. As the type of event varies widely, the variable allows for identifying possible patterns in the companies' available choices when drafting notifications.

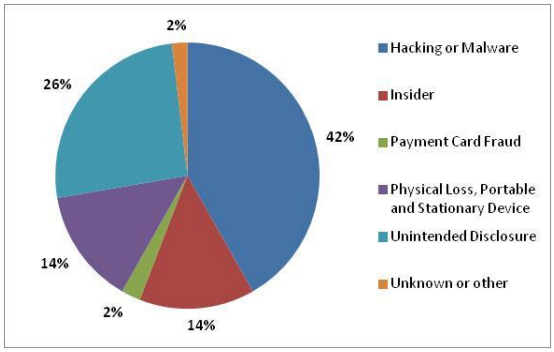


Figure 3.1 Data sample by type of event

Developed components

Table 3.1 describes how the notification components are represented in the sample. In most cases, the letters were transparent in describing data breach events (78,40%). Meanwhile, 14,08% of notifications transparently described the event, but omitted the date of the breach detection and indication of when the breach could have taken place or started. A neutral tone about the possible consequences of the breach was also used in the majority of cases (60,56%), while one letter out of four adopted a reassuring tone. Only 37,09% of the letters encouraged the readers to take action to mitigate risks. In addition, most organisations demonstrated availability towards customers in terms of supporting them in the post-event process (85,45%), while few fostered interaction (7,04%).

Characteristics	Options	Number of letters	%
Clarity of the incident description	Transparent	167	78,40%
	Transparent no dates	30	14,08%
	Opaque	16	7,51%
Communication tone on the possible consequences	Alarming	27	12,68%
	Neutral	129	60,56%
	Reassuring	57	26,76%
Approach to actions to be taken by the affected customers	Encouraging action	79	37,09%
	Neutral	134	62,91%
Interaction with affected customers	Fostering	15	7,04%
	Available	182	85,45%
	Neutral	16	7,51%

Table 3.1 Data breach notification characteristics – main components

Table 3.2 indicates the proportions of additional elements recorded. As listed in the table, organisations in the sample often emphasised both the importance of security (61,50%) and the steps taken to reinforce security after the breach (60,56%). In most cases, the letters address the individuals by name and surname (73,71%), not using a generic 'dear customer' or similar salutation. Moreover, in the vast majority of cases, the PII accessed or acquired was clearly specified (93,30%).

Characteristics	Options	Number of letters	%
Clarity of the PII involved	*Transparent*	*200*	*93,90%*
	Opaque	*13*	*6,10%*
Stated relevance of security	*Yes*	131	61,50%
	No	82	38,50%
Stated steps taken/planned to reinforce security	*Yes*	129	60,56%
	No	84	39,44%
Style in addressing consumers	*Personal*	157	73,71%
	Form	56	26,29%

Table 3.2 Data breach notification characteristics – additional components

The combination of the various letter elements defines the ultimate type of communication. We identified the **clarity** of the event, the **tone** regarding the consequences, the **action** suggested to the reader and the **interaction** fostered by the writer as drivers for identifying the overall letter type. Analysis of the letters in the sample resulted in the following six letter types, which cover almost 94% of the sample.

1. Cold: This style is detached, explaining the facts coldly and transparently. It remains neutral in all elements of the message, in particular when describing the consequences of the breach and the actions that might be initiated by the recipient of the letter. Within the sample, 27,70% of the letters belong to this type. As this group of letters refers to companies which did not take a strong position when communicating the data breach and did not actively foster contact with customers, this letter type cannot easily be linked to any of Coombs' (1995) strategies.

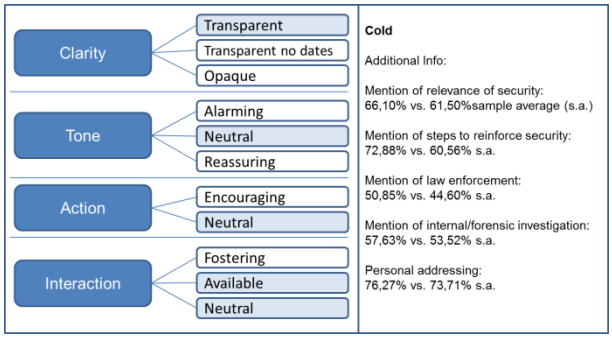


Figure 3.2 Cold letter

2. Routine: Companies electing this type present the event as a consequence of an unavoidable and relatively common risk. The company emphasises its actions, describing how all necessary steps were duly performed after the event. The consequences are represented with a neutral or reassuring tone, while still encouraging prompt action by customers. The company demonstrates availability or neutrality towards contact with customers. Within the sample, 23,47% of letters belong to this group. This letter type can be linked to Coombs' (1995) **distance strategy**, which attempts to weaken the link between the crisis and the organisation. In doing so, companies clearly acknowledge the crisis but present it as an unavoidable risk as a form of justification. The fact that the event is presented as unavoidable minimises the organisation's responsibility.

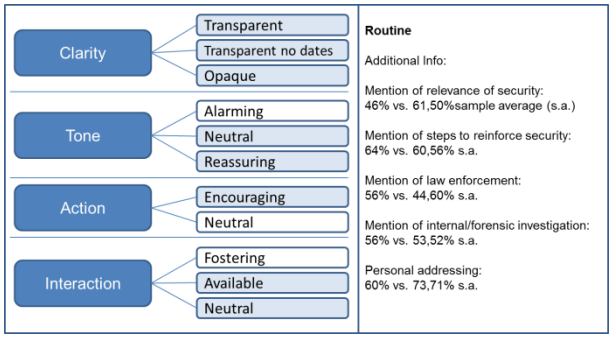


Figure 3.3 Routine letter

3. No worries: This letter type emphasises the minor nature of the risk generated by the event, reassuring the affected customer and listing options for possible action by customers, but not recommending any action. Given the reassuring tone of the letter about the consequences, interaction with the company is not fostered. Within the sample, 19,72% of letters belong to this group. This letter type reflects a **nonexistence strategy**, which attempts to eliminate the crisis by denying or belittling the current risks of concrete consequences for the consumer, thereby suggesting that no crisis exists.

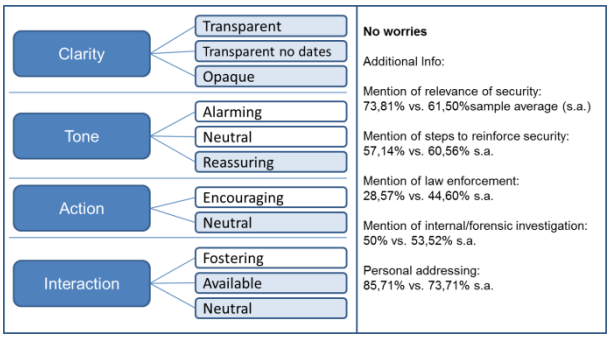


Figure 3.4 No worries letter

4. Junk: This letter type can be easily mistaken for a junk message and therefore discarded soon after the envelope is opened. The description of the incident is often not clear, or if the event description is transparent, no date about the occurrence of the incident is provided. The communication tone regarding the possible consequences and the approach to actions to be taken by affected customers are neutral. This type represents 10,33% of the sample. This type follows the **silence strategy**, in which organisations are obliged to notify, but try nonetheless to achieve a no-notification result by enhancing the possibility that affected consumers do not read the letter.

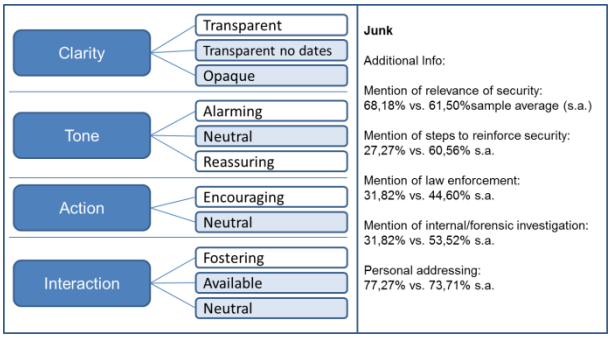


Figure 3.5 Junk letter

5. Cooperation: In this letter type, the facts are clearly described. It emphasises the actions taken by the organisation, while highlighting what actions need to be taken by individuals for their own protection. A statement about the increase of security measures by the organisation is often included and contact with the company is encouraged. One letter out of ten in the sample belongs to this type. This letter type is in line with the **mortification strategy,** which attempt to win forgiveness and create acceptance. Mortification strategies include remediation to offer full support to the victims, repentance to ask for forgiveness and rectification to clearly demonstrate that mechanisms are in place to prevent a similar crisis from occurring again.

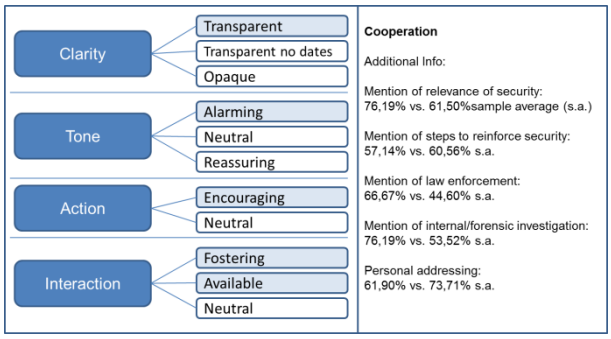


Figure 3.6 Cooperation letter

6. Supportive anyway: While the tone regarding the possible consequences of the data breach is reassuring or neutral and the approach to actions to be taken by individuals is also neutral in this letter type, the company still fosters contact with customers, highlighting its supportive attitude (2,35%). This letter type is in line with the **ingratiation strategy**, which focuses on ways to gain public approval. In this case, approval is sought by supporting consumers while giving the impression that it is a clear additional service the company is offering disproportionate to the actual risk.

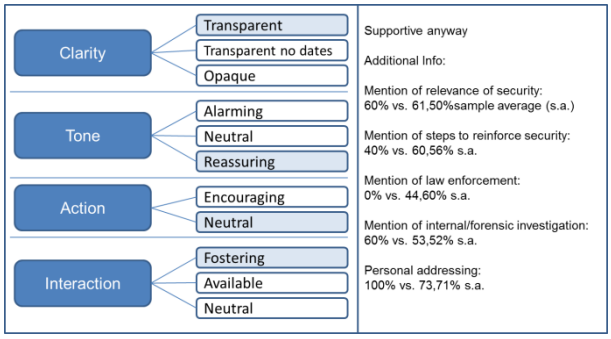


Figure 3.7 Supportive anyway letter

The remaining 6,57% of letters do not fall into any of the six groups.

Type of event	Cold	Routine	No worries	Junk	Cooperation	Supportive anyway	Other	Total
Payment Card Fraud	0	0	5	0	0	0	0	***5***
Unintended Disclosure	12	14	17	4	2	3	4	***56***
Insider	11	6	6	4	1	1	0	***29***
Physical Loss, Portable and Stationary Device	14	4	6	0	5	0	1	***30***
Hacking or Malware	20	26	7	13	13	1	9	***89***
Unknown or other	2	0	1	1	0	0	0	***4***
Total	**59**	**50**	**42**	**22**	**21**	**5**	**14**	***213***
Total (%)	**27,70%**	**23,47%**	**19,72%**	**10,33%**	**9,86%**	**2,35%**	**6,57%**	**100%**

Table 3.3 Data breach by event and letter type

Type of event	Cold	Routine	No worries	Junk	Cooperation	Supportive anyway	Other	Total
Payment Card Fraud	0,00%	0,00%	100,00%	0,00%	0,00%	0,00%	0,00%	***100,00%***
Unintended Disclosure	21,43%	25,00%	30,36%	7,14%	3,57%	5,36%	7,14%	***100,00%***
Insider	37,93%	20,69%	20,69%	13,79%	3,45%	3,45%	0,00%	***100,00%***
Physical Loss, Portable and Stationary Device	46,67%	13,33%	20,00%	0,00%	16,67%	0,00%	3,33%	***100,00%***
Hacking or Malware	22,47%	29,21%	7,87%	14,61%	14,61%	1,12%	10,11%	***100,00%***
Unknown or other	50,00%	0,00%	25,00%	25,00%	0,00%	0,00%	0,00%	***100,00%***
Total	**27,70%**	**23,47%**	**19,72%**	**10,33%**	**9,86%**	**2,35%**	**6,57%**	***100,00%***

Table 3.4 Percentage of data breaches by event and letter type

In the cases where a company could be easily identified as ultimately responsible for the data breach and thereby subject to legal actions, the use of no worries letters in order to minimise the problem was present in a high percentage. Not considering payment card fraud, for which only five cases were recorded, 30,36% of the cases for unintended disclosure, 20,69% of data breaches generated by insiders and 20% of cases of physical loss were associated with a no-worries letter. When the breach was generated by hacking or malware, the distance strategy was the most commonly used through the use of routine letters (29,21%). This finding affirms the notion that hacking and malware tend to be presented as an unavoidable external risk. Junk and cooperation letters were also a frequent option for companies in this case (14,61% each).

The decisions about each single element of notification and the resulting letter style represent the dilemmas that each breached organisation faces. Organisations must take into consideration the clashing motivations around breach notification: to develop clear and effective notification letters in order to comply with the law, and to mitigate the potential harm to the company. Furthermore, organisations often face the supreme dilemma of minimising concrete short-term reputational effects or minimising potential future damages due to customer churn and fines. There

is no universal solution that can be adopted by all organisations in all cases. However, from the analysed data, we can identify preferred behaviours by breached organisations. We will further investigate this aspect in the next chapter.

3.7 Conclusions

The presented data provide relevant insights on the actual achievements of the objectives of DBNLs and help us to answer our research question: '*What are the core elements of consumer notification letters and how do company decisions on what to include and how to express the message define specific letter types?*'

The data highlight how organisations, given the facts and company responsibility, make careful choices with consideration to the costs associated with each letter type. In fact, the data reveal interesting patterns related to the use of specific letters in response to specific events. For example, if the event generating the breach is easily imputable to phenomena that are not fully controllable, organisations tend to be clear about the facts. On the contrary, the identified patterns suggest a tendency to belittle the event (with no worries and junk letter types) when the responsibility of the firm is unquestionable. The type of event, therefore, appears to drive the type of formal response consumers receive from breached organisations.

Thus far, we looked only at the developed components of breach notifications, but it is important also to analyse the arrangements, use of apologies, timing and possible underreporting. We investigate these aspects in the following chapters.

Chapter 4 Slow and (not) safe: An investigation into company reactions to breaches

Based on the definition of letter types, we further the analysis with the identified time for DB detection and reaction in order to better shape possible ameliorative legal solutions. Using the notification database, we identify the differences in the choices organisations make – in terms of if, what and when they notify – also according to the different types of events that generate data breaches, and suggest what can be done to limit such choices and to minimise the risks associated with data breaches.

4.1 Introduction

On January 12, 2015, President Obama proposed the Personal Data Notification and Protection Act, intending to create a federal standard for data breach notifications. The draft bill follows many legislative proposals that have failed to gain passage despite the rising incidence of massive data breaches. In the previous two years, five data breach notification bills were introduced in the Senate alone, yet none garnered sufficient support for passage.

The implementation of a federal law raises certain questions, and different actors may perceive a federal law positively or negatively according to the features of the law. The key elements of the Personal Data Notification and Protection Act are as follows[37]:

- The definition of personal information is more expansive than most state breach notification laws, including home address, telephone number, mother's maiden name and date of birth as data elements;

[37] https://www.workplaceprivacyreport.com/2015/03/articles/identity-theft/the-data-security-and-breach-notification-act-of-2015/

- Companies would be required to implement and maintain reasonable security measures and practices to protect and secure personal information;
- Companies would not be required to provide notice if there is no reasonable risk of identity theft, economic loss, economic harm or financial harm;
- Companies would be required to provide notice to affected individuals within 30 days after discovery of a breach;
- The law would preempt all state data breach notification laws;
- Enforcement would be conducted by the Federal Trade Commission (FTC) or state Attorneys General; and
- No private right of action would be permitted.

The actors involved in the discussion include certain business groups which support federal legislation because it creates a single breach notification standard. Such groups argue that even a more stringent federal standard would be easier to comply with than the current patchwork of 47 different, and often conflicting, state laws.[38] Others include consumer protection groups and Attorneys General. They are concerned mainly because the federal legislation would preempt state data breach notification laws, including those that ensure better protection than the proposed federal standard. For example, with a letter sent on July 7, 2015, the National Association of Attorneys General (NAAG) addressed congressional leaders. NAAG urged them to consider the state laws that have been put in place to protect consumers and not to weaken the role that state Attorneys General play in enforcing data security and protection laws. The letter urges Congress not to make changes to federal data breach notification and data security laws that would reduce the protections that have been put in place by the states.[39] It calls for Congress to avoid to introduce data security and data breach notification laws that preempt those introduced in each state. In essence, it states that preemption interferes with state legislatures' democratic role as laboratories of innovation, and stresses how any federal legislation on data breach notification and data security should recognise the important role of State Attorneys. They are the ones on the front lines responding to data breaches, and helping the residents of their state.

[38] Brendan.

[39] https://www.hipaajournal.com/state-data-breach-laws-should-preempt-federal-laws-says-naag-8012/

4.2 Study approach

In order to contribute to this debate, our analysis followed an approach that was not based on past investigations about data breach trends or evaluation of data breach costs, but on a vast dataset of data breach notifications. However, the findings presented to date by other researchers on the impacts of breach notifications on breached organisations' performance provide a relevant context for our study (see chapter 2, section 5).

Our approach is based on the available content of all data breach notifications in the U.S. in 2014, 47[40] state data breach notification laws and selected extensive reports issued by law firms and available online.[41] These reports were thoroughly examined to identify any mandatory elements of the notification letters. The sample includes 445 notifications sent in 2014 from breached organisations to consumers[42] which were downloaded from the Attorney General websites of four different states and used to verify the choices made by the affected companies. The methodological steps followed in order to conduct the analysis are described below.

1) Identify the states that make the data breach notification letters issued by affected companies publicly available.

2) Download all available letters in the timeframe 1/1/2014-31/12/2014, identifying those sent out in more than one of the four states.

---

[40] Alabama, New Mexico and South Dakota were the only U.S. states that had not enacted a data breach notification law as of 31.12.2016. Since 2018 all states have DBNL.

[41] Data Breach Notification Laws by State (CLLA, 2012)
http://www.clla.org/documents/breach.xls
State Data Security Breach Notification Laws (Mintz Levin, 2012)
http://www.mintz.com/newsletter/2007/PrivSec-DataBreachLaws-02-07/state_data_breach_matrix.pdf
State Data Breach Stature Form (Baker Hostetler, 2013) http://www.bakerlaw.com/files/Uploads/Documents/Data%20Breach%20documents/State_Data_Breach_Statute_Form.pdf
Security Breach Notification Chart (Perkins, 2013) http://www.perkinscoie.com/files/upload/LIT_09_07_SecurityBreachExhibits2.pdf

[42] Additional 45 letters were discarded because either they were second communications or some information was not visible in the downloaded letter.

3) Based on the content of the letter, isolate specific letter elements and create a database to code each characteristic, both at the paragraph level to understand the order of the letter contents and at the sentence level to identify the content and purpose.

4) Perform a data analysis aimed at investigating:

- Possible common patterns in the notifications;
- The timing of the letters and their related usefulness for reducing consumer harm.

4.3 Sample Description

Our desk research revealed that only six states out of 47 made notifications available in 2014 through their Attorney General websites. These states are California, Maryland, New Hampshire, Vermont, Maine and Indiana.[43] The last two make the list of data breaches available to the state residents, but do not provide a copy of the notifications. Full availability of the notifications in the four states is the consequence of specific state laws which make notification in case of a breach mandatory, not only to residents but also to the office of the state Attorney General. The purpose of this requirement is to ensure that the Attorneys General have an overview of the state breach situation and can decide about the level of visibility of the notifications (18 states included such requirement in 2014). In this way, the Attorneys General act as collectors of all data breach notifications affecting state residents.

[43] Washington and Oregon started respectively from mid-2015 and 2016 to give such visibility, after law revision.

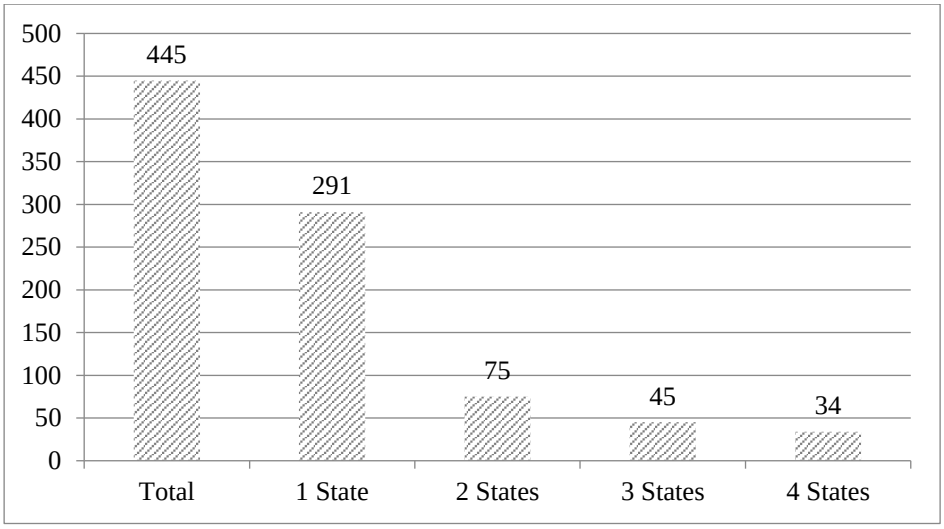


Figure 4.1 Data breach sample 1/1/2014-31/12/2014

The number of analysed letters after removing duplications (same letter sent to different states) amounts to 445. The sample comprises the following number of letters by state: 130 for Vermont, 169 for California, 250 for Maryland and 161 for New Hampshire. Figure 4.1 illustrates the overlapping between the four states: 291 notifications were sent in only one of the four states, 75 in two states, 45 in three states and finally 34 letters were sent to residents in all four states.

It is important to highlight the relevance of the sample. Although the number of the analysed letters may be perceived as low, the 445 letters represent 56,83% of the 783 total cases collected in the U.S. according to the 2014 Data Breach Report.[44] This total comes from the ITRC breach list: a compilation of data breaches known to the public through media operators, Attorneys General offices, other governmental bodies such as the U.S. Department of Health and Human Services and specific sectoral databases.[45]

Observation 1: The high percentage of notifications from four states among the total number of breaches in the U.S. raises the question of underreporting and emphasises the role of a legal notice requirement.

[44] Identity Theft Resource. 2014 Data Breach Reports. 2014.
[45] List of ITRC resources for data breaches available at http://www.idtheftcenter.org/index.php/id-theft/data-breaches.html

While the number of total notification letters affirms the representativeness of the sample analysed in this work, it also suggests the likely existence of a high number of hidden data breaches that are not publicly disclosed. As 43 states are excluded from the analysis (as they did not make notifications publicly accessible in the year of the investigation), we expect a much higher number than 783 for the total number of data breaches in the U.S. in the 12-month study period. In fact, the four states represent only 14,37% of the total number of firms in the U.S. according to the 2012 Economic Census statistics,[46] and 14,98% of residents according to the 2010 Census.[47]

In addition, based on the letters analysed in the four states and the sectors in which the breaches took place, we can identify that approximately 15% of notifications derive from local retail business, service or medical centres acting locally, in which case we can assume that the place of the breach and the residency of the affected individuals coincide. For example, on September 30, 2014 at Gold's Gym, a member was required by an associate to provide his or her credit card three-digit security number, even though Gold's Gym does not require this information. In another case, the online ordering software provider of BringItToMe.com, an online restaurant marketing and delivery service active in San Diego, informed the organisation that they identified unauthorised modifications in their software. This modification could potentially allow new payment credit card information entered between October 14, 2013 and January 13, 2014 to have been obtained by an unauthorised user. We can safely assume that similar events happen across the U.S., with a similar percentage of firms per sector affected by local data breaches which impact the residents of only one state.

The organisation that publishes data breach data, ITRC, itself states, 'we are certain that our ITRC Breach List underreports the problem'.[48] Fur-

[46] http://www.census.gov/econ/census/
[47] Population Distribution and Change: 2000 to 2010. 2010 Census Briefs.
[48] http://www.idtheftcenter.org/id-theft/data-breaches.html

thermore, given the current statistics about cyber-crime and cyberattacks,[49] it is hardly conceivable that less than 800 data breaches were registered across the U.S. in a given year.[50] According to a survey of approximately 300 attendees at the RSA Conference[51], more than 89% of security incidents went unreported in 2007.[52] Moreover, dedicated reports such as the 2014 Data Breach Investigation Report[53] utilise datasets which comprise all confirmed security incidents (over 63.000 globally in 2013), no longer restricting the analysis to confirmed data breaches only.[54] We focus in this research only on the breaches known by the affected organisation, not including unknown breaches, such as undetected malware, and the measures that could be taken to intercept such events.

It is important to distinguish between two possible reasons for the lack of public evidence of a data breach known by the affected organisation. Either the company decides not to disclose the breach, or the notified parties have no reason or incentive to inform the public about the received

---

[49] In 2001, the annual total loss of complaints referred to the IC3 (Internet Crime Complaint Center) amounted to approximately 17,8 million U.S. dollars and grew to 781,84 million U.S. dollars in 2013. In 2012 the amount was 581,44 million U.S. dollars. Statista 2015.

[50] Note that Maine Attorney General only lists data breaches without providing letters for consultation. Maine was therefore not included in the analysis. However, this list allows us to observe that adding a fifth state to the sample there would be additional 62 data breaches, bringing the total to 507 (64,75% of total data breaches then would be covered by 5 States out of 47).

[51] The RSA Conference is an international conference series on IT security that takes place in the United States, Europe, Asia/Japan, and the United Arab Emirates. The name RSA refers to the public-key encryption technology developed by RSA Data Security, Inc., which was founded in 1982. The abbreviation stands for Rivest, Shamir, and Adleman, the inventors of the technique.

[52] http://cybercrimeupdates.blogspot.it/2008/08/over-89-of-security-incidents-not.html /

[53] Verizon. 2014 Data breach investigations report. 2014

[54] Verizon uses the following definitions:

Security incident: any event that compromises the confidentiality, integrity, or availability of an information asset.

Data breach: An incident that resulted in confirmed disclosure (not just exposure) to an unauthorized party.

notification.[55] Regarding the first point, the topic of underreporting has been extensively discussed in previous studies, which suggest that organisations may prefer to focus on profit margins instead of the security of personal data. Therefore, they may underreport data breaches, primarily out of concern for their business liability and reputation. In this perspective, disclosure makes traceable an otherwise untraceable security breach, attracting publicity and potentially prompting costly legal action or regulatory scrutiny.[56] According to a white paper[57] from ThreatTrack released in 2013, which polled 200 security professionals in U.S. enterprises, 57% had experienced a data breach that they did not disclose.

Regarding the second reason, it is clear that companies, once they have fulfilled the legal obligation to inform affected consumers, have no incentive to inform the media or other third parties about the breach in order to avoid reputational damages. However, it is less clear why Attorneys General in 14 states do not make such information public, even after notification by breached companies according to the state data breach notification laws. We can expect a delay in informing the public if investigations are ongoing, but a complete lack of information has no apparent motivation, apart from preventing additional organisational burden. AG offices need to prioritise public access. They must adequately manage the incoming notification flows and establish procedures for the publication of the letters on their websites, possibly increasing interaction with the public.

From the percentages highlighted above, with four states reporting more than 50% of the total breaches reported in the U.S., those AG offices in the notification flow that do not publicly disclose known data breaches via their websites or in other ways may generate a counterproductive effect on the public perception of the issue. From the presented numbers, we can assume that in those states where Attorneys General do not

[55] There is also a third reason, but it is a temporary one, notifications may in fact be delayed if a law enforcement agency informs the business that notification may impede a criminal investigation or jeopardize national or homeland security.
[56] Schwartz, P., & Janger, E. (2007). Notification of Data Security Breaches. 105 Michigan Law Review 913. 2007.
[57] Malware Analysts Have the Tools to Defend Against Cyber-Attacks, But Challenges Remain, Threattrack security. White Paper, November 2013.

disclose because they are not in the flow or because they decided not to do so, the media and the other actors largely fail to identify and record data breaches, even when they are reported directly to the customers.

Attorneys General can play a decisive role in the emergence of non-reported data breaches if supported by the necessary legal requirements (government notice requirement). However, it is also a matter of their willingness to foster the visibility of the data breach notifications. In fact, 12 state AG offices currently prefer not to disclose to the public such information, limiting the effect of the data breach notification laws. A federal law would, therefore, facilitate the opportunity to centrally manage the visibility of notifications received from companies and would allow for the collection of accurate national data breach statistics.

4.4 Letter content

The requirements of the DBNLs in the 47 States vary from one state to another. These differences generate complexity for organisations dealing with customers residing in multiple states. Unfortunately, there is no single form letter that guarantees compliance with all state laws, and most state breach notification laws do not establish specific requirements for the notice's content.[58] However, an assessment of the state breach notification statutes that do create minimum requirements can identify the most frequently required elements, which allows for general recommendations to organisations of what to include. Fifteen of 47 states with DBNLs include such minimum requirements in their statutes. The requirements are listed in Figure 4.2.

[58] Some organizations opt for filling the gap with an annex which fulfils case by case each state legislation.

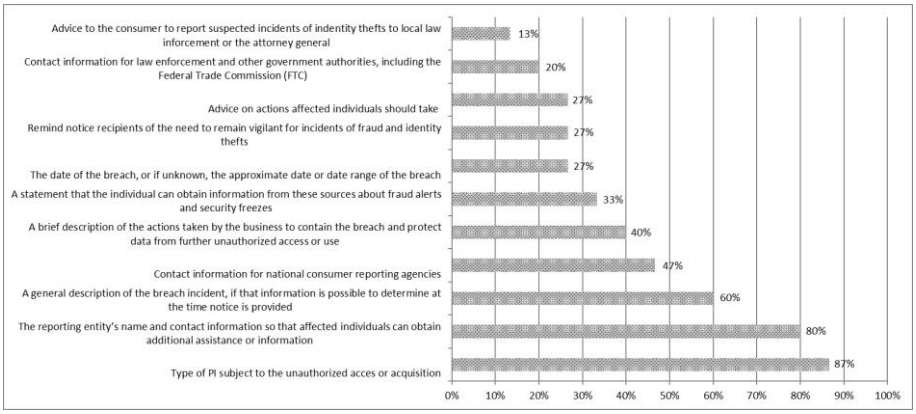


Figure 4.2 Mandatory elements of data breach notification

State	No. of elements included in legislation	
California	7	63.64%
Hawaii	5	45.45%
Illinois	3	27.27%
Iowa	4	36.36%
Maryland	5	45.45%
Massachusetts	2	18.18%
Michigan	5	45.45%
Missouri	5	45.45%
New Hampshire	4	36.36%
New York	2	18.18%
North Carolina	8	72.73%
Oregon	6	54.55%
Vermont	5	45.45%
Virginia	5	45.45%
West Virginia	4	36.36%

Table 4.1 Mandatory elements of data breach notification by state

While in 32 states the content of notifications is not formalised in any way by the data breach notification law in place, Figure 4.2 highlights that 13 out of 15 states that do list requirements (87%) require the letters to specify the type of personal information subject to unauthorised access or acquisition. A large number of states (80%) also require the notifications to

specify the reporting entity's name and contact information so that affected individuals can obtain additional information. Only 60% of state laws require that companies provide consumers with specific information on what has happened (a general description of the breach incident). Meanwhile, general advice on actions that affected individuals should take is mandatory in only four states out of 15. Some state statutes provide for more explicit requirements. For example, a statement indicating that individuals can obtain information from specific sources such as the Federal Trade Commission and consumer reporting agencies and a reminder notice of the need to remain vigilant for incidents of fraud and identity theft are mandatory in five and four states, respectively.[59] Only four states mandate that an organisation specify the date of the breach. The specification of the date of the breach and the date of the breach discovery would effectively support the achievement of the purposes of DBNLs, *right to know* and *sunlight as disinfectant*. The first date is essential in order to support the consumer in evaluating the seriousness of the situation and the need for a prompt reaction. The second date highlights the organisation's speed in communicating breaches to consumers. Both dates enable assessment of the organisation's capacity to detect breaches. Within the sample, 272 letters out of 445 indicated at least the date of the breach discovery within the organisation, while 268 indicated at least the date of the breach or, if unknown, the approximate date or date range of the breach; 166 letters specified both and 70 none.

From the 272 letters in which the time of the breach discovery is specified, we calculated the average time in days from the discovery of the event to the moment of the communication to consumers and other parties. We define this variable as **notification time**: the time the organisation needs to assess the situation after breach detection, to finalise the notification letter and to activate the necessary communication channels towards customers and other relevant parties (e.g., Attorney General, customer credit reporting agencies). The average[60] notification time was 38 days (see Table 4.2), with only 124 cases under 30 days. The median value was 32,50. The data presented in Table 4.2 also indicates that some sectors are more reactive than others.

[59] Table 4.1 does not include a requirement set in California, where the letter has to specify whether notice was delayed as a result of law enforcement investigation.
[60] Once we eliminated 6 outliers according to the z score rule.

Table 4.2 Notification time

Sectors	Notifications	Average (days)	Over 15 days	Over 30 days	Over 45 days	Over 60 days	Median (days)
Financial and Insurance Services	42	34,19	83,33%	47,62%	19,05%	9,52%	29,00
Other Business	67	34,27	77,61%	47,76%	26,87%	11,94%	28,00
Retail/Merchant	48	34,92	79,17%	52,08%	27,08%	8,33%	33,00
Educational Institutions	25	50,28	84,00%	64,00%	44,00%	32,00%	41,00
Government and Military	17	41,35	82,35%	47,06%	29,41%	17,65%	28,00
Healthcare - Medical Providers	59	41,51	84,75%	64,41%	44,07%	11,86%	39,00
Nonprofit	8	36,25	87,50%	37,50%	37,50%	25,00%	22,00
Total	**266**	**38,00**	**81,58%**	**53,38%**	**31,58%**	**13,53%**	**32,50**

Types of event	Notifications	Average (days)	Over 15 days	Over 30 days	Over 45 days	Over 60 days	Median (days)
Hacking or Malware	120	39,03	82,50%	52,50%	33,33%	14,17%	32,50
Insider	26	44,92	80,77%	65,38%	38,46%	15,38%	40,50
Payment Card Fraud	2	39,00	100,00%	100,00%	0,00%	0,00%	39,00
Physical Loss, Portable and Stationary Device	46	38,80	89,13%	63,04%	41,30%	13,04%	36,00
Unintended Disclosure	69	32,87	75,36%	42,03%	20,29%	11,59%	28,00
Unknown or other	3	41,67	66,67%	66,67%	33,33%	33,33%	34,00
Total	**266**	**38,00**	**81,58%**	**53,38%**	**31,58%**	**13,53%**	**32,50**

PII	Notifications	Average (days)	Over 15 days	Over 30 days	Over 45 days	Over 60 days
SSN	59	35,41	81,36%	49,15%	28,81%	10,17%
account / credit card or debit card number	57	34,37	78,95%	50,88%	26,32%	10,53%
Email / Password / User / ID card number	9	23,00	55,56%	22,22%	11,11%	0,00%
Personal Health Information	11	31,55	72,73%	54,55%	36,36%	0,00%
SSN and account / credit card or debit card number	41	38,88	75,61%	39,02%	31,71%	21,95%
Other combinations	89	43,95	89,89%	67,42%	38,20%	16,85%
Total	**266**	**38,00**	**81,58%**	**53,38%**	**31,58%**	**13,53%**

We classified breaches across seven primary industries: financial and insurance services (BSF), retail/merchant (BSR), educational institutions (EDU), government and military (GOV), healthcare and medical providers (MED), nonprofit (NGO) and other business (BSO). Financial and insurance services and retail/merchant sectors exhibited similar behaviour, taking 34 days on average to complete the notification process. In comparison, government and military and healthcare and medical providers required 41 days on average. Education institutions reacted even

slower (50 days). A nonparametric k-sample test on the equality of medians[61] revealed that the k samples (6 sectors[62]) were drawn from populations with different medians with probability = 0,040 and Pearson chi2(5) = 11,6503. For the type of event (4[63]), the Pearson chi2(3) = 10,9090 and probability = 0,012.

The type of PII accessed or acquired does not seem to generate a relevant impact on the notification time. In fact, when only SSN were accessed, the average notification time was 35 days. Similar values result when only bank accounts or credit or debit card numbers are the breach target.

Finally, we investigated the role of the event type in determining the notification time. The definition of the type of event is derived by privacyrights.org, which classifies the events that generate notifications as follows: unintended disclosure (sensitive information posted publicly on a website, mishandled or sent to the wrong party via email, fax or mail), physical loss (lost, discarded or stolen non-electronic records or portable or stationary device), insider (someone with legitimate access intentionally breaches information, such as an employee or contractor), hacking and malware (electronic entry by an outside party, malware and spyware), payment card fraud (fraud involving debit and credit cards that is not accomplished via hacking), unknown or other (all other cases).

The analysis revealed that organisations need more time from the breach discovery to assess the situation and initiate the notification process in cases of insider breaches (45 days), and less time in cases of unintended disclosure (33 days). This finding is likely related to the internal investigation dynamics, which are very straight forward in cases of human error and more complex in cases of fraud.

Within the sample, 268 letters indicated the date of the breach, either when the generating event took place or when it started (and thereby

[61] Shapiro-Wilk W test confirmed that group data (grouped both by sector and type of event), specifically notification time, do not show a normal distribution.

[62] NGO sector not taken into consideration given the limited number of observations (8).

[63] Payment card fraud and Others not taken into consideration given the limited number of observations.

when the potential harm was induced). In cases of unintended disclosure, the date could be when the file was disseminated. In cases of insider breach, this date could be the date when the employee developed criminal intentions. We define the time between the breach and the notification date as **uninformed exposure time**. During this period, customers are not aware of the risk they are exposed to and cannot undertake any defensive action. The data reveal a worrying situation, with an average of 132 days[64] (see Table 4.3) between the communication and the day when the potential harm started, and with 29% of the cases[65] having uninformed exposure times of over three months.

Both a nonparametric k-sample test on the equality of medians and a Kruskal-Wallis equality-of-populations rank test were performed on the sectors.[66] The first produced the following result: Pearson chi2(5) = 20,0929 and probability = 0,001, highlighting that the k samples (6 sectors[67]) were drawn from populations with different medians. The second test further revealed that there is a statistically significant difference in uniformed exposure time between the six sectors, with chi-squared = 20,914 with 5 d.f., probability = 0,0008. The results also confirmed a statistically significant difference by event type, with chi-squared = 40,397 with 3 d.f.[68] and probability = 0,0001.

[64] Once eliminated 3 outliers represented by 4 insider cases, discovered more than 3 years after the potential data breach.

[65] Information extracted from the created database.

[66] Shapiro-Wilk W test confirmed that group data (grouped both by sector and type of event), specifically uninformed exposure time, do not show a normal distribution

[67] NGO sector not taken into consideration given the limited number of observations (8).

[68] Payment card fraud and Others not taken into consideration given the limited number of observations.

Sectors	Notifications	Average (days)	Over 30 days	Over 60 days	Over 120 days	Over 180 days	Median (days)
Financial and Insurance Services	49	60,43	55,10%	22,45%	14,29%	10,20%	36,00
Other Business	67	113,60	62,69%	47,76%	25,37%	16,42%	41,00
Retail/Merchant	58	166,14	87,93%	65,52%	39,66%	27,59%	98,00
Educational Institutions	17	214,41	76,47%	64,71%	47,06%	47,06%	102,00
Government and Military	14	128,07	64,29%	50,00%	35,71%	28,57%	47,50
Healthcare - Medical Providers	56	168,84	83,93%	46,43%	30,36%	21,43%	60,00
Nonprofit	4	29,50	25,00%	25,00%	0,00%	0,00%	21,00
Total	**265**	**132,90**	**71,70%**	**47,55%**	**29,06%**	**21,13%**	**58,00**

Type of event	Notifications	Average (days)	Over 30 days	Over 60 days	Over 120 days	Over 180 days	Median (days)
Hacking or Malware	122	157,38	80,33%	63,11%	37,70%	30,33%	88,50
Insider	24	258,38	83,33%	75,00%	54,17%	33,33%	147,50
Physical Loss, Portable and Stationary Devi	55	50,47	58,18%	20,00%	5,45%	3,64%	34,00
Unintended Disclosure	62	112,44	62,90%	32,26%	24,19%	14,52%	36,00
Unknown or other	2	35,50	50,00%	0,00%	0,00%	0,00%	35,50
Payment Card Fraud	0	-	0,00%	0,00%	0,00%	0,00%	-
Total	**265**	**132,90**	**71,70%**	**47,55%**	**29,06%**	**21,13%**	**58,00**

Table 4.3 Uninformed exposure time

Finally, it is also important to consider the delay between the date of discovery and the start of the potential harm, which was calculated in 163 cases where both dates were available. We define this gap as the **breach detection time**. The average is 113,10 days, with significant variation across the data breach types, as outlined in Table 4.4. This variation suggests that the approach and regulations to breaches should be differentiated according to the data breach type. Notifications sent for data breaches generated by insiders and hacking arrive to customers already late, even if sent on the same date of the discovery. In fact, for these breaches, the breach detection time is over six months. In contrast, data breaches due to physical loss and unintended disclosure can be better addressed through prompt notification, as organisations discover these data breaches more rapidly (in 18 and 78 days, respectively).

Type of event	Notifications	Average (days)
Hacking or Malware	71	158,10
Insider	12	249,83
Physical Loss, Portable and Stationary Devi	33	17,70
Unintended Disclosure	46	78,33
Unknown or other	1	26,00
Total	**163**	**113,10**

Table 4.4 Breach detection time

Observation 2: Understanding and open communication of the breach detection time, notification time and resulting uninformed exposure time is essential to enable consumers' reaction to breaches and effective sectoral intervention.

The timing analysis alone demonstrates that the first objective of DBNLs, to protect the consumers' right to know, is not adequately fulfilled. Indeed, the timing poorly matches individuals' need to defend themselves against potential identity theft promptly. Criminals may even use the late notifying reaction by breached organisations to their advantage. Furthermore, the fact that many state statutes do not yet mandate minimum information for the content of notifications provides organisations with discretion, which may not support customers' conscious reactions to a breach.

This analysis of timing can also raise company awareness about the risks related to different types of events that generate data breaches and about specific dynamics associated with these events that put customers' data at risk for various periods of time. As we estimated, in cases of hacking or insider breach, organisations need at least 90 days more to identify a data breach in comparison to cases of physical loss or unintended disclosure.

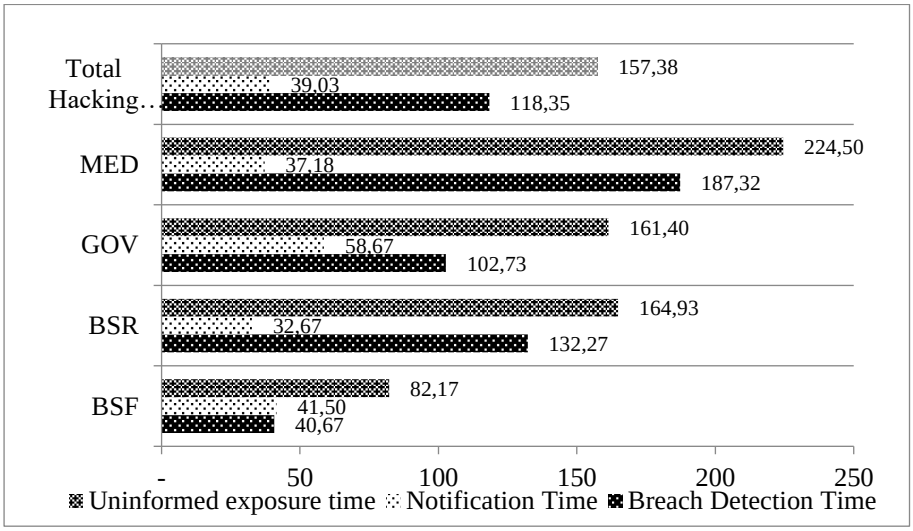


Figure 4.3 Hacking or Malware

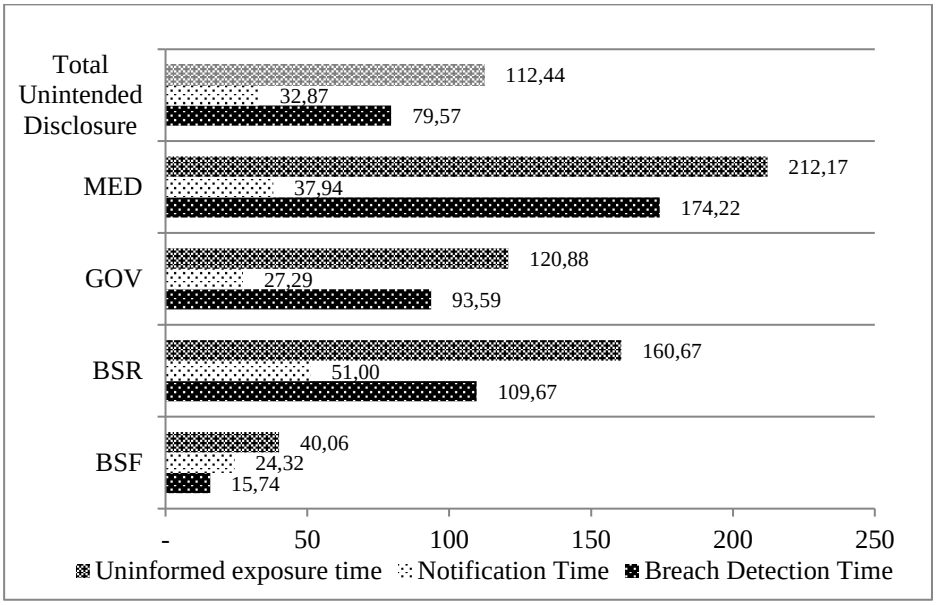


Figure 4.4 Unintended disclosure

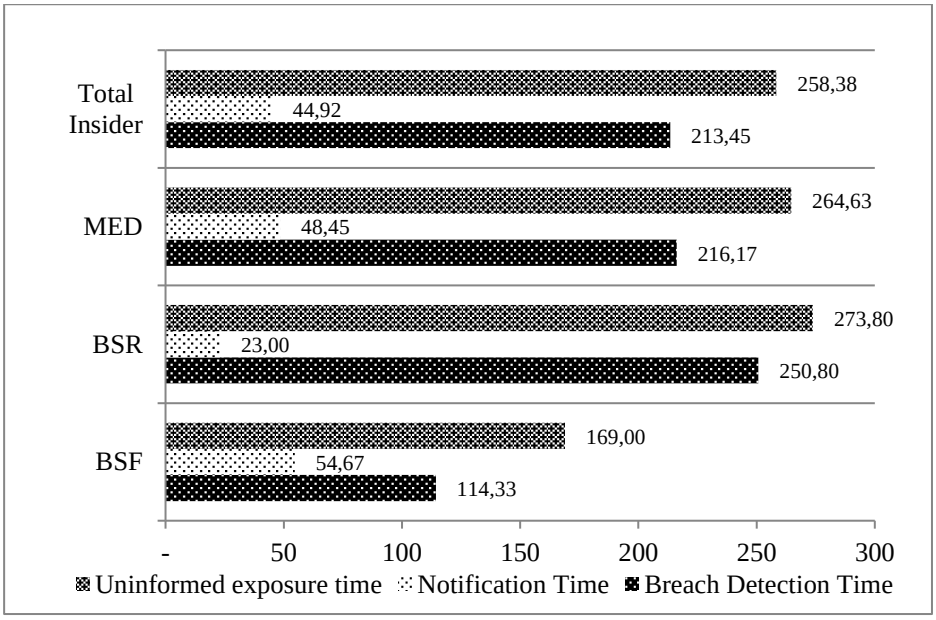


Figure 4.5 Insider breach

Based on the data summarised above, Figures 4.3, 4.4 and 4.5 illustrate the different dynamics related to three types of breach-generating events (hacking or malware, unintended disclosure, insider) applied to specific sectors. The specific breach detection time, notification time and the resulting uninformed exposure time highlight the strong performance of the financial sector in comparison to other sectors, but also reveal how the retail sector is the most reactive once the breach is detected in cases of hacking or malware and in cases of insider breach.

The Personal Data Notification and Protection Act announced by President Obama does not provide for the mandatory inclusion of any date in the content of breach notices to individuals. The analysis conducted in this chapter suggests that the consequences of this lapse may be to limit citizen risk awareness when receiving notifications.

4.5 Letter style

The predefined letter elements, if they aim to be effective, should make public notices useful and easy to understand, meaning that they should contribute to mitigating the risks of unauthorised and uncontrolled access to customer personal information. In cases of data breaches, a prompt notification to customers can help them to mitigate the damage caused by identity theft[69] or provide them with the opportunity to take preventative steps to protect themselves from possible identity theft, such as through suggesting placing fraud alerts and activating credit monitoring services.

The form of breach notification is therefore essential to ensure that the right message is sent, sufficient information is provided, and motivational incentives for precautionary actions are given. The fact that many state statutes do not mandate minimum information to be included in notification letters highlights the poor regulations in place to guarantee the quality and the appropriateness of the means – the notification – to

[69] Data breaches and identity theft. Prepared statement of the Federal Trade Commission before the Committee on Commerce, Science and Transportation. U.S. Senate 109th Congress. 2005.

achieve the goal of timely alerting consumers to trigger a prompt reaction against identity theft and other negative consequences of data breaches.

In the few cases where content is specified by law, some of the mandatory elements cannot be modulated, as they are objective details such as the date of breach or contact information. However, the majority of the components can be manipulated, resulting in messages with different tones – whether alarming or reassuring, clarifying or confusing – about the event and its consequences. The following paragraphs focus on those elements and their sequence in sample letters. According to the analysis in the previous chapter, there are four key features:

1. **Clarity:** Clarity of the incident description and of the PII involved
2. **Tone:** Communication tone on the possible consequences given the organisation's reaction
3. **Action:** Approach to actions to be taken by the affected customers
4. **Interaction:** Stance toward interaction with affected customers

We benefit from the previous analysis of how 'conventional' components in business communication – such as bad news, explanation, apology, prefatory and closing buffers – are embedded in these four elements. By using this classification, we can also indirectly leverage the analysis performed by Veltsos on the traditional bad-news components applied to data breach notifications. In essence, we can identify both negative messages that tend to focus on low-risk, routine situations and the approaches used when negative news is not about refusals or rejections. In recent times, we encounter many variations on negative messages, such as notices of cancelled flights (Jansen & Janssen 2011), product recalls, negative policies or organisational news (Alred, Brusaw & Oliu 2011, Bovée & Thill 2012, Shwom & Snyder 2012), rate increases and price hikes (Guffey & Lowey 2011). Bisogni's and Veltsos's research benefit from these previous works in investigating the intersection of business communication and information security in the form of breach notification letters.

We applied our classification to the letters in the sample based on the analysis of a full year of notifications, examining the traditional bad-news

elements under the perspective of fostering better communication towards consumers affected by data breaches. The existence of options at the disposal of the breached organisation demonstrates that companies have specific opportunities to belittle breach events even while complying with the law. These elements were analysed for each of the 445 letters sent in 2014. It is also relevant to examine the sequence used to communicate bad messages. These sequences can be interpreted with the support of the existing research in the field of communicating negative messages, as many lines of research inquiry have centred on arrangement as a key aspect of composing and disseminating negative messages.

The order or **arrangement** of components within a negative message can follow two patterns: indirect and direct. An indirect arrangement starts with an explanation, delivers the bad news and finally closes with an expression of goodwill. In contrast, a direct arrangement presents the bad news, delivers an explanation and then ends with a statement of goodwill. The indirect or inductive pattern is strongly recommended by most authors (Hynes 2008, Kolin 2007, Alred et al. 2011). Such authors suggest avoiding negative words altogether, highlight how diplomacy and 'reader psychology' are fundamental elements in corporate correspondence and present the indirect arrangement as more effective especially when stakes are high (Alred et al. 2011). The consensus among textbook authors is that the indirect pattern should be used when the problem is considerable or when there is the likelihood that the reader will be shocked or upset (Bovée & Thill 2012, Shwom & Snyder 2012).

Conversely, the fact that the stakes are high may be the driver for using a direct pattern in data breach notifications (Veltsos 2012). Readers must know that their PII has been breached and their privacy is in danger. With the bad news in the opening paragraph writers can capture the readers' attention immediately and 'shake' them into action (Lehman & DuFrene 2012, p. 105). The direct pattern clearly provides a stronger incentive to keep reading about protective measures and is considered 'good ethics and good business' (Locker and Kienzler 2010, p. 437). Below is an example of the two types of arrangement (direct and indirect respectively), the first sent by Dreslyn and the latter sent by Liberty Tax.

Dear [INDIVIDUAL NAME]:

We deeply value your business. Your security is our top priority, which is why, as a precautionary measure, we are writing to inform you of a data security incident that involves your personal information. [250]

Dear Liberty Tax Customer:

Liberty Tax makes every effort to protect the confidentiality and integrity of our customers' confidential information. The state of Maryland requires that if a business experiences a security breach where personal information that, combined, may pose a threat to a consumer if misused, that business must notify any affected consumers residing in Maryland. Once a security breach is detected, a business must also conduct in good-faith a reasonable and prompt investigation to determine whether the information that has been compromised has been or is likely to be misused, i.e. for identity theft. If the investigation shows that there is a reasonable chance that the data will be misused, that business must notify the affected consumers.

Unfortunately, our office has discovered some tax returns that may have been filed with the IRS and respective states without the consent of the taxpayers. [282]

The combination of the four letter elements defines the ultimate form of communication towards consumers and the type of message that is received. Meanwhile, the decision on the arrangement may provide a relevant indication of the organisation's willingness to capture and direct the attention of consumers toward the negative event and its consequences.

Finally, the selected strategy is also strongly linked to the role of apology and its relation with corporate responsibility. Essentially an apology is offered by the organisation accepting responsibility for the crisis and asking for forgiveness (Benoit & Drew, 1997, Fuchs-Burnett, 2002). A variety of additional components can complement this definition, including expressions of remorse/sympathy, expressions of regret, preventative measures, and reparation (Benoit & Drew 1997, Cohen 1999, Fuchs-Burnett 2002, Patel & Reinsch 2003). However, companies clearly have at their disposal a wide range of communication strategies, from apology strategy to less accommodative strategies, for example giving no comment, denial, excuse or justification (Bradford and Garrett 1995, Dean 2004, Lyon & Cameron 1998). Less accommodative strategies (partial apologies) are likely to resolve disputes in which the extent of each party's fault is unclear and difficult to establish. (Patel & Reinsch 2003). Coomb and Holladay (2008) advised:

> Given the higher costs associated with apologies, crisis managers can confidently offer compensation and/or express sympathy in the lower to moderate responsibility crises rather than relying on apology as the default. Ethically and pragmatically, if management knows it is at fault, an apology is advised. It is unethical to evade responsibility when it is known. However, not accepting responsibility (expression of sympathy and/or compensation) is an important and viable option to an apology when responsibility is unknown or ambiguous.

The apology strategies adopted by organisations can be classified in three main groups: accepting responsibility by apologising, using expressions of sympathy and not using any form of apology or sympathy. Below are two examples of how apologies and regrets were respectively formulated within data breach notifications in the sample.

'We apologize sincerely for this incident and hope the steps we have instituted help allay any concerns you may have.'

'We deeply regret any concern or inconvenience this incident may cause'

Observation 3: Data breach notification laws ensure that organisations contact customers after the discovery of a breach affecting PII, but offer poor indications for the style and content of the notification. Even in states where some letter elements are mandatory, companies have leeway in delivering bad news related to the breach, which offers the opportunity to belittle the actual risk and the possible consequences.

Table 4.5 illustrates how the letter characteristics are represented in the sample. In most cases, the letters were transparent in describing data breach events and accessed PII, even if relevant dates were not specified in some cases. The analysis further reveals that most of the organisations described the event in a very transparent manner. However, none of the analysed letters provided the number of the breached records: information that could directly reveal the extent of the breach and therefore the extent of the company's failure in ensuring data security. A neutral tone about the possible consequences of the breach was also used in the majority of cases (60%), and 30% of letters reassured the customers. Fi-

nally, organisations generally demonstrated availability towards customers in terms of supporting them in the post-event process (85,45%), but only a few fostered contact with customers (8,54%).

Clarity - Event	Notifications	%	Junk	No worries
Opaque	36	8,09%	√	√
Transparent	354	79,55%		√
Transparent no dates	55	12,36%	√	√
Total	**445**	**100%**		
Tone	**Notifications**	**%**		
Alarming	46	10,34%		
Neutral	267	60,00%	√	
Reassuring	132	29,66%		√
Total	**445**	**100%**		
Action	**Notifications**	**%**		
Encouraging	219	49,21%		
Neutral	226	50,79%	√	√
Total	**445**	**100%**		
Interaction	**Notifications**	**%**		
Available	382	85,84%	√	√
Fostering	38	8,54%		
Neutral	25	5,62%	√	√
Total	**445**	**100%**	**29**	**74**

Table 4.5 Data breach notification main components

The combination of the letter characteristics defines the ultimate form of communication. We identified the clarity of the event, the tone regarding the consequences, the action suggested to the reader and the interaction fostered by the organisation as drivers for identifying the letter type. In the previous chapter, we proposed six letter types according to the combination of these elements, which represent different strategies organisations can select when composing a notification letter.

The analysis reveals that companies belittled the event in 23,15% of the cases by sending one of the following two letter types:

- No worries letter: This letter emphasises the minor risk generated by the event, reassuring the affected customer, and lists options for possible action by the customer. However, it does not recommend any action. Given the reassuring tone adopted about the consequences, interaction with the company is not fostered. In the sample, 74 letters belong to this group, which includes notifications with the following characteristics: opaque or transparent no dates clarity of the event, neutral tone, neutral action, available or neutral interaction.

- Junk letter: This letter can be easily mistaken for a junk message and therefore discarded soon after the envelope is opened. The description of the incident is not clear, or if it is transparent, no dates about the occurrence of the incident or its discovery are provided. The communication tone about the possible consequences and the approach to actions to be taken by affected customers is neutral. Within the sample, 29 letters belong to this group, which includes notifications with the following characteristics: opaque or transparent or transparent no dates clarity of the event, reassuring tone, neutral action, available or neutral interaction.

A further element of discretion that provides a clear indication of the type of message that the company wants to deliver to customers is represented by the use of apologies. To better analyse this element, we classify the type of data breaches according to the assumed company responsibility for the event.[70] In particular, we investigated the role of apology in order to understand the different options available better. We assume that if a company decides to apologise, then it has admitted its responsibility for the event ask for forgiveness.[71] We analysed this aspect at the sentence level. The use of expressions such as 'we apologize' and 'accept our apologies' were coded as apology, while expressions such as 'we are sorry' and 'we regret' were classified as regrets. In a few cases neither apologies nor regrets were offered (labelled as none in Table 4.6).

Type of event	Apology	Regret	None	Total	% Apologies
Payment Card Fraud	8	0	0	***8***	100,00%
Unintended Disclosure	53	37	11	***101***	52,48%
Insider	24	15	7	***46***	52,17%
Physical Loss, Portable a	34	31	9	***74***	45,95%
Hacking or Malware	63	96	51	***210***	30,00%
Unknown or other	3	3	0	***6***	50,00%
Total	**185**	**182**	**78**	***445***	**41,57%**

Table 4.6 Use of apologies

The results detailed in Table 4.6 were translated into three levels of responsibility: *** high level of responsibility, with over 50% use of apologies; ** medium, with over 33% use of apologies and * low, with less

[70] Ibid

[71] Benoit, W. L., & Drew, S. (2002). Appropriateness and effectiveness of image repair strategies. Communication Reports, 10, 153–163. 1997 and Fuchs-Burnett, T. Mass public corporate apology. Dispute Resolution Journal, 57(3), 26–32. 2002

than 33%. We can consequently organise data breach event types by the associated level of responsibility. The result is the following:

1. Payment card fraud: Fraud involving debit and credit cards that is not accomplished via hacking, but primarily through mishandling of information by the personnel of the organisation involved. ***
2. Unintended disclosure: Sensitive information posted publicly on a website, mishandled or sent to the wrong party via email, fax or mail. The employees' lack of attention and poor process control often play a decisive role. ***
3. Insider: Someone with legitimate access intentionally breaches information, such as an employee or contractor. Lack of control and screening in the recruiting/partnership phase is one of the reasons behind such data breaches. ***
4. Physical loss: Lost, discarded or stolen non-electronic records or portable or stationary devices. A lapse in the security of the premises or lack of attention by personnel may facilitate such events. **
5. Hacking and malware: Electronic entry by an outside party through malware and spyware. Easy to be presented as unavoidable. *

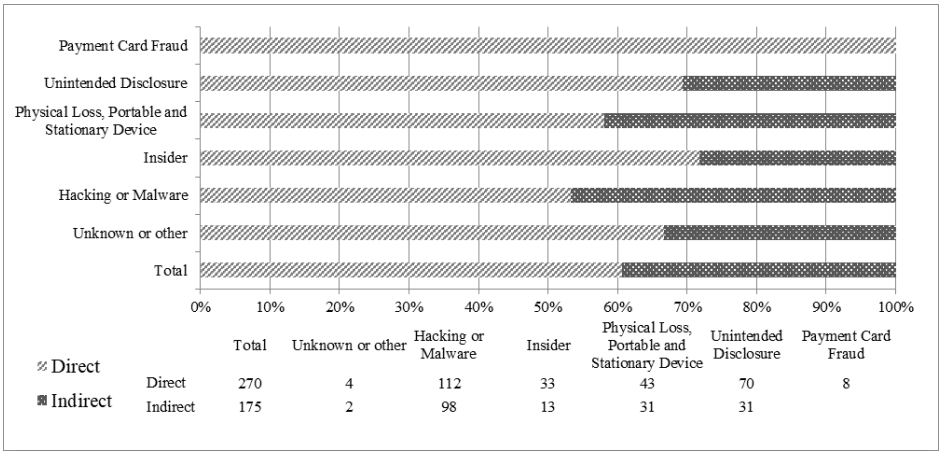
Similar trend is present analysing the tone element alone, reflected by the decision to reassure consumers regarding the consequences of the breach. In the cases in which a company could be easily identified as ultimately responsible for the data breach and therefore possibly subject to legal actions, the use of a reassuring tone in letters in order to minimise the problem was present in high percentage. As per Table 4.7, this tone was present in 100% of the cases of payment card fraud, 44,55% of the cases of unintended disclosure, 40,54% of physical loss cases, 21,74% of insider cases and 18,10% of cases of hacking or malware.

Tone vs. Event	Alarming	Neutral	Reassuring	Total	% Reassuring
Payment Card Fraud	0	0	8	***8***	100,00%
Unintended Disclosure	6	50	45	***101***	44,55%
Physical Loss, Portable a	8	36	30	***74***	40,54%
Insider	5	31	10	***46***	21,74%
Hacking or Malware	26	146	38	***210***	18,10%
Unknown or other	1	4	1	***6***	16,67%
Total	**46**	**267**	**132**	***445***	**29,66%**

Table 4.7 Tone and events

Finally, we analysed the arrangement of letters through coding the use of direct and indirect order patterns in the analysed sample. We compared the use of patterns with the consensus of the related debate in business communications textbooks. The analysis reveals (Figure 4.6) that the need to immediately capture the attention of readers to foster their action is not in line with the suggestion offered by business communication authors to employ an indirect pattern in high-stakes situations for the writer or reader. The rationale behind this contradiction is that the stakes may become even higher if the reader is not 'shaken' into action. Otherwise stated, breached organisations must convince consumers that a potential problem occurs and encourage them to act, particularly when their action could be useful. In the sample, 60,67% of the letters employed the direct pattern as an instrument to overcome optimism bias and rational ignorance.

In line with the previous finding that in cases of hacking or malware the time span between the data breach and the notification exhibited a conspicuous delay, the results demonstrate that the direct approach was used the least (53,33% vs. 47,67% indirect) in cases of hacking or malware. This finding suggests that there is no urgency to capture the attention of the reader in order to foster his/her reaction if the event happened more than three months before the notification. In cases of payment card fraud or unintended disclosure the use of the direct approach was consistently higher (100% and 69,31%, respectively). This finding suggests that companies may consciously decide to use the direct approach when they feel it is useful given the short detection time. At the same time, they may opt more frequently for the indirect approach when they are aware that it is already too late for consumers to protect themselves against the consequences of the data breach.



	Total	Unknown or other	Hacking or Malware	Insider	Physical Loss, Portable and Stationary Device	Unintended Disclosure	Payment Card Fraud
Direct	270	4	112	33	43	70	8
Indirect	175	2	98	13	31	31	

Figure 4.6 Direct and indirect patterns

In short, organisations clearly exploit the fact that many state statutes do not yet mandate minimum information for the content of breach notifications, providing them with significant elements of discretion. Companies often use such elements in order to limit their eventual reputational damage or the short-term costs posed by the activation and management of communication channels (e.g., call centres, higher rate of activated credit monitoring). Thus, an organisation's exercise of discretion may not always support customers' conscious reactions to a breach, and the results of such 'flexibility' can produce suboptimal effects for society.

The notice-based approach of the state breach notification statutes in the U.S. represents an important step toward increasing a widespread corporate culture towards data security. The fear of reputational sanction is a significant stimulus, and recognising its value, it is important to limit any easy 'way out' for companies. Nevertheless, consumers may not open and read notification letters or act on their content because they are already overwhelmed by communications from commercial companies or because the letters themselves do not convey their content unhindered and effectively. Consequently, letters under the current legal framework may not provide particularly useful information about a company's security practices, or about the measures customers should take to protect themselves from harm.

A federal law represents a unique opportunity to regulate the content of notification letters and to ensure that letters properly convey the content.

Based on the series of prior analyses, we recommend that a federal law mandate the following notification elements to serve consumers' interests better:

- **Full transparency** on the clarity of the incident description and of the breached PII involved, also indicating the number of residents affected by the breach to allow consumers to self-evaluate the size of the breach.[72]
- **Avoidance of a reassuring tone** in depicting the possible consequences of the data breach. This restriction would no longer enable organisations to sugarcoat the consequences of a breach, which can discourage consumers from acting.
- **Clear recommendation** to the affected customers of actions to mitigate breach-related risks. This may include encouraging customers to carefully review bank and credit card statements or activate credit monitoring and credit freeze services.
- **Foster interaction** with affected consumers by communicating full company availability in supporting affected individuals and in clarifying any unclear aspects of the notification and of the breach.

By mandating or promoting these elements, policymakers can ensure that companies will have less leeway in drafting notifications and that they will support consumers in better engaging in post-breach self-protection.

4.6 Conclusions

The presented analysis was performed following an innovative approach which was not based on a traditional investigation of data breach trends or evaluation of data breach costs, but on a vast dataset of data breach notifications. The research was feasible thanks to the letters made available by four Attorney General offices. It is clear that in order to reinforce the role of information disclosure in combating misaligned incentives and information asymmetries, this level of visibility should not be limited to California, Maryland, New Hampshire and Vermont. If a federal data breach law is implemented, we can expect a much higher number of no-

[72] Based on the letter sample analysed this information is never reported in the notification letters to consumers but often present in the notification letter to Attorney General sent in the same timeframe, indicating a clear intention of the firms not to disclose such element.

tifications made public, fostering the emergence of 'hidden' notifications. This would also support a more precise estimation of the total number of breaches experienced in the country. Awaiting developments in federal law, states in which Attorneys General are already in the communication flow for notifications can powerfully contribute by making these notifications publicly available. This effort would also support the second goal of data breach notification laws: to provide sunlight as disinfectant. Besides, it can produce a better analysis of the phenomenon of data breaches, and also help to investigate more deeply the causes of the statistical mismatch between data breaches and the prevalence and magnitude of cybercrime.

Additionally, it is essential to have knowledge of the actual scale of the breach, of the timing of breach detection and of notification drafting. Therefore, we suggest that the disclosure of both the date of breach and the date of discovery should be mandatory in the notifications made by breached organisations to consumers and relevant authorities. The analysis of such information enables us to study sectoral dynamics generated by the different types of events in order to inform better prevention and response in case of a data breach. We found that organisations belonging to certain sectors are significantly slower in reacting after a breach discovery. Thus, the relevant differences in the breach detection capabilities various industries should be taken into account.

Regarding the content of notification letters, few states legally require a minimum set of elements to be included in notifications. The consequence is that consumers must rely entirely on the letter style of the breached organisation to understand the seriousness of the situation and to be adequately alerted about the breach. Organisations, in turn, may focus on profit margins instead of the security of personal data, using the leeway in the law to belittle the event or to reassure consumers in order to reduce costs in the short term. Therefore, we conclude that data breach notification laws should dictate extensive mandatory elements regarding the content of notice to individuals. Under the current Personal Data Notification and Protection Act, the foundation for the possible forthcoming federal law, only three elements are mandatory: a description of the categories of sensitive personally identifiable information accessed or acquired, a toll-free number to contact the business entity or the agent

of the business entity from which the individual may learn what types of sensitive personally identifiable information the business entity maintained about that individual and the toll-free contact telephone numbers and addresses for the major credit reporting agencies. Such weak restrictions for communicating breaches will enable companies to manage almost independently the level of alert communicated to the consumer, not safeguarding the latter.

Under the current framework, data breach notification laws serve more as 'sunlight as disinfectant' in the medium-to-long run than as effective and prompt preventative measures against identity theft. The prevalence of reassuring tone, underreporting and time delays together demonstrate that businesses cannot work without strict supervision in this arena. Mandatory data breach notifications, regulation of notification content and timing and associated penalties for non-compliance are fundamental pillars for achieving more responsible data management practices which embody the right-to-know and sunlight-as-disinfectant principles. The implementation of a federal law or ad hoc reviews of state laws that can define stricter rules and better control the described elements represent two options to reinforce the effects of the current legislative framework towards a better safeguard against identity theft. Apart from specific features that a federal data breach notification law can enact, the added value of a federal solution includes providing uniform indications to consumers and companies, which can resolve the issues related to the current patchwork of data breach notification laws. In cases of breaches affecting the residents of different states, the current patchwork framework results in a notification system that is challenging for companies to navigate. Such regulatory complexity increases the consumer's risk of remaining unprotected. Replacing the current mix of state laws with a single comprehensive and standardised federal law would also enhance the response time of firms through outlining equal and clear steps to follow after a breach. Organisations would no longer have to undergo time-consuming cross-state analysis to answer questions regarding what information is covered and when and how notification must be provided. Finally, a federal approach would allow for centralisation of data collection, enabling the government to develop and maintain accurate national data breach statistics to monitor the dynamics of the data

breach phenomenon and to promptly react by means of audit, penalties or legislative revisions.

Chapter 5 Estimating the size of the iceberg from its tip: An investigation into unreported data breach notifications

Leveraging on the findings of the previous chapter, we investigate in this chapter the prevalence of unreported data breach notifications, focusing on measures to be taken in order to unveil known breaches that do not become public. The research questions that we address are 'what effects do specific DBNL provisions have on reported data breaches?' and 'how large is the portion of data breaches that we are unaware of?'

5.1 Introduction

A decade has passed since the enactment of data breach notification laws in numerous U.S. states. These laws require companies that have suffered a data breach to inform the customers whose data might have been exposed. The intent of DBNLs can perhaps be best summarised in the phrase 'sunlight is the best disinfectant'. Whether the goal of incentivising better security practices has been realised is the subject of ongoing debate (e.g., Romanosky et al. 2011, Bisogni 2016). What is clear, however, is that DBNLs have offered greater visibility into the state of data breach events in the United States.

Nevertheless, it is also clear that an unknown number of breaches are hidden from view. The Identity Theft Resource Center's (ITRC) Breach Report and similar databases only contain breaches that have become public knowledge. As Figure 5.1 illustrates, a breach must first be detected by the affected organisation (move from 4 to 3) and then one or more relevant parties must be notified (move from 3 to 2) before the breach can become publicly reported (move from 2 to 1). Many breaches never make it past the last hurdle. Indeed, the notification letters that were made public by the Attorneys General in four U.S. states

account for approximately 40% of all reported breaches in 2014 according to the ITRC, while these states host only 14% of U.S. firms and 15% of the population.

This chapter seeks to provide an enhanced understanding of the submerged part of the iceberg (i.e., hidden breaches). We first leverage the differences among DBNLs in different U.S. states to estimate the impact of certain provisions on how many breaches triggered notifications, yet did not become publicly reported. In other words, we estimate level two of the iceberg (see Figure 5.1). We model the number of reported breaches as a function of the different DBNL provisions across the states, while controlling for the size of different sectors in each state and other factors.

Our model also includes the impact of the 'risk-of-harm' exemption in some DBNLs, which allows breached organisations not to notify affected consumers if after a reasonable investigation they determine that there is no reasonable likelihood of harm to customers stemming from the breach. States with this exemption report fewer breaches. This means that affected organisations never notify anyone in the first place. By modelling the impact of the risk-of-harm exemption on the number of reported breaches, we estimate one way in which how breaches are detected but not notified: a portion of level three of the iceberg.

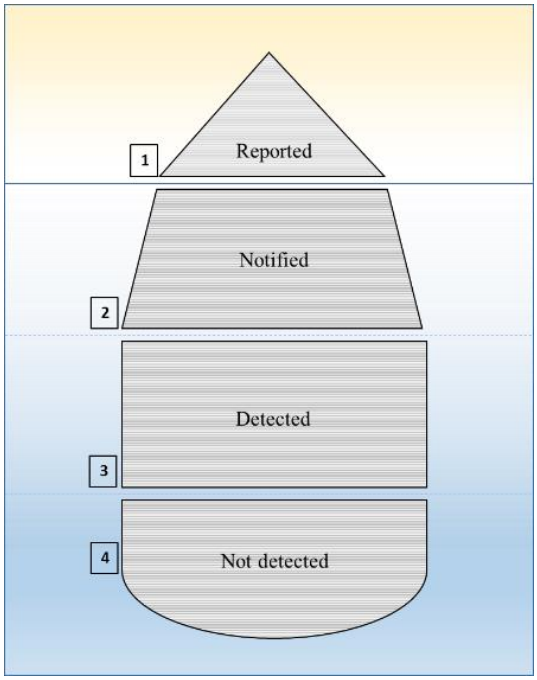


Figure 5.1 Data Breach Iceberg

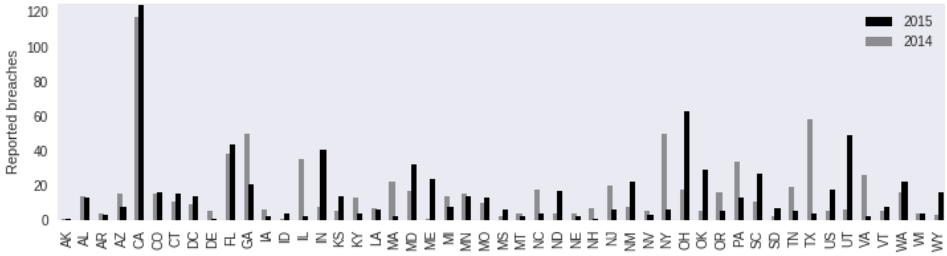


Figure 5.2 U.S. data breach statistics 2014–2015 by state

Finally, we catch a glimpse of the deepest part of the iceberg – level four – through analysing the notification letters in four states. In those states, the Attorneys General publicly report all notifications. We coded all breach causes mentioned in those letters. Interestingly enough, the sector with the lowest breach rate ('retail and other business') is also the one with the highest ratio of breaches caused by 'hacking' and lowest ratio of 'unintended disclosure'. This finding suggests that security practices in this sector do not detect a significant number of breaches, contributing to a breach rate that is between two and 12 times lower than other sectors. The notification letters also allowed us to look at notification and detection times by modelling the time span between the notification and the breach discovery by the organisation and/or the breach event. By doing so, we identified the breach causes that more than others require notification times, not in line with the individuals' need to defend themselves against potential harm promptly.

Our analysis reveals that there is quite a lot that is not known about U.S. data breaches. That being said, the security community knows much less about breaches in Europe. This is evident from browsing public databases that gather known data breaches, such as the ITRC report, which contains only breaches affecting U.S. residents. The EU has recently introduced its own industry-wide DBNLs: a directive[73] and regulation[74] will extend the weaker and sector-specific security breach notification laws

[73] Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016

[74] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016

that previously applied to the telecom sector. Our analysis can help the EU to learn from the results of almost 15 years of regulation in the U.S. since the enactment of the first DBNL in California,[75] offering relevant insights in view of the adoption of the Data Protection Package.[76] In short, the contributions of this chapter are as follows: (i) to model the impact of DBNL provisions on the number of known data breaches and breach notification times, while controlling for sector and state differences; (ii) to estimate the number of breaches for which notifications have been issued but that are not publicly reported and (iii) to discuss key elements of DBNLs that make those laws effective in view of the implementation of the European regulation on security and data breaches.

5.2 Objectives of data breach notification law

Data breach notification laws are typically justified with two objectives. The first is to protect customers' right to know when their personal information has been stolen or compromised. As Schwartz and Janger (2007) described, informing customers allows them to protect themselves – for example, by changing their passwords or monitoring their credit card statements for signs of abuse. A second objective is to create incentives for organisations to take adequate steps to secure the personal information they store. The reputational damage resulting from a reported breach activates 'the sunlight as disinfectant' principle, leading companies to invest more in cybersecurity, and disinfecting organisations of shoddy security practices (Ranger 2007).

The ideal final result of pursuing these two objectives is summarised in the Federal S.177 - Data Security and Breach Notification Act of 2015: 'to protect consumers by requiring reasonable security policies and procedures to protect data containing personal information, and to provide for nationwide notice in the event of a breach of security'. DBNLs also contribute to improving the security of the overall Internet ecosystem by in-

[75] California Civil Code § 1729.98 enacted in 2003.

[76] Consisting of the General Data Protection Regulation (GDPR) and the Directive for data processing by law enforcement for the purposes of prevention, investigation, detection or prosecution of criminal offences. The Directive is to be implemented by 6 May 2018, and the Regulation will apply from 25 May 2018.

creasing transparency for the security community, policymakers and citizens. In this respect, ENISA (2011) believes that the introduction of data breach notification requirements is an important development with the potential to increase the level of data security and foster reassurance among citizens regarding how their personal data is being secured and protected. The introduction of DBNLs acts in an environment with two opposing trends: increased breach risk due, among other factors, to greater digitalisation and increased investment in security due to better awareness of the risk.

We build on prior work by examining the degree to which DBNLs support the public visibility of data breaches. Previous research has studied specific sectors (e.g., the medical sector investigated by Kwon and Johnson 2015) or state-level differences in the overall number of reported breaches (e.g., Faulkner 2007). However, no study to date has looked at both simultaneously. This combined focus is essential, as there are key differences among the sectors in terms of the use of information technology and the presence of more specific laws on how to deal with personal data, as is the case for the finance and health sectors.

5.3 Data

Our research objective in this chapter is to study the impact of key provisions of DBNLs on the number of breaches that move from detection to notification to being public reporting, while controlling for sector and state differences. We also study notification letters for additional insights into underreporting and the timing of breach notifications. The following paragraphs describe the datasets we used.

Breach datasets. There are numerous initiatives aimed at providing details on data breaches, such as the Privacy Rights Clearinghouse (PRC) database, the Identity Theft Resource Center Breach Report and the Veris Community Database.[77] Given the current coverage of data breach

[77] Appendix I provides links to these breach datasets, as well as the data sources used by ITRC.

notification laws in the U.S., one might expect a joint institutional repository for data breaches, but this does not exist. As of December 31, 2016 22 states require notifications to the Attorney General, but only seven states publish details of the events and the notification letters.[78]

We use two datasets of breach incidents: the ITRC list and a self-compiled dataset based on the notification letters made available by the AG offices in four states (see Appendix I). We identified the ITRC as the most comprehensive source of breaches in the U.S.: in 2014, the PRC reported 330 breaches versus 783 reported by the ITRC. From the ITRC dataset, we collected the date of each breach, the sector and state of the breached firm and the number of records breached.

From the notification letters, we manually extracted the date of the letter and the breached company. We then determined the company's sector. Other details include the type of incident, number of affected records and the incident date, which can be compared to the notification date (i.e., when the letter was sent). Table 5.1 lists the summary statistics from these datasets for five primary sectors. The data we extracted from the notification letters covers all of 2014, and we selected the 2014 and 2015 events from the ITRC dataset.

[78] California, Maryland, New Hampshire, Vermont (included in our analysis), Washington, Oregon and Montana. Washington and Oregon started, respectively, from mid-2015, beginning of 2016 to give such visibility, after law revision. Montana from mid- 2015.

Table 5.1 Datasets

	ITRC dataset (for breach count model)	**Notification letters (for breach time model)**
Educational institutions (EDU)	108	21
Financial and insurance services (FIN)	106	97
Retail and other Business (BSO) [79]	561	222
Medical & healthcare providers (MED)	590	71
Government and military (GOV)	137	19
Total breaches	1.502	430[80]
Median records breached[81]	2.500	6 (only NH/MD customers)
States covered	47[82]	CA, MD, NH, VT
Dates covered	1-Jan-2014 to 31-Dec-2015	1-Jan-2014 to 31-Dec-2014

Sector size. We built a breach rate for each *state and sector* by dividing the number of breaches per state and sector by the number of firms active in that state and sector. The firm data was extracted from the 2012 U.S. Census. As the census data excludes governmental offices, we used the number of medical centres as the denominator for breaches in the

[79] We merged BSR (Retail/Merchant) with BSO (other business).

[80] The majority of these breaches, that is 311, are also in the ITRC. The missing records are due to ITRC grouping smaller breaches together, and ITRC recording some breaches in 2013, while the letters were sent in 2014. However, we use the datasets in separate models, so the overlap, or lack of, does not matter.

[81] ITRC reports breached records, or customers, for 55% of the incidents. For our dataset, it is stated in 39% of incidents, and only in letters to the Attorneys General of New Hampshire and Maryland.

[82] The ITRC dataset includes all U.S. states except West Virginia, plus District of Columbia. We further remove the records for the states of Alabama, New Mexico, and South Dakota, as do not have DBNL.

governmental sector. Our assumption is that the number of medical centres is driven by the number and size of cities in a state, which similarly influence the number of governmental offices.[83]

Control variables. When modelling the relationships, we controlled for the size of various sectors in different states and attributed the remaining variation to differences among the DBN laws. However, there might be other systematic reasons that lead to less, or more, data breaches occurring or being reported in a state. We used a number of variables to control for such differences, such as crime rates, household income and the concentration of firms per population. These controls ultimately had no decisive impact on our models (see Appendix IV).

DBNL provisions. We selected a set of DBNL provisions to include as variables in our models. We did not include all provisions for substantive and statistical reasons. The substantive reason is that some provisions could not be codified clearly among the states: for example, the definition of personal information has too many variations across states. The statistical reason is that some of the categories were too sparse and included only a few states, which would bias the regression results. The selected provisions are as follows:

inform_credit, inform_ag_np & inform_ag_p: All DBNLs require affected consumers to be notified. In order to identify the effect of additional notification flows on the number of reported breaches, we coded two variables for whether the law also requires informing credit agencies and/or informing Attorneys General. In the latter case, we distinguished between *inform_ag_np* and *inform_ag_p,* where the difference is whether the Attorney General publishes the notification letter on their website (ag_p) or not (ag_np). These provisions also affect the probability that a specific event can be known by additional actors, such as banks and the media. If more actors are aware of the breach, it becomes more likely that the event will reach the public domain. The breaches included in the ITRC list are not only those reported by the Attorneys General, but also breaches that the media reported, with or without the AG being notified or reporting about them.

[83] Except for the District of Columbia, which we exclude due to concentration of governmental offices.

penalty_cap & priv_cause: All DBNLs include penalties for not notifying about a breach. Some, however, include a cap on the financial penalties. This cap can be fixed per breach (e.g., in Oklahoma) or per single violation (e.g., in the District of Columbia) or both (e.g., in Utah). The existence of a cap on the penalty for not complying with the law defines a priori the risk for not notifying. Some laws include a so-called 'private right of action': the possibility for consumers to sue entities for failing to comply with the data breach notification statute. This provision increases the potential penalty for non-compliance in terms of breach notification.

risk_harm: The safe harbour provisions in different DBNLs are difficult to bring into a common set of categories.[84] We focus solely on the presence of a risk-of-harm exemption, which states that a breached organisation only has to notify if the organisation determines that the breach constitutes a reasonable likelihood of harm to the customer.

Limitations. A major limitation stemming from the ITRC and other breach datasets is that a breach is reported in the location of the headquarters of the company. However, this might not be where the breach actually occurred if the company is active in multiple states. In addition, breach notification procedures are tied to the residency of the affected customers. In such cases, a company active in several states might follow the strictest DBNL among all the states to simplify its processes. This limitation is common to all studies which conduct similar analyses. One solution presented in Appendix II is to rerun our models with a dataset that excludes the financial and business sectors, which contain the most multi-state companies. The direction of the coefficients does not change even in this case, indicating robust results.

5.4 Explaining the number of reported breaches per state

We now model the impact of the different DBNL provisions on the number of reported breaches. Our approach assumes that the probability of

[84] Several statues include encryption as a safe harbor provision, but some do this with a definition of encryption, while other leave it undefined. The exemption due to the application of sectoral specific regulation, i.e., Financial and Medical sectors is already pictured by the sectoral analysis we performed.

a breach in a specific sector (i.e., the number of breaches per organisation in that sector) is the same across different states. In other words, we assume that differences in the number of reported breaches per state and sector are caused by differences in the DBNLs and the control variables, rather than by systematic differences in security practices among states or by attacker preferences for companies in certain states over others.

Figure 5.3 plots the breach count versus firm count for each combination of state and sector (colour-coded by sector). Given the distribution of the data, we used a *negative binomial regression* to model *breach rates*—the number of breaches per state-sector, offset by the number of organisations in that state-sector. This method is the widely recommended way to model rates (Hilbe 2011). Using a negative binomial distribution is also consistent with prior work on breaches (Edwards et al. 2015).

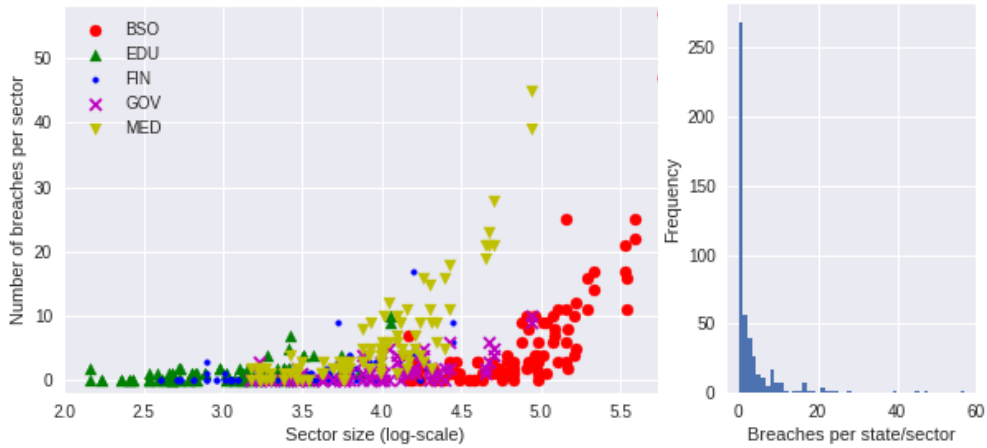


Figure 5.3 Left: Breach count versus organisation count per sector, Right: Histogram of breaches per sector/state/year

The regression results are presented in Table 5.2. In general, regression coefficients in a negative binomial distribution can be interpreted as 'incident rate' ratios, and they are also 'multiplicative'. That is, they tell us how much more or less likely an incident (here, a data breach) is likely to be counted (here, detected and reported).

The results reveal that the sectoral differences are much stronger than DBNL provisions, and among the DBNL provisions, Attorneys General

who publicly report notifications cause, on average, a 43% increase in reported breaches in that state. This effect was to be expected, though not perhaps its magnitude. More surprising is the fact that the requirement to report to credit agencies leads to a 34% increase, all other things being equal. Allowing the risk-of-harm exemption significantly decreases reports, by 21%. The penalty variables have no significant effect. The following paragraphs discuss these main findings in more detail.

Table 5.2 Breach count regression model

Variable	**Coefficient (Std Err)**	**Incident Rate Ratio (95% CI)**
inform_ag_p	0,361 (0,149) *	1,43 (1,07 -1,93)
inform_ag_np	0,044 (0,101)	1,05 (0,86 – 1,28)
inform_credit	0,289 (0,094) **	1,34 (1,11 – 1,61)
penalty_cap	-0,053 (0,085)	0,95 (0,80 – 1,12)
priv_cause	0,119 (0,094)	1,13 (0,94 – 1,35)
risk_harm	-0,231 (0,104) *	0,79 (0,65 – 0,97)
Fin	1,376 (0,136) ***	3,96 (3,03 – 5,15)
Med	2,186 (0,098) ***	8,90 (7,33 –10,80)
Edu	2,482 (0,135) ***	11.97 (9,13 –15,58)
Gov	0,745 (0,127) ***	2.11 (1,64 – 2,70)
Bso	NA—base sector	NA
sector_size	*Offset*	*Offset*
(Intercept)	-9,963 (0,135) ***	NA

Negative binomial disp: 5,179. N= 478. Deviance null/residual: 1.179/563. McFadden pseudo R^2: 0,23. (Full diagnosis available in Appendix II)

Effects of notification flows. Reported data breaches increase by more than one third when Attorneys Generals publish notification letters or when breached organisations must notify credit agencies. In the first case, the contribution of Attorneys General in improving the level of visibility of breaches from notified to reported is clear and direct, as they themselves publish the received notification letters.

For credit agencies, the mechanism is less clear. One explanation is that agencies contribute to increasing the number of reported breaches by informing other actors outside the communication flow dictated by the state DBNL, who then contribute to making the breach public. Another explanation is that these agencies provide an additional notification to consumers, in addition to the one they receive directly from the

breached organisation. This may cause consumers to take the breach more seriously and may increase the probability that consumers report to the media. This interpretation is consistent with the findings of Ablon et al. (2016), who found that a surprising 44% of consumers learned of the breach from other sources before receiving an official breach notification. The most common method of discovery that participants recalled was through media reports (28%), followed by notifications from a third party, such as a bank (16%).

Consumers can play an important role in informing the media, in addition to the Attorney General and credit agencies, of a breach. To explore the role of consumers in making breaches public, we compare the size of reported breaches across states with different notification authorities.

Thus far, the level of our analysis has considered all data breaches equal, not taking into account the number of records affected. However, we know that data breaches come in all shapes and sizes, from a breach enabling access to a few records containing personal information to impactful mega breaches.[85] The number of accessed records defines the size of the breach.

Figure 5.4 presents cumulative distribution functions (also known as CDFs, or cumulative histograms) of *the number of records per reported breach*[86]—a proxy for breach size—in six scenarios related to the notification flow. The x-axis is the number of records in a breach (cropped at 6.000 for readability), and the y-axis is the cumulative percentage of all breaches with that number of breached records or lower. The scenarios include the Attorney General not being informed, being informed but not publishing or being informed and publishing notification letters for the rows, and credit agencies being informed or not for the columns.

In two combinations, consumers are the main, if not only, actor that can make the media aware of the breach: *not_inform_ag/not_inform_credit* (no authority informed) and *inform_ag_np/not_inform_credit* (AG is in-

[85] A mega-breach is commonly defined as a breach of more than 10 million records.
[86] As stated earlier, the ITRC has the number of records for only 55% of the breaches.

formed but does not publish the notifications). In these scenarios, the median number of records affected by the breach is higher than all other combinations, at 2.929 and 6.000, respectively. This finding is consistent with consumers serving as the main source for reporting breaches: larger breaches means more affected consumers, which increases the probability that one or more of them makes the breach public.[87] [88]

We can make two additional observations. The scenarios in the top row, in which the AG is notified and publishes the notifications, have the smallest median records affected. This means that in these states we know about both small and larger breaches. Similarly, in comparing the two columns, the column where the credit agency is notified consistently has smaller median records. This too indicates some public reporting mechanism after the credit agencies are notified.

[87] Given that our analysis is based on medians, the impact of Mega breaches is limited. In the ITRC database only six breaches have more than 10 million records in the timeframe 2014-2015.

[88] It is important to note that the reputational effects of (missing) notifications may also depend on the nature and significance of the PI breached, in addition to the size of a breach. However, we cannot say much about the nature of the PI from the data, and assume breaches to be similar in this regards.

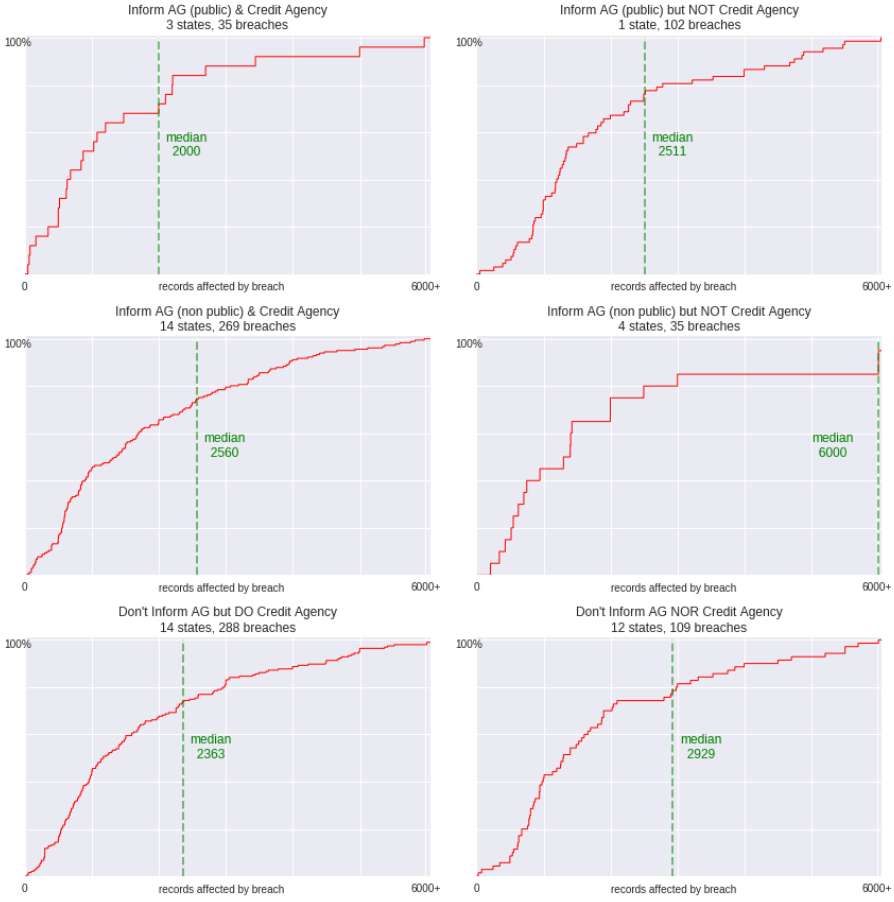


Figure 5.4 Cumulative histogram (or CDF) of records affected by breach

The x-axis is cut at 6.000 for readability; the y-axis is the percentage of breaches with x or less records affected. Each plot represents one notification-flow combination.

Effects of penalties. When companies become aware of a data breach they face two options:

1) They decide not to notify and bear the risk of penalties, of a private right of action (if present) and of potential reputational damage should the breach become public at a later stage. Inaction will generate immediate savings, avoiding costs associated with the notification process, customer services operations and customer redress.
2) They decide to notify and accept the consequence of such a disclosure, such as the costs of notification, call centres, customer

support, identity theft insurance or credit monitoring, legal fees, regulatory fines and the potential loss of market value or lost business. This behaviour avoids the penalties and reputational risk related to a breach becoming public at a later stage.

When assessing the two options, organisations must be aware that data breaches are not just breaches of security. However, breaches of trust between companies and their customers, and can result in not only negative publicity, but also lost business, lawsuits and fines that can threaten the viability of the business.

The breach of trust that may result from intentionally hiding a breach that then becomes known to the community can be restored less easily than if companies follow option two. In the latter case, organisations can communicate that data breaches are a common phenomenon in the sector and are not necessarily dependent on the company's security investment and practices.

Our model findings suggest that companies evaluate whether to issue notifications based primarily on the reputational damage that results from the lack of notification, and less on the tangible consequences of not notifying under DBNLs. The direct financial consequences, tested via the private cause of action and penalty cap provision, were both insignificant in the model. In contrast, the coefficient for credit reporting agencies being notified (inform_credit) suggests that companies fear the reputational consequences of a lapse in notification. If a hidden breach becomes public (through other channels), organisations will have misled not only consumers, but also other organisations.

Risk-of-harm exemption. The negative impact of the risk-of-harm exemption on the number of reported breaches confirms that when the option is given, companies tend to use it in one out of five breaches. Thus, 21% fewer data breaches are notified when the exemption is offered.

Sectoral differences. The differences across sectors demonstrate a much stronger impact on reported breaches than the different DBNL provisions. For example, the finance, medical, education and government sectors have respectively 4, 9, 12 and 2 times higher rates of reported breaches per firm than the 'retail and other business' sector.

Such sectoral differences are to some extent expected, as some sectors are subject to additional federal laws which govern breach notifications (e.g., the Gramm–Leach–Bliley Act for financial institutes, or the Health and Accountability Act). The financial sector, in particular, also has a higher level of security than other sectors (Security Scorecard 2016). However, it is less clear why retail and other businesses experience lower breach rates. Possible explanations include that this sector is targeted less (suggesting that its data is less attractive than that of the financial and health sectors), that often does not detect breaches (due to underinvestment in security), or that the breaches often do not become public (due to their size). The typical causes of breaches in each sector, which is available from the AG notification letters, can shed some light on this question. Table 5.3 presents the difference between the *observed* and *expected* causes of breaches in each sector.[89] The expected value is calculated by multiplying the row total and column total for a cell, and dividing by the grand total. We observe the following patterns:

- In retail and other business, hacking represents a larger proportion of all breaches than in other sectors, which may indicate a lower level of network security. Incidents of unintended disclosure are lower than in other sectors, which points to either underreporting or less vigilant monitoring and fewer process controls in place to identify these kinds of events.
- In contrast, unintended disclosures cause a very high proportion of breaches in the governmental sector, highlighting either weak personal data handling processes or particularly effective monitoring of the processes, or a combination of both. Insider attacks represent the lowest proportion of breaches, which is in line with the fact that more security background checks are performed compared to other sectors.
- In the medical sector, physical losses are most prevalent, possibly reflecting the unique nature of health services, where data physically travels more during service delivery than elsewhere. Besides, the high

[89] We use the breach causes from privacyrights.org, which are as follows: unintended disclosure (sensitive information posted publicly on a website, mishandled, or sent to the wrong party via e-mail, fax, or mail), physical loss (lost, discarded, or stolen nonelectronic records, or portable or stationary devices), insider (someone with legitimate access intentionally breaches information—such as an employee or contractor), hacking and malware (electronic entry by an outside party, malware, or spyware), and unknown or other (all other cases, including payment card fraud).

proportion of insider theft likely derives from the fact that many professionals must have access to the data.
- In the finance sector, physical loss is the least prevalent, possibly due to the greater use of digital capabilities than in the rest of the economy (McKinsey Global Institute 2015).

Table 5.3 Contingency table with the difference between observed and expected breaches by cause and sector.

Cause / Sector	Hacking	Insider	Physical loss	Unintended disclosure
BSO	+24,3% (133 vs 107)	-8,7% (21 vs 23)	-5,3% (36 vs 38)	-43,1% (29 vs 51)
EDU	-10,0% (9 vs 10)	-50,0% (1 vs 2)	0% (4 vs 4)	+40,0% (7 vs 5)
FIN	-2,1% (46 vs 47)	0% (10 vs 10)	-43,7% (9 vs 16)	+36,4% (30 vs 22)
GOV	-66,7% (3 vs 9)	-100,0% (0 vs 2)	-33,3% (2 vs 3)	+250,0% (14 vs 4)
MED	-50,0% (17 vs 34)	+85,7% (13 vs 7)	+83,3% (22 vs 12)	+12,5% (18 vs 16)

Alternative model specifications. In addition to the presented breach count regression model, we attempted a number of different model specifications, namely including interaction terms (between laws and sectors), limiting the dataset to the three local sectors and adding control variables. These more complex models, however, do not perform better based on the Akaike information criterion (AIC) than our parsimonious model. They are presented in the Appendices for interested readers.

5.5 Estimating the total number of data breaches

As the model identifies the effect of DBNL provisions on the number of reported breaches, we can estimate how many breaches would be notified and reported across the U.S. if all DBNLs – or a federal law – required credit agencies to be notified, the Attorney General made all notifications public, and the risk-of-harm exemption was removed. In this scenario, underreporting would be limited to those cases where the breached companies do not detect the breach or where they do not disclose detected breaches.

The results of the estimation are represented in Figure 5.5, with one subfigure for each sector. The green dots represent the observed breach counts (each dot is one state/year). The blue line is the fitted model, depicting the breach count that the model predicts for that combination of independent variables (inform_credit, inform_ag_p, risk_harm). The red pluses are the predicted counts if the laws in all states were stricter (i.e., require the credit agencies to be notified, the Attorney General would make all notifications public, and the risk of harm exemption is removed. The Pearson correlation between the predicted and observed values ranges from 0,53 for finance to 0,92 for BSO and medical).

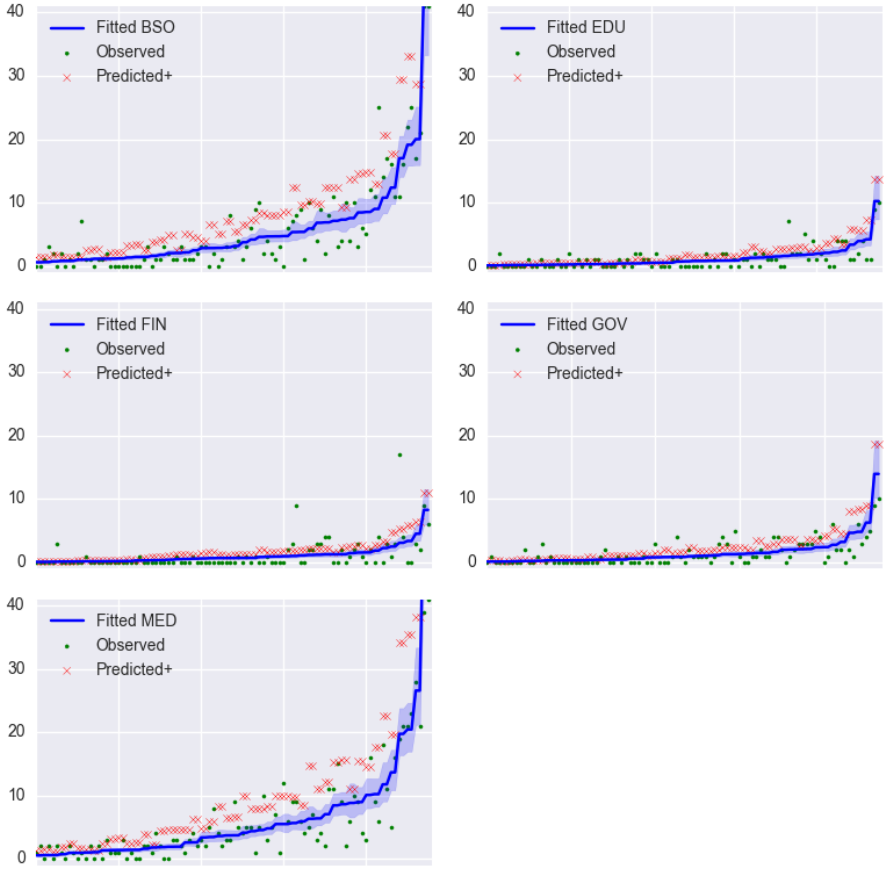


Figure 5.5 Prediction results by sector

Each point is a state; the x value represents the observed/fitted/or predicted+ number of breaches in that state; the number of breaches sorts states.

The prediction result indicates that 1.264 data breaches would have been publicly reported for 2015 if notification to credit reporting agencies had been mandatory, all Attorneys General had published notification letters on their websites and the risk-of-harm exemption did not exist.[90] An additional 46% of reported data breaches is generated by applying the first two provisions to states where they do not exist. In comparison, 17% is generated by the exclusion of the risk-of-harm exemption. This pre-

[90] If we instead set all states to allow the risk of harm analysis exemption, the number of breaches will be 1.005 more severe breaches.

dicted total is 483 breaches more than the actual 781 data breaches reported. In other words, the current patchwork of data breach notification laws in place in the U.S. hides from the public more than 500 data breaches per year.

5.6 Modelling time regression

Given that informing customers faster is another aim of DBNLs, we also model the *uninformed exposure time* (the time between a security breach and the firm's notification) and the *notification time* (the time the organisation needs to assess the situation after breach detection, to finalise the letter and to inform the customer and relevant parties). During both periods, customers are not aware of the risk they are exposed to and cannot undertake any defensive action.

This data is available in the 2014 notification letters from AG websites in four states. The histograms for both variables are presented in Figure 5.6, with an average of 44 days for the notification time and 102 days for the uninformed exposure time. We used a negative binomial regression to model how sector, state and breach cause influence these times.

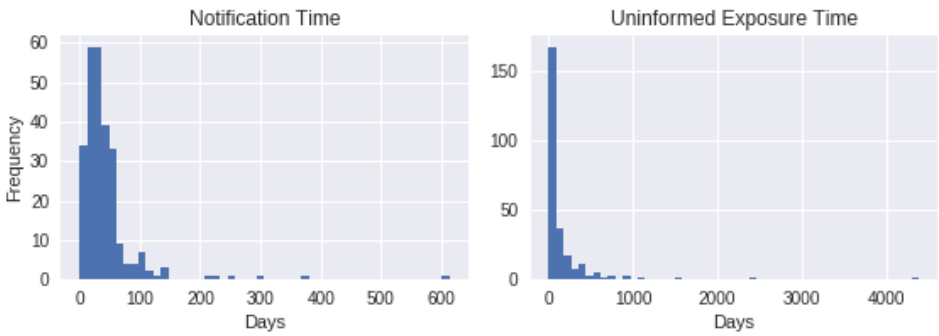


Figure 5.6 Histogram of notification (n) and uninformed exposure (ue) time, public AG dataset

Table 5.4 reports the results for the uninformed exposure time. According to these results, the financial sector detects breaches in about half the average time of 102 days across all industries and breach events, which may reflect the maturity of the sector in terms of information security (see Security Scorecard 2016). Compared to the baseline unintended disclosure time, hacking events take 71% longer to detect, and insider events

take five times longer. When the affected consumers reside in more than one state in our dataset, breaches take 30% longer to detect, possibly highlighting the negative effect of organisational complexity. Given its strong link with the industry and breach event, uninformed exposure time can generally serve as a proxy for competency.

For the notification time, no variable – state, sector or breach type – improves the intercept-only model based on the Akaike Information Criterion (AIC) or pseudo R^2. Thus, we can only say that the notification process on average takes 44 days (see appendix V for details).[91] However, we can conclude that no DBNL element in the four states of California, Maryland, New Hampshire and Vermont causes a notification to be sent faster.

Table 5.4 Uninformed exposure time regression Model

Variable	**Coef. (Std Error)**	**Incident Rate (95% CI)**
(Intercept)	4,629 (0,298) ***	**102 days** (59,9-195,9)
Hacking	0,537 (0,207) ***	**1,71** (1,19-2,44)
Physical	-0,588 (0,207) **	**0,65** (0,37-0,83)
Insider	1,594 (0,252) ***	**6,00** (3,62-8,21)
Unintended	*baseline*	--
BSO	-0,342 (0340)	0,71 (0,35-1,32)
FIN	-0,747 (0,352) **	**0,47** (0,23-0,92)
EDU	-0,268 (0,467)	0,77 (0,30-2,03)
MED	0,318 (0,351)	1,38 (0,67-2,60)
GOV	*Baseline*	--
Multistate	0,258 (0,142) *	**1,30** (0,97-1,74)

Negative binomial disp: 0,948. N=257. Deviance null/residual: 440/296. McFadden pseudo R^2: 0,04.

[91] Interestingly enough, five of the seven states that indicate a time frame for notification in U.S. indicate a limit of 45 days, in line with the notification time average. Specifically, Ohio, Rhode Island, Vermont, Washington and Wisconsin.

5.7 Conclusions

We modelled the impact of DBNL provisions on the number of known data breaches and breach notification times, while controlling for sector and state differences. We concluded that the data breaches that are publicly known are just the tip of the iceberg. The dimensions of what is visible and what is hidden below the surface are dependent on how DBNLs are designed. In this vein, we calculated the number of breaches that could be reported if certain provisions were uniformly adopted across the U.S.

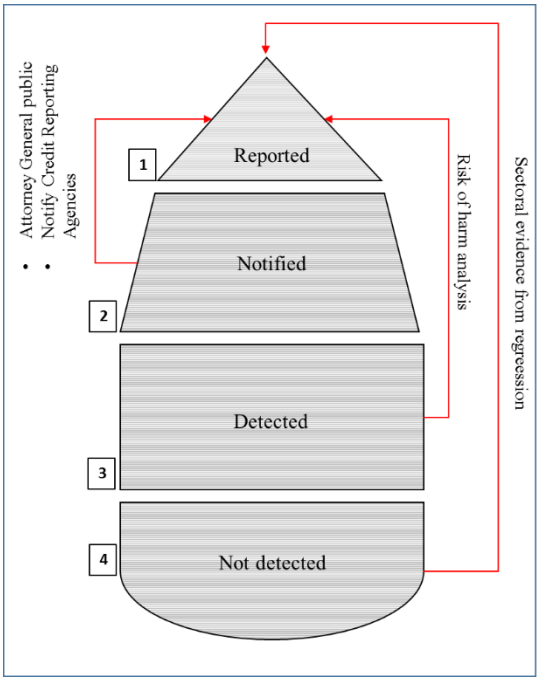


Figure 5.7 Increase in reported data breaches

Breaking down the iceberg structure, we estimated that (i) 46% more breaches would be reported because of the *inform credit agency* provision and the provision *notification publication by informed Attorneys General,* moving a portion of the breaches in block two to block one; (ii) 17% more breaches would become known from block three (detected, but not notified) as the effect of the elimination of the *risk-of-harm analysis provision* and (iii) an undefined percentage of undetected breaches could be identified from the sectoral results of the regression model.

By examining the uninformed exposure time, we also identified the breach causes that more than others represent an obstacle for a safer data security environment. For breaches in the categories of 'hacking' and 'insider' in particular the detection and notification timing is poor, dramatically eroding the utility of the notification for helping individuals to defend themselves against potential harm such as identity theft promptly.

In short, two core elements play an important role in bringing detected and notified breaches into the light: the inclusion of more actors in the notification flow and the publication by notified actors of the notifications. While the first factor requires a revision of the law and therefore a lengthy legal process, the second factor can be easily implemented in those states that already include the Attorney General in the notification flow. These authorities are well-positioned to foster the visibility of data breach notifications by publishing them, consequently supporting both DBNL objectives (sunlight as disinfectant and consumers' right to know).

We also noticed key sectoral differences in how breaches lead to public breach notifications or not. Certain sectors lag behind in terms of detection time, possibly as a result of weak security measures, and therefore lag behind in terms of the notification itself. For example, organisations in the 'retail and other business' sector report fewer breaches overall and more breaches caused by hacking. This finding strongly suggests shortfalls in detection and in notification for breaches with other causes.

The findings of our analysis raise two main implications for previous studies that have relied on datasets of publicly reported breaches: (1) the breach frequency has been underestimated, and (2) the comparisons of breach counts and magnitudes across different sectors are likely to be biased by systematic and substantial sector differences in whether and how breach events are reported. Indeed, uncertainty about actual breach frequency is clearly visible among the different existing datasets. Many previous studies relied on the PRC dataset, while we used the ITRC data. In 2014, the two databases exhibited a significant discrepancy in total figures: 330 breaches reported by the PRC versus 783 by the ITRC.

Both sources use a similar rationale when intercepting data breaches.[92] However, the ITRC also includes single-used data sources, as highlighted in annex 1. We consequently relied on this source, which is able to identify a larger number of breaches.

---

[92] ITRC: Each selected incident is required to have been reported to a state Attorney General's office or published by a credible media source, such as TV, radio, press, etc. The item will not be included at all if ITRC is not certain that the source is real and credible.
PRC: PRC's Chronology includes breaches reported through either government agencies or verifiable media sources.

The implications of our research do not apply to all previous research in the field. Romanosky et al. (2011), for example, focused solely on the date of implementation of DBNLs and its relation with identity theft. Event studies on the effect of a single data breach on the affected organisation's market values are also not relevant (Gordon et al. 2011, Acquisti et al. 2006, Cavusoglu et al. 2004, Campbel et al.2003). In addition, the implications are limited for studies that are confined to a single sector regulated by federal law (e.g., Kwon et al. 2015), in which the data are less impacted by sectoral differences.

Nevertheless, many studies would benefit from taking our findings into account, specifically the systematic underreporting bias in certain sectors (e.g., for BSO) and states (e.g., for states where Attorneys General do not publish breach notifications). Studies that explore trends using the PRC dataset are particularly vulnerable. This includes the work of Garrison et al. (2011), who presented a longitudinal analysis of data breaches focused on the analysis of time series of data breaches, of Edwards et al. (2015), who modelled breach frequency and Romanosky et al. (2014), who used DatalossDB[93] to identify the subset of data breaches that were publicly 'reported' and analysed which breaches generated litigation. In the last case, the use of the dataset in this research could enlarge the sample of reported breaches that are subsequently classified under non-litigated, federally litigated, or state litigated. In addition, the fact that a large number of breaches would be reported if Attorneys General publish them could result in a different distribution among those three classes.

Achieving more breaches reported is not a goal in itself; rather, it is a mechanism to improve the security of the Internet ecosystem by making the state of affairs transparent to the security community, policymakers and citizens. In the long run, careful monitoring is needed to determine whether these outcomes are indeed achieved. Will reported breaches become background noise, the inevitable consequence of a digitising society; will they generate increasing tangible negative consequences in terms of 'naming and shaming' or companies going after the custom-

[93] DataLossDB.org operated until mid-2015.

ers of breached competitors or will other hitherto unanticipated consequences emerge? The employment of DBNLs as a means to improve security will remain an important topic of study for the foreseeable future.

Chapter 6 More than a suspect: An investigation into the connection between data breaches, identity theft and data breach notification laws.

The final study in this dissertation focuses on the ultimate effect of data breaches, namely on identity theft. We investigate the relationship between data breaches and identity theft and study what effects DBNL enactment and revisions have on incidences of both data breaches and identity theft.

6.1 Introduction

Information technology enables the collection and storage of large amounts of personal data. While these activities provide unquestionable economic benefits, it has also proven impossible to keep personal data fully secure against criminal misuse. Surveys report that in 2017, identity thieves fraudulently obtained approximately $16,8 billion from 16,7 million American consumers (Javelin 2018). According to the same study, in the past six years, identity thieves have stolen over $106 billion from their victims. Having access to personally identifiable information[94] is a

[94] The U.S. government defined the term *"personally identifiable information"* in 2007 in a memorandum from the Executive Office of the President, Office of Management and Budget (OMB),[M-07-16 SUBJECT: Safeguarding Against and Responding to the Breach of Personally Identifiable Information FROM: Clay Johnson III, Deputy Director for Management (2007/05/22)] and that usage now appears in US standards such as the NIST Guide to Protecting the Confidentiality of Personally Identifiable Information (SP 800-122).["Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)" (PDF). Special Publication 800-122. NIST.]
The European Regulation (EU) 2016/679 (General Data Protection Regulation - GDPR) defines in its art. 4 personal data as any information relating to an identified or identifiable natural person being an identifiable natural person one who can be

prerequisite for perpetrating identity crime. Data breaches are a key source for this access (Garrison and Ncube 2011, Roberds et al. 2008).

California was the first state to enact a data breach notification law[95] (hereafter also DBNL), emphasising the potential criminal harm of identity theft as their main rationale for the *duty to notify* (Skinner 2003, Draper 2006, Bisogni 2016). Other U.S. states have since enacted DBNLs. In Europe, the General Data Protection Regulation, similarly recognises (in its preamble) that identity theft is a major risk when a data breach is not addressed in an appropriate and timely manner.

Despite these legal rationales, little research exists to date on the relationship between data breaches, identity theft, and the impact of DBNLs on related trends over time. It is clear that data breaches are numerous and increasing: the Identity Theft Resource Center (ITRC) reported 1.579 data breaches in the U.S. in 2017, an increase from 1.091 in 2016 and only 421 in 2011. Nevertheless, there is no definitive estimate of how many cases of identity theft have resulted from data breaches. In a small-scale effort, the U.S. Government Accountability Office (2007) examined 24 large data breaches between 2000 and 2005, and conclusively linked four of them to subsequent outbreaks of fraud. Romanosky et al. (2011) have done one of the few studies on the impact of DBNL on identity theft, measuring and estimating this effect using panel data (from 2002 to 2009) from the US Federal Trade Commission.

This chapter addresses this research gap by investigating the relationship between data breaches and identity theft in more depth, including the impact of DBNL enactments (and revisions) on these incidents (using empirical data and Bayesian modelling). We collected incident data on breaches and identity theft over a 13-year time span (2005-2017) in the U.S.. The databases we used included those of the ITRC (data breaches), Privacy Rights Clearinghouse (data breaches), Consumer Sentinel Network (identity theft) and Perkins Coie (DBNLs).

---

identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

[95] California Civil Code § 1729.98 enacted in 2003.

Our analysis reveals that the correlation between data breaches and identity theft is driven in large part by the size of the state. Enacting a DBNL still slightly reduces rates of identity theft in the enacting states, while Attorneys General publishing breach notifications helps the broader security community learn about the breaches.

6.2 Background

In this section, we explore general aspects related to identity theft, data breaches and the laws adopted in the U.S. and Europe to control these two issues. It is generally acknowledged that identity theft can take many forms. The U.S. Government Accountability Office, in its report to congressional requesters dated July 2007,[96] divided identity theft into two categories: existing-account fraud and unauthorised creation of new accounts. Examples of these categories are, respectively, the misuse of credit card numbers (credit card information is stolen) and opening a credit card account in someone else's name (personal information is stolen). The Identity Theft Resource Center provides another classification; ITRC identifies five categories of identity theft:

- Financial identity theft: when the imposter uses another individual's personal identifying information, primarily a Social Security number, to establish new credit lines;
- Criminal identity theft: when a criminal gives another person's personal identifying information, in place of his or her own, to law enforcement;
- Identity cloning: when the imposter uses the victim's information to establish a new life. He or she actually lives and works in the victim's identity;
- Medical identity theft: use of someone else's data in order to obtain medical services or goods;
- Commercial identity theft: similar to financial identity theft except the victim is a commercial entity.[97]

[96] GAO-07-737, a report to congressional requesters.

[97] Di Ciccio (2014) indicates also Synthetic Identity Theft (use of different subjects' personal data combined in order to create a new identity), Ghosting (creation of a new identity, different from the original one by exploiting the data of a deceased person), Cyber Bullying (Impersonation: impersonation in a different person, by

In all of the cases mentioned above, data, mostly personal identifiable information, in the hands of thieves is a prerequisite to perpetrate the crime. Therefore the means to access this information plays a central role. Data breaches appear to be the primary source for accessing personal information and thereby the primary source of identity theft (Garrison and Ncube 2011).

However, we do not have a definitive estimate of how many cases of identity theft have resulted from data breaches. This type of estimation was the goal of the U.S. Government Accountability Office (2007) in examining 24 data breaches between 2000 and 2005 in which large amounts of data were compromised. The GAO conclusively linked four massive breaches to subsequent outbreaks of fraud. However, the sample was very limited; thus, their findings cannot be generalised. An additional study providing evidence that a significant proportion of identity theft can be attributed to inadequately secured commercial data is the one conducted by Gordon et al. (2007). The study examined 274 cases of identity theft prosecuted by the Secret Service from 2000-2006 and found that 50% of the cases resulted from compromised data at a business.

The nature of a causal connection between security breaches and concrete harm suffered by consumers is not always easy to determine. A data breach does not necessarily result in identity theft, as data may be stolen without being used for fraudulent purposes. Moreover, identity theft can occur without a data breach. In consumer surveys, victims of identity theft who know how their information was stolen commonly attribute their loss to channels that are not linked to technology, such as lost or stolen wallets (43% of cases reported in Javelin 2009), fraud by acquaintances (13%) or stolen mail (3%). Only 11% of cases are reported to be linked to data breaches and 11% to online methods.

It is evident from the existing literature that most of the analysis performed on data breaches and identity theft have been carried out in the U.S., which is a pioneer country in terms of data breach notification laws. DBNLs in the U.S. were promulgated under the main objective of reducing identity theft. However, the measurement of this specific effect has

means of cellular phones or web services, with the purpose of sending messages with objectionable contents).

been the subject of limited research. The work of Romanosky et al. (2011) is the only empirical study measuring this effect; using panel data from the U.S. Federal Trade Commission, the researchers estimated the impact of data breach disclosure laws on identity theft from 2002 to 2009. They found that the adoption of data breach disclosure laws reduces identity theft caused by data breaches by 6,1% on average. Our study not only updates this analysis with a wider time span, but also extends it to the effect of specific law provisions and legal revisions not only on identity theft but also more directly on data breaches. Moreover, we test different statistical models to identify the strongest model for such estimation. We conclude that the Bayesian model is more adequate.

In order to lay proper foundations for a U.S.-Europe comparison, it is important to highlight that data breach notification laws not only attempt to fulfil a specific purpose, the mitigation of identity theft, but also confront conflicting goals of consumer protection and corporate compliance-cost minimisation (Burdon 2011). In contrast, comprehensive information privacy legal frameworks, such as that of Europe, have an extensive aim of ensuring legal protections related to the protection of personal information.[98] Information privacy laws set minimum standards that relate to fair information practices and provide individuals with a series of limited rights of involvement in the process of personal information exchange.[99] The relation between laws protecting privacy and laws addressing concerns about identity theft is complex and sometimes antagonistic. For example, Towle (2003: 261–264) described the dilemma as follows: customers argue both for and against more privacy, creating tension under identity theft statutes and attribution procedures. Vendors and organisations generally find themselves between a rock and a hard place. They are asked to increasingly respect more privacy in not forcing customers to provide extensive identification data before entering into a transaction, but also less privacy in ensuring that no one is violating their customer's identities.

---

[98] Information privacy law is based on the notion that individuals have rights relating to control over their personal information (Kang 1998), or at least, have rights pertaining to who can access their personal information (Gavinson 1980) or a combination of both (Moor 2010).

[99] See Privacy Rights Clearinghouse, Why Privacy, https://www.privacyrights.org/why-privacy-0 (last visited 13/1/2019)

Identity theft and data breaches have become a relevant issue in the EU not only for individual member states, but also in the broader EU agenda. The main result is the General Data Protection Regulation 2016/679, which entered into force on May 24, 2016 and applied, after a two-year transition period, from May 25, 2018. Contrary to its predecessor, Directive 95/46/EC,[100] the GDPR equally applies directly to every citizen and organisation falling within the scope of European Union law. Hence, the GDPR is well placed to become a significant piece of legislation. The connection between identity theft and data breaches is clearly defined in the preamble of the GDPR (EU) 2016/679, point (85):

> *A personal data breach may, if not addressed in an appropriate and timely manner, result in physical, material or non-material damage to natural persons such as loss of control over their personal data or limitation of their rights, discrimination,* ***identity theft*** *or fraud, financial loss, unauthorised reversal of pseudonymisation, damage to reputation, loss of confidentiality of personal data protected by professional secrecy or any other significant economic or social disadvantage to the natural person concerned.*

In the European context, the situation is partially different from the U.S., where a general right of information privacy does not protect consumers. Indeed, the breach notice is not associated with any right to compensation.[101] Also, the General Data Protection Regulation extends the notification duty to all data controllers. Apart from these factors, the GDPR mainly follows the approach of the U.S. DBNLs, with one important difference: the regulatory environment that it creates includes a much-improved enforcement mechanism for data protection violations compared

[100] Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31 (Data Protection Directive).

[101] Winn 2009: "Attempts to establish a right to damages following receipt of a security breach notice through class action lawsuits have generally only succeeded in clarifying the degree to which no such right exists, although many businesses suffering breaches have chosen on a voluntary basis to provide their customers with credit monitoring services to reduce the risk of harm from identity theft."

to the U.S. scheme. This also means that companies reporting a breach may face substantive fines by the regulators, in addition to any possible action by the individuals affected.[102] In the European context, while class actions are not typically found in European jurisdictions, fines can be levied by the data protection authorities without necessarily proving a concrete loss for individuals.

The GDPR will, therefore, not only reaffirm the general right to information privacy, but also provide an enforcement mechanism following the evolution of privacy regulations. (This issue is further analysed in section 6.5).

6.3 Research Method

The remainder of this chapter investigates the relationship between DBNL enactment and revisions, reported data breaches and identity theft. We start by investigating the causal connection between data breaches and identity theft, with the aim of identifying the strength of correlation among the two variables. We then move to the effects that DBNL enactment and revision have on both, also considering the level of notification publicity that these laws may introduce. As illustrated in Figure 6.1, for this analysis we take into account other important predictors related to state wealth and infrastructure, to digital threats (for data breaches), and to crime and breached records (for identity theft).

.

[102] In the US, there is no general tort of privacy violation, however individuals affected by a data breach can sue if they can prove that they suffered economic harm through the negligence of the breached entity. The availability of class actions in the US legal system gives this opportunity.

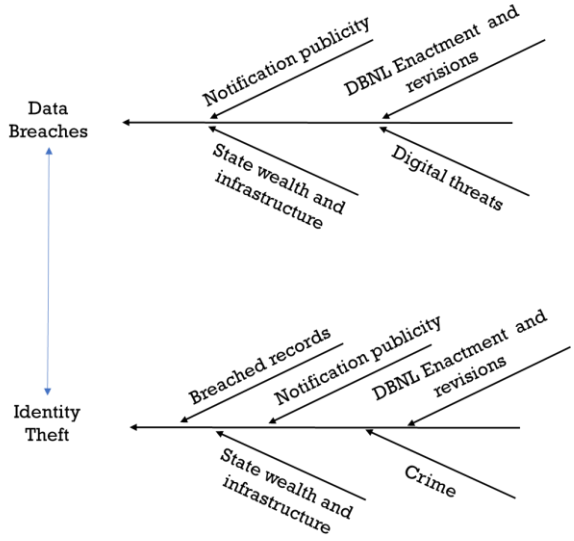


Figure 6.1 Causal diagram

The effects of DBNLs can be studied thanks to specific datasets. The introduction of legislation to address the threat of data breaches in the U.S. has indeed fostered a number of databases which gather information about data breaches and identity theft at state level.

Table 6.1 Summary statistics

Years (2005-2017)	13
States	50
N. of observations (state value per year)	650

	Variables	Total	min	max	average
control	Data breaches (DB) 2005-2017	8.171	133	1.557	626
	per state per year	*8.171*	-	*231*	*13*
	Identity theft (IDT) 2005-2017	3.879.919	238.107	440.068	295.455
	per state per year	*3.879.919*	*156*	*69.795*	*5.972*
	Employer Firms (DB model offset)	6.073.017	16.952	740.303	121.460
	Population (IDT model offset)	320.903.064	586.555	38.993.940	6.418.061
	Notification publicity: n. of states publishing	10	1 (2010)	10 (2018)	n.a.
	Records (x1.000)	998.882	12.471	179.465	76.837
predictors	GDP per capita		33.658	75.852	49.109
	property crime per capita		0,0178	0,0391	0,0286
	DBNL enactment	*see table 6.2*			
	DBNL revisions	*see table 6.2*			

We combine the following sources as summarised in Table 6.1:

- The data breaches come from the Identity Theft Resource Center (ITRC)[103] and Privacy Rights Clearinghouse (PRC)[104] databases. The main sources of the two databases are the same (i.e., media, State AGs, US Dpt. Health and Humanities). For the timespan 2013-2017 we used ITRC data, as they include a higher number of data breaches for the analysed period (4.851 vs. 3.546), retrieving a higher number of data breaches from media. For the time span 2005-2012 ITRC data were only available at aggregated level, so we used PRC data. The similar number of data breaches collected by the two sources in this time span (only c. 5% of difference) suggests that potential data heterogeneity between datasets is very limited.
- The identity thefts data come from the Consumer Sentinel Network database.[105] These statistics are consumer reported and collected by the Federal Trade Commission for each state. (They reflect reports by individuals once they discover a theft, and not an automated check and balance by other agencies such as consumer credit bureaus).

Our dataset comprises 650 total records, with each record containing the number of data breaches and identity theft in one of the 50 states from 2005 to 2017. We add a number of common predictor variables to this dataset, including the population of states,[106] number of firms per state[107] and GDP per capita.[108] (These variables are used to normalise, as predictors and as controls; further explanations are provided for each use.)

---

[103] https://www.idtheftcenter.org/data-breaches/ reporting 9.774 data breaches in the time span 2005-2018

[104] https://www.privacyrights.org/data-breaches/ reporting 9.002 data breaches in the time span 2005-2018

[105] https://www.ftc.gov/enforcement/consumer-sentinel-network/reports

[106] https://www.ftc.gov/enforcement/consumer-sentinel-network/reports

[107] US Census Bureau. Number of Firms, Number of Establishments, Employment, and Annual Payroll by Enterprise Employment Size for the United States and States, Totals: 2016. https://www.census.gov/data/tables/2016/econ/susb/2016-susb-annual.html

[108] US Bureau of Economic Analysis. Last updated: May 1, 2019-- new statistics for 2018; revised statistics for 2010-2017.

Finally, we added the date when DBNL laws came into effect for each state, and dates of their subsequent revision/amendment, based on data from Perkins Coie.[109] California was the first U.S. state to enact a DBNL (in 2003); 33 states enacted their DBNLs before 2015; Alabama and South Dakota were among the last (enacting in 2018). The majority of states have revised or amended their DBNLs a number of times since enactment, as provided in Table 6.2.[110]

Table 6.2 States enacting DBNLs and subsequent revisions

For example, as of 31 December 2018, of the 16 DBNLs enacted in 2006, 5 had no revision, 9 had one revision and 2 had two revisions.

Year of enactment	N. of enacting States	States with 0 revision	States with 1 revision	States with 2 revisions	States with 3 revisions	States with 4 revisions
2003	1	-	-	-	-	1
2005	10	1	5	3		1
2006	16	5	9	2	-	-
2007	9	3	4	-	2	-
2008	5	2	2	1	-	-
2009	4	2	1	1	-	-
2011	1	1	-	-	-	-
2014	1	-	1	-	-	-
2017	1	1	-	-	-	-
2018	2	2	-	-	-	-
Total	**50**	**17**	**22**	**7**	**2**	**2**

We begin our analysis with descriptive statistics, followed by a difference-in-differences (DiD) analysis for the effects of DBNLs (Angrist &

[109] https://www.perkinscoie.com/. Given that we consolidated our dataset at yearly level, we considered enactments and revisions in the last quarter of a year for the subsequent year.

[110] The average time for the first revision (or amendment) is 6 years and 2 months. Among the states, 10 went through a second revision, with an average time (from the previous change) of 3 years and 3 months; 4 went through a third revision (within 2 years and 2 months); and 2 through a fourth one.

Pischke, 2015). DiD models are, in short, not suitable for our analysis, as they generate high standard error, and cannot reliably estimate the enactment effect. This is because DiD requires that we assume parallel trends for states before (or only after) enactment, which does not hold upon inspection. A further problem is that DiD treats the year and state intercepts (dummies, or fixed-effects) as entirely independent of each other. This conceptualisation ignores the fact that external events may impact data breaches across all states.[111] Nevertheless, as DiD models are used by a number of prior studies involving data breaches, we included them as a baseline.[112]

Our main analysis employed multi-level (also known as hierarchical or random-effects) Bayesian regression models.[113] The detailed model specifications are presented in the following Findings section (where we also explain the variables and interpret the results). The motivation for using Bayesian multi-level modelling is that it can more precisely estimate the effects of a common intervention (DBNL) while allowing for differences among states by pooling together the varying intercepts (see McElreath 2016, ch. 12) for each state (and similarly pooling the varying intercepts for each year). The two classic approaches to modelling interventions across multiple states, which are opposites on a spectrum, are to specify the model with only the intervention variable and no additional dummies for the states, or to specify the model with the intervention and add a unique dummy (or intercept) for each state. On the one hand, the first option (no unique intercepts) ignores structural differences among states that may affect the observation, and results in very poor model fit and estimates. The second option (per state intercept), on the other hand, assumes that the states are completely independent from each other, and may lead to the intercept capturing too much of the variance (and noise) in the data. Pooling the intercepts is the more realistic and accurate middle ground that Bayesian multi-level modelling allows. In essence, the states have unique intercepts, but those intercepts are kept as close to

[111] Some prior work has attempted to resolve the fact that the year and state dummies are not completely independent in this instance using *robust and cluster-corrected* error terms (e.g., Romanosky, Telang and Acquisti 2011). However, the Bayesian multi-level method that we present next is a more flexible and robust approach.

[112] The DiD models are estimated using non-Bayesian MLE methods.

[113] The multi-level refers to stacking of distributions in the model definitions, due to the pooling of the intercepts.

each other as possible in the fitting process. Given the increase in computational power, Bayesian models are increasingly recommended for problems with inherent clusters.

We follow Bayesian inference and reporting procedures, as recommended by McElreath (2016) and Kruschke (2015). Our *Jupyter notebooks,* which make use of *PyStan*[114] and *ArviZ*[115] packages, in addition to the classic Python analysis toolkits, are available upon request. [116]

6.4 Findings

6.4.1 Correlation driven by size

Figure 6.2 plots data breach and identity theft trends from 2005 to 2017.[117] The figure depicts clear and parallel growing trends until 2015.

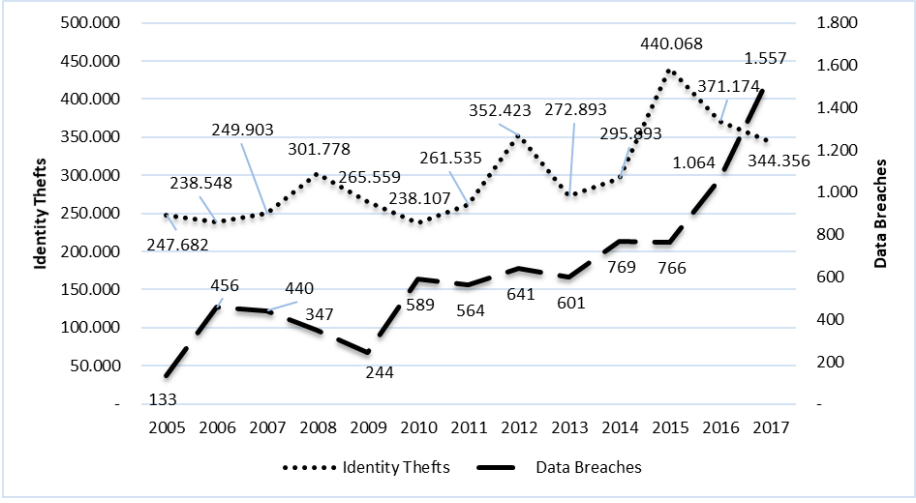


Figure 6.2 U.S. trends in data breaches and identity theft

[114] PyStan (https://github.com/stan-dev/pystan) provides a Python interface to Stan, a package for Bayesian inference using the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo.

[115] ArviZ (https://arviz-devs.github.io/arviz/) is a Python package for exploratory analysis of Bayesian models.

[116] We also contemplated the use of Bayesian multi-level ARMAX models. However, given the fact that data breaches and identity thefts, have clear time trends, and given that the random variation is less important than the overall magnitude of these incidents, ARMAX models are not informative for our study.

[117] See appendix I for a comparison with breached record counts.

The causes of this growth can be traced to many factors, such as growing digitalisation, and the increasing ease with which financial transactions are conducted electronically and processes are managed digitally. This development enlarges the opportunity for criminals to act in the digital arena. In addition, the ability to monetise personal information has increased as an incentive to perpetrate data breaches. At the same time, data breaches have been more frequently publicised over time, with higher numbers of Attorneys General publishing the notifications received from breached organisations and therefore increasing the number of breaches fed into relevant databases (see Bisogni, Asghari and Van Eeten 2017). As such, the number is not necessarily growing due solely to more frequent data breaches. However, it could also be growing due solely to the increased reporting of data breaches through public channels.

Figure 6.2 also illustrates that the number of reported incidences of identity theft follows a more unstable trend: the phenomenon has generally been growing over the 13 years, but with positive and negative peaks. For example, the number of reported identity thefts in 2017 (344.346) was slightly lower than the value reported in 2013 (352.423). The differences in the two-time trends indicate the existence of data breaches that do not lead to identity theft, and incidences of identity theft that are not a result of data breaches.

As the scatterplots illustrated in Figure 6.3—left reveal, the correlation between data breaches and identity theft is significant (Spearman correlation coefficient of 0,78).[118] However, the correlation *weakens if we normalise the variable 'identity theft'* with a state's population (Spearman coefficient 0,58; see Figure 6.3—right).[119] The correlation *further weakens*

[118] The correlation is much stronger (0,89) if we take into consideration only the subset of states (and years), where Attorneys General report notifications received from breached organizations, shrinking the number of data breaches not known to the public and therefore reducing the gap between current data breaches and reported ones (Bisogni, Asghari and Van Eeten 2017). As of 2019, 22 states require notifications to the Attorney General, but only 10 states publish details of the events and the notification letters. These ten states include California, Indiana, Maine, Maryland, Montana, New Hampshire, Oregon, Vermont, Washington, and Wisconsin.

[119] This correlation is similarly stronger when considering only states with AG reporting (0.75).

if we also normalise 'data breaches' by the number of firms in a state (0,29). In other words, the *strong correlation is driven by the size of the state, and once we control for size, the unexplained variance increases*. (This finding is in line with the fact that the causes of identity theft are not limited to data breaches, and that not all breaches are publicly known.)

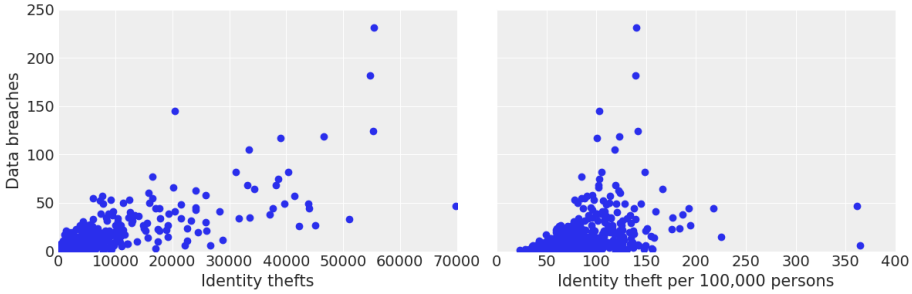


Figure 6.3 Scatter plots between data breaches and identity thefts (left); and data breaches and identity theft per 100.000 persons (right).

The Spearman correlation coefficient is 0,78 for the left and 0,58 for the right scatter plot.

However, another factor may also be at play: DBNLs may have different effects on the two variables. We investigate this scenario in the following sections.

6.4.2 The Impact of DBNL on data breaches

We employed regression analysis to model the impact of DBNL on data breaches.

Difference in Differences. As explained in the Methods section, we first used the *difference-in-differences (DiD*) method. DiD is a well-established method that *under the right assumptions* mimics an experimental design using observational data (Angrist & Pischke 2014). The basic requirement is a longitudinal and cross-sectional dataset, with treatments applied at various points in time. The key assumption is that the control and treatment outcomes move *in parallel* in the absence of treatment.

In our case, this means assuming that data breach trends run in parallel across states. As a number of prior studies which examined the impacts

of data breaches have used DiD (e.g., Kwon & Johnson 2015, Choi & Johnson 2017), we *temporarily* accept this assumption. The regression formula is expressed as follows:

$$Breaches_{s,y} = \delta_{DD} Enacted_{s,y} + \sum_{k} \beta_k \, State_k + \sum_{j} \gamma_j \, Year_j$$

The formula includes the enactment effect (δ_{DD}) and add dummies to control for the difference by state (β_k) and by year (γ_j).

The left part of Figure 6.4 depicts the density plot for the dependent variable, data breaches[120]. This variable may be fitted with a negative binomial curve. This choice is conceptually sound because data breaches are rare,[121] discrete events, and counts of such events are best modelled using the negative binomial distribution (see among others: Edwards, Hofmeyr and Forest 2015).

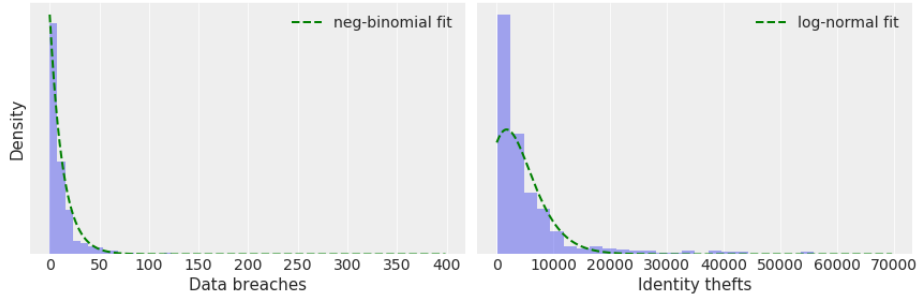


Figure 6.4 Density plots for the dependent variables (data breaches, identity thefts, and the normalised versions)

The dashed lines represent fitted distributions: negative-binomial (also known as the gamma-Poisson) for data breaches; log-normal for id thefts.

The regression results are summarised in Table 6.3. The `enactment effect` (*hasdbnl*) is $e^{0.02 \pm 0.40}$ (coefficients must be interpreted as $e^{coef \pm 2stderr}$

[120] We *normalise* breaches in our models (that is given its correlation with the number of firms in a state) using a regression *offset* and keep the dependent variable as breaches.

[121] Rare considering the number of data breaches reported relative to the millions of firms processing data.

due to the GLM specification). This results in 68% to 152% change in the odds of a breach being reported after enactment.

Table 6.3 DiD model results

Breaches as the dependent var.; uses negative binomial regression; see Appendix for full results

Generalized Linear Model Regression Results

Dep. Variable:	**breaches**	No. Observations:	650
Model:	GLM	Df Residuals:	587
Model Family:	NegativeBinomial	Df Model:	62
Link Function:	log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1929.6

	coef	std err	z	P>|z|	[0.025	0.975]
hasdbnl	**0.0173**	**0.203**	0.085	0.932	-0.381	0.415
Intercept	-1.0892	0.408	-2.672	0.008	-1.888	-0.290
States dummies		*(see appendix)*				
Years dummies		*(see appendix)*				

The high standard error means that we cannot reliably estimate the enactment effect using DiD. In fact, the assumption of parallel trends also does not hold if we plot the trends for states before (or only after) enactment. Another problem with the DiD specification is that it assumes that the year and state intercepts (dummies, or fixed-effects) are completely independent of each other. This assumption ignores the fact that certain external effect may impact data breaches across all states.[122]

[122] Some prior work has attempted to resolve the fact that the year and state dummies are not completely independent in this instance using *robust and cluster-corrected* error terms (e.g., Romanosky, Telang and Acquisti 2011). However, the Bayesian multi-level method that we present next is a more flexible and robust approach.

Bayesian multi-level model. Bayesian multi-level modelling effectively resolves the deficits of DiD. We built a model in which each state (and year) has its own intercept, but the intercepts are *pooled* together by assuming they come from a common underlying distribution (with tight prior variance).[123] We specifically opted for a *multi-level Poisson model*, because it yields more efficient results (that is better model fits) than the negative binomial distribution *when combined with pooled varying intercepts*.[124]

We modelled the regression using *Stan* platform and programming language. The complete model code can be found in the Appendix. The key lines of the model are the following:

```
// priors
alpha    ~ normal(0, 10);
betas    ~ normal(0, 1);
a_years  ~ normal(0, sigma_y);
a_states ~ normal(0, sigma_s);
sigma_y  ~ cauchy(0, 1);
sigma_s  ~ cauchy(0, 1);

// linear relation
mu = intercept + a_years + a_states + betas*X + offset;
Y  ~ poisson_log(mu);
```

[123] McElreath (2016) refers to this as 'partial pooling', which is in between 'no pooling' (assuming each state acts fully independent of the other) and 'complete pooling' (ignoring differences among states and having only a common intercept). Partial pooling strikes a balance by allowing some state differences while assuming there still is a common pattern. These models are also referred to as *random effects models*.

[124] A Poisson distribution is (also) a distribution of counts events, but it requires the sample mean and variance to be equal. This isn't the case if we look at all the data breaches together, but it holds if we assume each state to have its *own* rate. The *mixture* of Poisson distributions leads to the negative-binomial (or gamma-Poisson) distribution. It has the advantage of not needing negative binomial's *dispersion* parameter, which makes the model estimations more efficient. This better fit can also be tested with *the widely applicable information criteria (WAIC)*, which indeed holds in this case. Also see McElreath (2016) pp 350-383.

In the model, *Y* is the observed data (breaches per state/year); *X* represents the regression predictors (e.g., whether DBNL has been enacted, DBNL provisions, and control variables) and *mu* is the Poisson rate. *mu* is modelled using a linear relationship between a common intercept, varying year and state intercepts (*a_years*, *a_states*),[125] the predictor coefficients (*betas*) and an *offset*[126] that limits the Poisson rate (here, the number of firms in the state). The year and state intercepts have *weakly informative priors,* in this case, a shared normal distribution and a tight sigma. We plugged in the following predictors:

- *b_enacted* indicates whether a state in a given year enacts a DBNL;
- *b_agp* indicates whether a state's Attorney General publishes breach notification letters;
- *b_revised* captures whether a state has revised (or amended) its DBNL in a given year;
- *b_ytrend* captures the yearly trend of data breaches;
- *b_gdp_pcap* captures the yearly trend of GDP per capita.

We then run the model:[127]

$$\begin{aligned} Breaches_{s,y} \sim\ & \alpha + \beta_{enacted}.Enacted_{s,y} + \beta_{agp}.AGP_{s,y} \\ & + \beta_{revised}.Revised_{s,y} + \beta_{y\_trend}.year \\ & + \beta_{gdp\_pcap}.GDP\_pcap_{s,y} + \alpha_{state\_pooled} \\ & + \alpha_{year\_pooled} + log(Firms_{s,y}) \end{aligned}$$

[125] Another common approach is to use a varying intercept *per observation*. This of course risks over-fitting the model; a point that is also reflected in a worse WAIC score here.

[126] Using an *offset* is the recommended method for setting limits on Poisson rates (here number of firms). An offset basically fixes the coefficient for the limiting factor to 1. If we use firms as a predictor instead, the model will estimate its coefficient still close to 1, but the estimates will be less efficient (and take much longer to compute).

[127] A model specification with only a single enactment effect as the predictor—basically the same as the DiD specification—yields similar results for that parameter as the full model explained here; the only difference is that the common and pooled intercepts have a larger spread since less of the variance is captured by the other predictors.

The Bayesian models converge well[128]. In Bayesian analysis, the *posterior distribution* of the parameters provides the same results as the coefficient estimates in non-Bayesian regression analysis. The resulting posterior distributions are shown in Figure 6.5. The mean of each parameter, and the 94% highest posterior density (HPD) interval, also known as the credible interval, are also marked. The HPD visualises the parameter uncertainty.[129] The variance of the varying intercepts is considerably small, which indicates successful pooling (i.e., the states differ, but not too much).[130]

Posterior *b_enacted* reveals an approximate 11% (±12%) increase[131] in reported breaches after a state enacts a DBNL. In other words, the model is not entirely certain about the enactment effect; passing a DBNL may have no effect (-1%), or some increase (23%).

The uncertainty around enactment effect (i.e., the coefficient's spread) across states may be explained, foremost, by the fact that the provisions of the DBNL matter, a point we shall return to in the next paragraph. An alternative (or compounding) explanation might be that the effect of enacting DBNLs decreases over time, as more states enact them, states that enact a DBNL later will experience less of an impact, given that larger firms active across multiple states will have already adopted breach notification duties (i.e., have procedures and systems in place for it), a phenomenon known as the 'California Effect' (Vogel 1995, Vogel and Kagan 2004).[132] We will return to this point in the identity theft model.

---

[128] See the Appendix for more convergence details; a posterior predictive plot is presented later in this section as well.

[129] The HPD functions somewhat similar to the standard errors in non-Bayesian regression results. The 94% interval is chosen on purpose by the ArviZ package so as not to be confused with the 95% frequentist significance levels. If one selects a different credible interval (e.g., 80%), then the reported parameter range becomes smaller.

[130] The unique year and state intercepts are presented in the Appendix.

[131] The coefficients for a Poisson models need to be interpreted as *change in the odds* by $e^{\text{mean } (\pm\text{range})}$. Here this is $e^{0.10\ (\pm 0.11)}$, which translates to a breach rate *change* of 99% to 123%, or 11% (±12%) increase.

[132] Due to its large market share, and preference for strict consumer and environmental regulations, California often leads with regulations which all firms active in California must implement. For larger firms, once they have implemented these

AGP indicates whether a state's Attorney General publishes breach notification letters,[133] which we know from prior research plays an essential role in the public's knowledge of a breach having occurred (Bisogni, Asghari, Van Eeten 2017). The effect of *b_agp* is quite strong: the number of (known) breaches in a state increases on average by 28% (±12%) if it enacts its DBNL with the additional condition that the AG be notified of any breach, and the AG subsequently publicises breaches.

The effect of *b_revised* captures whether a state has revised (or amended) its DBNL in a given year. As previously noted, approximately two-thirds of states followed their DBNL enactment with a revision (or amendment).[134] We hypothesised that in the absence of enforcement[135], revisions might help maintain a vigilant environment among actors involved in the notification process. On average there are 4% (±7%) more breaches reported in years that DBNLs are revised (excluding revisions that lead to the AG publicising the notifications, as that is captured by *b_agp*). As the uncertainty around this parameter's estimates are high, much cannot be said about it.[136]

---

changes in their operations, they might prefer to streamline their operations and de facto implement it in other jurisdictions as well.

[133] This is set only once the AG starts publishing these letters. NH, MD, and VT where the first states to do so, publishing the letters since 2010; CA followed suit in 2012; In 2017, this number increased to 9 states.

[134] In the context of constitution and law, an amendment is a change or addition to an existing law. A revision, on the other hand, is through re-examination of the entire law. This is done to make changes or alterations in the law.

[135] Concerning DBNLs the connection between legal sanctions and notification remains indirect and, in practice, weak (Schafer 2017).

[136] We also tried an alternative manner of operationalizing DBNL revisions, by creating b_dbnl_version variable, which we defined as the square root of the number of times this law has changed (0 = no DBNL, 1 = DBNL enacted, 1.41 = one revision/amendment, and so on. If we use this variable in place of all existing law variables (b_enacted, b_agp, b_revised), we find that every unit increase yields a 10% (±8%) increase in breaches. If, however, we add the variable to the model next to the b_agp variable, its effect disappears (while other parameter coefficients stay approximately the same). In other words, a key success factor for a DBNL is that the regulator is placed in the notification loop, and it publicizes the notifications (whether as part of the original law or added in a revision).

From *b_ytrend,* we see that on average the number of (known) data breaches increases every year by 14% (±7%), even after controlling for changes to the regulatory environment (DBNLs). This increase has two potential sources. On the one hand, the growing number of digital threats that are not countered by proper measures produces more breaches. On the other hand, as organisations become better equipped to detect breaches, more reports are produced. Unpacking these two sources requires more data (and maybe of interest for future research).

Finally, GDP per capita is a common state-level control that is a proxy for wealth and infrastructure, among other variables.[137] The number of data breaches increases by approximately 5% (±7%) for every $9.600 increase in GDP per capita[138]. This effect may simply reflect that companies in more prosperous areas are more attractive targets for hackers.

[137] We exclude some other common controls that are either correlated with GDP_pcap, as they can lead to multicollinearity; Or are unrelated to firm behavior, such as crime (which is about the population of a state, while breaches happen over state lines), since they can lead to inefficient estimates. (Multicollinearity and inefficient estimates can mask the actual effects of interest).

[138] The GDP per capita variable has been centered and standardized. Thus, the parameter's value is the increase caused by one standard deviation change in GDP per capita (from the mean GDP per capita), which is approximately $9.600.

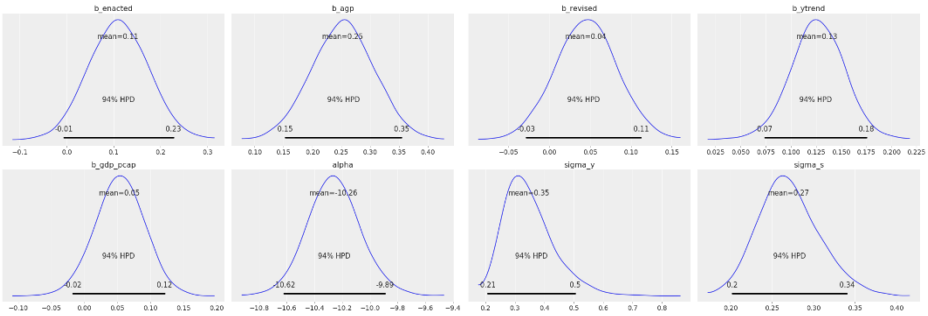


Figure 6.5 Posterior distributions for the multi-level data breach model

The highest 94% posterior densities are marked; The betas are the predictors; other parameters are the common intercept, and the variances of the two varying/random intercepts; All betas need to be interpreted as e^{mean} due to the Poisson log link function. See appendix for the unique intercepts.

6.4.3 The Impact of DBNL on Identity Theft

We followed a similar reasoning process to model the impacts of DBNLs on identity theft –using, once again, a multi-level Bayesian model with pooled varying intercepts.

A key computational difference between identity theft and data breaches is the choice of distribution for the dependent variable (see Figure 6.3 right). Empirically, a *log-normal* distribution offers the best fit, not a gamma-Poisson distribution. Conceptually, a log-normal distribution points to an underlying multiplicative process instead of an additive one (Limpert, Stahel & Abbt, 2001). This process can be understood by the fact that a fraudster will (typically) target multiple victims within a single fraud campaign, which explains identity thefts strong correlation with the state population (data breaches, on the other hand, are rarer and more independent).

The key lines of the Stan model are as follows; as in the previous model, we used weakly informed priors (the additional sigma parameter captures the overall variance for the log normal distribution):

```
mu = intercept + a_years + a_states + X*betas + offset;
Y  ~ lognormal(mu, sigma);
```

This time, the population size is used as the offset. The predictors we use are as follows:

- *b_enacted* indicates whether a state in a given year enacts a DBNL;
- *b_agp* indicates whether a state's Attorney General publishes breach notification letters;
- *b_revised* captures whether a state has revised (or amended) its DBNL in a given year;
- *b_records_pcap* is the number of records breached per capita (estimated for state/year);
- *b_pcrime_pcap* captures the yearly trend of property crime per capita;
- *b_gdp_pcap* captures the yearly trend of GDP per capita.

The model includes two new predictors, *breached records per capita*, and *property crime per capita*[139], both which are both expected to increase identity theft. We estimate the number of breached data records per state by summing up the total data records breached in all reported breaches across the U.S. in a year (an average of 76 million records) and dividing this total by each state's population. (The rationale is that the sum of records breached per year is strongly driven by the so-called 'mega' breaches—breaches that impact millions of customers. These customers are likely spread over all the U.S. states). Property crime we include since it can be a cause of identity theft; and also, the socio-economic factors that lead to a rise of property crime in a region may also lead to increased identity theft. We exclude the yearly trend variable in this model, since identity theft does not show a strong trend in the logarithmic form[140]. Thus, placing the predictors in the model, it produces the following:

[139] Property crime results from the sum of burglary, larceny and motor vehicle theft. Data source: Summary (SRS) Data with Estimates at https://crime-data-explorer.fr.cloud.gov/downloads-and-docs. Another possible control is internet penetration (e.g., internet users per capita). This variable is highly correlated with GDP per capita; and substituting it in the model does not affect any of the reported predictors.

[140] Additionally, breach records have a strong yearly component to them, and having both predictors would mask the other's effect. If we use yearly trend (in place of records), we find a slight annual growth of 4% (±2%).

$$IdentityTheft_{s,y} \sim \alpha + \beta_{enacted}.Enacted_{s,y} + \beta_{agp}.AGP_{s,y} + \beta_{revised}.Revised_{s,y} + \beta_{records_{pcap}}.Records_{pcap_{s,y}} + \beta_{crime_{pcap}}.PCrime_{pcap_{s,y}} + \beta_{gdp_{pcap}}.GDP_{pcap_{s,y}} + \alpha_{state\_pooled} + \alpha_{year\_pooled} + log(Population_{s,y})$$

Figure 6.6 presents the posterior distributions of this model, which again converges well.[141]

Adopting a DBNL results in a 2,5% (±3%) decrease in identity theft. While the direction of this effect is negative as expected, the decrease is quite small, and the credible interval crosses zero, making it also uncertain. *What is interesting is that the effect size is less than half the 6,1% that Romanosky et al. (2011) reported for the period 2002-2009.* This contrast may be evidence of the California effect—larger firms active across multiple states may have already adopted breach notification duties and practices in all states by choice, thus decreasing the effects of DBNL enactment by later states[142]. Another explanation for this small effect size is that data breaches are only a portion of identity theft, e.g., Javelin Report (2009) estimated that data breaches are the source of 11% of identity theft. In other words, the decrease in breach-related identity theft is several-fold larger. If all incidents of identity theft were driven by data breaches, the magnitude of the identity theft decrease would be about 22,7%, applying the 2,5% decrease to an 11% subset. Nonetheless, it also highlights that *being notified of a breach does not guarantee one can stop the resulting identity theft in time*.[143]

The credible interval for *b_revised* and *b_agp* spreads widely around zero, indicating no clear effect. The fact that the AG publicising breaches

[141] The posterior predictive plot for this model can be found at the end of the previous section.

[142] To make this more concrete: in 2005 (the start of our dataset), eight states had passed DBNLs, and these states held approximately a third of all U.S. firms; In other words, a third of U.S. firms were already subject to some DBNL in that year. By 2008, this had increased to 40 states and 84% of all U.S. firms.

[143] With more precise data on identity theft causes, which currently are not available, this idea can be further explored (future work).

does not further reduce identity theft is, paradoxically, a positive finding: it suggests that firms notify breach-affected customers, as required by law, irrespective of the publicity.[144]

The posterior for *b_records_pcap* suggests that a one per cent increase in the number of breached records equates to a 1% (±1%) increase in identity theft.[145] Finally, *b_pcrime_pcap* is a strong predictor of identity theft: a one percentage point increase in property crime per capita equates to an 11% (±5%) increase in the identity theft rate. (As the mean property crime per capita is 2,8%, a 1% increase is substantial).

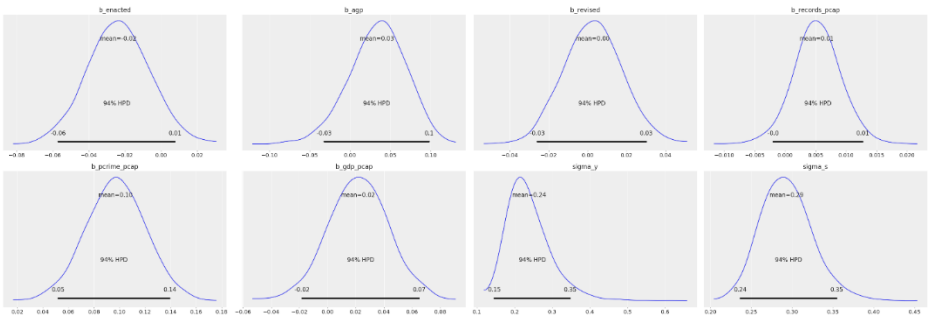


Figure 6.6 Posterior distributions for the multi-level identity theft model

The 94% highest posterior densities are marked. Betas must be interpreted as e^{mean}.

Counterfactual Plots for Identity Theft. We can use 'counterfactual plots' to visualize and better understand how the three key predictors (b_dbnl, b_records_pcap, b_pcrime_pcap) impact identity theft. This is shown in Figure 6.7: the model's predicted outcome (identity theft) is shown for imaginary states with varying degrees of breached records and property crime, with and without a DBNL. The plots make it clear that any benefits that come from enacting a DBNL (in terms of decrease in

[144] Note that using the *b_dbnl_version* variable (explained in a footnote of the data breach model) in place of the three separate law variables yields a similar effect as *b_enacted* alone.

[145] The credible interval for this predictor also touches zero, reflecting uncertainty in the effect. This uncertainty is in part because we use a rough estimate for the number of breached records per state (as the actual number of data subjects affected in each state aren't reported). Examining the counterfactual plot for this parameter makes this point evident (e.g., the alignment of the dots).

identity theft) are by far outweighed by a significant increase in the number of breached records (e.g., resulting from a mega breach). In other words, the drop from the black line to the red line is small, compared to the overall upward slope that shows the effect of additional breached records on identity theft.

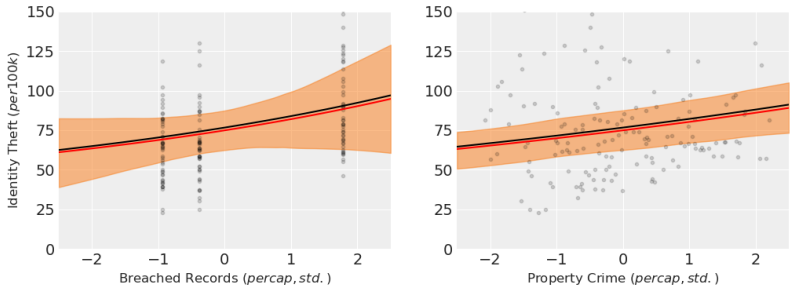


Figure 6.7 Counterfactual plots for states with different breached records (left) and property crime (right).

The two solid lines are whether a DBNL is enacted or not (with enactment being the lower line). The shaded area is the 94% HPD; The dots are observed data (plotted for 2005, 2011, and 2017); Counterfactual variables are standardized.

Posterior predictive checks. It is customary in Bayesian statistics to check, as an additional robustness measure, whether the posterior predictions of a model mimic the observed data with reasonable accuracy. The plots of Figure 6.8 (respectively for data breaches-left and identity theft-right) depicts 100 *simulations* of breaches overtime for California, New York, Virginia, and New Hampshire versus the actual trends for these states. The simulations are in light grey, and the observed trends in solid colours. As visible in the figure, the predictions and observations are reasonably well matched.[146]

[146] These four states were chosen simply because they have very different baseline levels. A more classic posterior predictive plot of *y* to *y_hat,* which includes all the states and years, can be found in the Appendix.

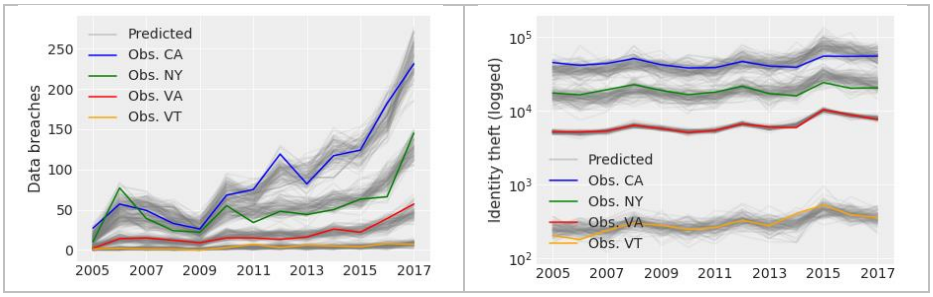


Figure 6.8 Posterior predictive plots for the multi-level data breach (left) and identity theft (right) models

They show observed vs. predicted over time for select states. The grey shades indicate 100 simulations, and the solid colours indicate observed values.

6.5 Conclusions

We analysed the correlation between identity theft and data breaches, and found that the size of a state primarily drives the correlation between the two variables. That is, the correlation decreases (but does not disappear) when we control for population or number of firms.

We next used multi-level Bayesian modelling to examine the effects of DBNLs on data breaches and identity theft over 15 years in the U.S.. We observed an increase in reported (and known) breaches rates after DBNL enactment, and a considerable increase if the Attorney General publicises the notifications. DBNL enactment also slightly reduces identity theft rates, and if we consider that data breaches are not the only source of identity theft, the decrease is considerable. These findings are very relevant for the European context, particularly in the present period of implementation of the GDPR. Moreover, in addition to relevance to other regions, the collection of identity theft statistics is also important to better measure the impact of legislation with aims to mitigate the consequences of data breaches.

Currently, it is impossible to perform a similar analysis in Europe, considering the current inadequate state of data collection, and the level of accessibility to the limited information collected on data breaches and identity theft. This is a relevant weakness of the European system. On one side, since GDPR implementation, we have a high societal cost of reporting and notifying to specific bodies, including supervising authorities. On the other side, the relevant information collected remains too often

'behind the gate' of DPOs. Companies are requested to make incredible efforts in notifying, but not enough learning is coming from fulfilling this requirement, as researchers have not access to data that are also not adequately collected and/or organised.

There is another additional barrier for performing a similar exercise in Europe. Currently, there is no common way for EU Member States to internally identify identity theft and no procedure to report these cases at the European level centrally. Member States define, record and subsequently report identity theft in diverse ways, generating differences in the number of cases from one country to the other. We will explore more in detail what is the state of the art of collecting in the conclusions.

Whether the focus is information disclosure or regulation, a central question about data breach notification policy in the 21st century is whether we have appropriately designed institutions and processes to foster and monitor the desired outcomes. There is little question that the current mix of public policies does not always live up to expectations. Information disclosure policy, such as the GDPR or the U.S. DBNL, have played a fruitful role in the mix of contemporary policy and regulations. However, much can be done to improve program effectiveness, particularly in ensuring that the information collected is easily accessible, understandable and meaningful in terms of real public and private risks faced.

Chapter 7 Conclusions

This chapter offers a review of the findings in the preceding chapters, and connects them to the dissertation's central question and objective. The dissertation focused on three themes for strengthening cybersecurity: the significance of data breaches, the use of disclosure policies and the role of enforcement. These themes were combined in the following question, which was examined via a literature review, four empirical studies and two theoretical papers.

> *What are the effects of the provisions of data breach notification laws on (1) communications issued by breached organisations to their customers; (2) the timing of breach detection and reaction; (3) the number of data breaches reported and (4) the volume of identity theft stemming from data breaches.*

Section 7.1 reflects on the analysis of data breaches and related findings and on the commonalities among the studies conducted in this dissertation. Section 7.2 then explores the policy implications for Europe to address the final research question: *'What are useful lessons learned for EU regulators on managing and monitoring the implementation of the GDPR?'.* Section 7.3 reports on the main limitations of my research and suggests future study. Finally, section 7.4 reflects on broader themes related to data breaches which were not investigated in this dissertation.

7.1 Summary of the empirical findings

The findings and implications of this dissertation offer diverse contributions to the field of economics of information security. The studies have also received significant attention outside of academia. For example, two of the studies[147] were quoted in the *Economic Report of the President – the White House (2019)* and a third[148] in the *Federal Reserve Bank of Kansas*

[147] Bisogni (2016) and Bisogni, Asghari and van Eeten (2017)

[148] Bisogni (2015)

City Economic Review (2016), possibly increasing U.S. government insight into its data breach notification legislation.

We instituted some methodological innovations in the studies to answer the primary research and policy questions. We constructed unique databases, such as a database of one year of notification letters analysed at the paragraph level, and we correlated time series of 13 years regarding data breaches, identity theft and implementation and revision of DBNLs.

Our research contributes to solving contemporary challenges in data and information security in an interdisciplinary manner. The findings on communication styles in notifying data breaches, timing for breach identification and notification, hidden breaches and generated identity theft offer significant insights for further policy discussions and the development of data breach notification laws.

We analysed breaches affecting companies which collect and process personal information. These companies may maintain weak security practices or lack proper staff behavioural controls or technical solutions, such as data encryption, anti-malware solutions and software updates. Such vulnerabilities increase the risk of events generating data breaches.[149]

Our analyses affirmed that the market alone might not fully drive organisations to increase their security practices. Firms' ex ante incentives to invest in proper security measures are misaligned, as the harms from data breaches are not entirely borne by the firm, but passed on to the affected individuals (Anderson and Moore, 2006). Thus, firms are likely to consider only their private costs and to thereby underinvest in security.

The clear growing trend in both data breaches and identity theft (see Figure 6.1) indicates that despite the enforcement of DBNLs by regula-

[149] unintended disclosure (sensitive information posted publicly on a website, mishandled or sent to the wrong party via email, fax or mail), physical loss (lost, discarded or stolen non electronic records, portable or stationary device), insider (someone with legitimate access intentionally breaches information - such as an employee or contractor), hacking and malware (electronic entry by an outside party, malware and spyware), payment card fraud (fraud involving debit and credit cards that is not accomplished via hacking), Unknown or other (all other cases).

tors, the effects of such legislation are no match for the progress of related threats. The reasons for an increasing trend in reported breaches include the growing number of digital threats that are not countered by proper measures and organisations becoming better equipped to detect breaches.

Figure 7.1 visualises how we investigated all of these aspects. Across the chapters, we examined different aspects of notifications (including timing and communication style) and their relationship with the different DB causes and DB effects.

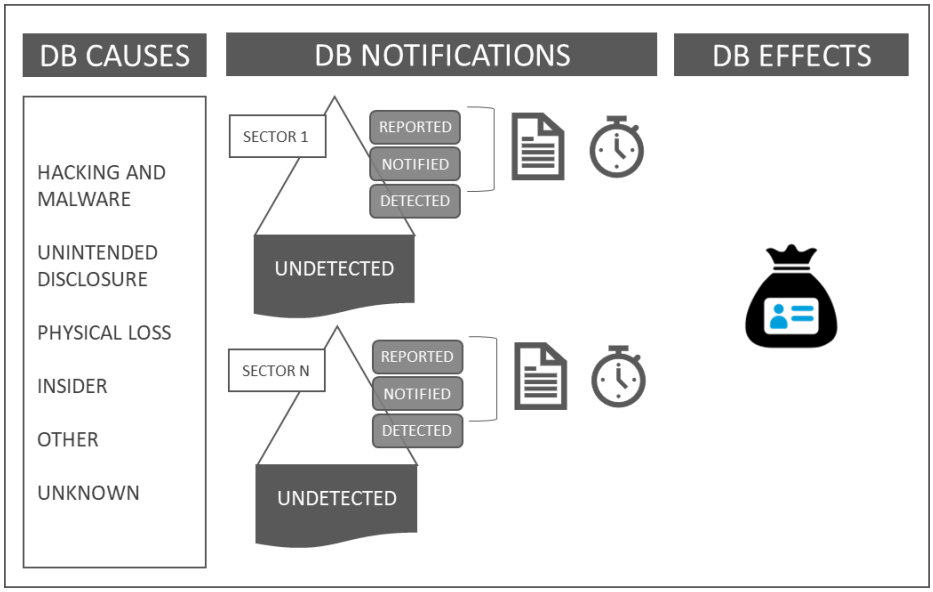


Figure 7.1 DBNL investigation field

Each chapter answered part of the above research question, and also relevant sub-questions. In chapter 2, we introduced the information asymmetry issue. We demonstrated that the lack of complete information on cyberattacks might lead to insufficient awareness of the related risk, causing more significant economic damages than expected by operators. In order to improve cybersecurity, policymakers can establish an incentive framework through disclosure policies. We then focus on DBNL as disclosure policy to fight identity theft. The theoretical evaluation model presented in chapter 2 disentangled factors associated with data breach notification laws to address the sub-question 1: *What are the different elements to consider in evaluating data breach notification laws?* The

model guided our review of state of the art in researching data breaches and related regulations.

In chapter 3, we focused on sub-question 2: *What are the core elements of consumer notification letters and how do company decisions on what to include and how to express the message define specific letter types?* We studied the communication styles employed in data breach notifications, classifying notifications in our sample of 213 notifications into six different letter types. The six letter types vary in the degree to which they activate customers' reactions to the breach and thereby in organisational customer support. The data highlight how organisations make careful choices when assessing the higher costs associated with letter types that clearly disclose facts and company responsibility. In the cases where a company could be easily identified as ultimately responsible for the data breach and thereby subject to legal actions, the use of no worries letters in order to minimise the problem was present in a high percentage. In essence, 30,36% of the cases for unintended disclosure, 20,69% of data breaches generated by insiders and 20% of cases of physical loss were associated with a no-worries letter. No worries letter emphasises the minor nature of the risk generated by the event, reassuring the affected customers. When the breach was generated by hacking or malware, the distance strategy was the most commonly used through the use of routine letters (29,21%). Routine letters present the event as a consequence of an unavoidable and relatively common risk. This finding affirms the notion that hacking and malware tend to be presented as such.

In chapter 4, we addressed sub-question 3: *How different are the choices organisations make in terms of if they notify, what they notify, and when they notify?* The investigation included a more comprehensive analysis of a full year of data breach notifications in California, Maryland, New Hampshire, Vermont.
Focusing on the element 'if organisations notify', the number of retrieved letters confirms the presence and relevance of underreporting. The number of the analysed letters in 2014 (445) represents 56,83% of the 783 total cases collected in the U.S.. Nevertheless, the four states represent only 14,37% of the total number of firms in the U.S. according to the

2012 Economic Census statistics,[150] and 14,98% of residents according to the 2010 Census.[151] In addition, based on the letters analysed in the four states and the sectors in which the breaches took place, we can identify that approximately 15% of notifications derive from local retail business, service or medical centres acting locally, in which case we can assume that the place of the breach and the residency of the affected individuals coincide. We can safely assume that similar events happen across the U.S., with a similar percentage of firms per sector affected by local data breaches which impact the residents of only one state.

Focusing on what is notified, we confirmed that organisations clearly exploit the fact that many state statutes do not yet mandate minimum information for the content of breach notifications, providing them with significant elements of discretion. Companies often use such elements in order to limit their eventual reputational damage or the short-term costs posed by the activation and management of communication channels (e.g., call centres, higher rate of activated credit monitoring). Thus, an organisation's exercise of discretion may not always support customers' conscious reactions to a breach, and the results of such 'flexibility' can produce suboptimal effects for society.

Finally, we analysed the time between detection and notification, and discovered significant delays by organisations in identifying breaches. The average breach detection time was 113 days, with significant variation across breach types. Data breaches generated by insiders and hacking require over six months for their identification. Such delays prove that notifications arrive to customers already late even if sent on the date of the discovery. In comparison, organisations discover data breaches due to physical loss and unintended disclosure more rapidly (in 18 and 78 days, respectively). Prompt notifications can better address these breaches. We also calculated the average time needed by organisations to notify, finding an average of 133 days between the breach and consumers' awareness about it (uninformed exposure time). This delay clearly jeopardises the achievement of DBNL goals. Here the results revealed differences according to affected sectors. The time between the detection of the breach and notification to regulators and consumers is the shortest in the financial sector. In addition, the retail sector is the most

[150] http://www.census.gov/econ/census/

[151] Population Distribution and Change: 2000 to 2010. 2010 Census Briefs. 2010

reactive once the breach was detected in cases of hacking or malware and insider.

In chapter 5, we answered the sub-questions 4 and 5 looking at the law provisions that could better support the public disclosure of data breaches. By mandating notification to *credit reporting agencies* and requiring the free accessible publication of the notifications received by Attorneys General, 46% more notified breaches would be publicly reported (sub-question 4: *What effects do specific DBNL provisions have on reported data breaches?*). An additional 17% more breaches would become known as the result of eliminating *risk-of-harm exemptions*[152], and an undefined percentage of undetected breaches could be identified from the sectoral results of the regression model in chapter 5 (sub-question 5: *How large is the portion of data breaches we are unaware of?*). We also analysed the events causing data breaches in detail. The results affirmed the importance of the human factor and of behavioural controls, with unintended disclosure, physical loss and insider breaches playing a central role in enabling access to personal information (accounting for more than 50% of the total breaches). We also revealed interesting patterns related to DB causes by sector. In retail and other business, hacking comprises a larger proportion of all breaches than in other sectors. This may indicate a lower level of network security in the sector or the fact that they are a more attractive target for profit-driven criminals. Besides, unintended disclosures are less prevalent than in other sectors, which points to either underreporting or less vigilant monitoring and process controls.
In contrast, unintended disclosures cause a high proportion of breaches in the governmental sector, highlighting either weak personal data handling processes or ineffective monitoring of these processes, or a combination of both. Insider attacks comprise the lowest proportion of breaches, which is aligned with the fact that more security background checks are performed in this sector than in others. In the medical sector, physical losses are the dominant source, possibly reflecting the fact that data physically travels during service delivery in this sector more than

[152] This provision requires a breached organization to notify customers only if the organization determines that the breach constitutes a reasonable likelihood of harm to the customer.

others. This may explain the high proportion of insider theft as well, as many professionals must have access to the data. In the finance sector, physical loss is the least frequent source of breach, possibly due to the more extensive use of digital capabilities than in the rest of the economy.

In chapter 6, we further examined the relation between DBNL, data breaches and identity theft (sub-question 6: *What effects do DBNL enactment and revisions have on incidences of data breaches and identity theft?*). Identity theft is a concerning effect of data breaches; indeed, we found a strong correlation among the two phenomena (Spearman correlation coefficient of 0,78). However, this correlation can also be traced to confounding factors: the number of firms and people in each state. In fact, the correlation *weakens if we normalise 'identity theft'* with a state's population and *'data breaches'* by the number of firms in a state. In other words, the *strong correlation is driven by the size of the state, and once we control for size, the unexplained variance increases*. This finding is in line with the fact that the causes of identity theft are not limited to data breaches, and that not all breaches are publicly reported. We also observed an increase in reported (and known) breach rates after DBNL enactment, and a considerable increase if the Attorney General publicises the notifications. DBNL enactment also reduces identity theft rates, which adds credibility to the positive effect of DBNLs in revealing unreported data breaches and helping fight identity theft.

7.2 Implications for European data breach notification policies

The studies included in this dissertation focus on the U.S., the pioneering country in terms of disclosure policies related to data breaches. This focus enabled us to highlight the successes and failures of DBNLs and areas for improvement in the future. The results also generate interesting implications for Europe, where the GDPR and related notification duties have only recently been implemented. The final research sub-question is, therefore, the following: *What are useful lessons learned for EU regulators on managing and monitoring the implementation of the GDPR?*

7.2.1 The GDPR and its data breach notification obligation

As the normative framework plays a relevant role in how organisations behave, we focus in the next paragraphs on lessons learned from the U.S. experience regarding the content of data breach notifications, timing, notified actors, penalties and the importance of data collection. We first briefly present the main characteristics of the GDPR.

The GDPR introduces new notification obligations for organisations regarding personal data breaches. Under the GDPR, both data protection authorities and data subjects must be notified without undue delay, or if this is not possible, the corresponding information has to be conveyed through public communication. Data protection authorities must be notified of any data breach. In contrast, data subjects must be notified only if the breach is likely to result in a high risk to the rights and freedoms of natural persons.

The GDPR represents the core of the European regime for the regulation of data protection and the flow of personal information. The requirements apply to the personal data of any individual EU resident that is used by any person or organisation located anywhere in the world. The GDPR applies to almost every type of personal information about an individual, including name, address, computer IP address, bank details, social networking content and medical information.

Under the GDPR, there are two categories of data users: 'controllers' and 'processors'. Data controllers determine the purposes for which and the means by which personal data is processed, and data processors are those processing data on behalf of a data controller. Both controllers and processors can be organisations that are managing the personal information of EU residents, whether they are located inside the EU or not. The GDPR obliges both the controller and the data processor to implement appropriate measures for the security of data processing. The GDPR clarifies the processor's and controller's obligations (Article 32) *to implement appropriate technical and organisational measures to ensure a*

level of security appropriate to the risk.[153] Specifically, following an evaluation of the privacy risks, the controller and the processor must take the necessary measures to protect personal data against accidental or unlawful destruction or accidental loss and to prevent any unlawful forms of processing, particularly any unauthorised disclosure, dissemination or access, or alteration of personal data.

The GDPR also introduces the obligation of both the controller and the processor to provide notifications of personal data breaches (Articles 33 and 34):

> *In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent[154], unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay.*

The data processor is similarly obliged to alert and inform the controller without undue delay after becoming aware of a personal data breach. The controller must document any personal data breaches, and this documentation must comprise the relevant facts, the effects of the breach and the remedial actions taken.

The GDPR also obliges the controller to report the personal data breach to the data subject when it is likely to harmfully affect the protection of his or her personal data or privacy. This notification must be done *without undue delay, using clear and plain language* (Article 34). However, the communication of a personal data breach to the data subject is not re-

[153] Among others: the pseudonymisation and encryption of personal data; the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services; the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident; a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing. (art. 32)

[154] in accordance with Article 55

quired if the controller has taken proper protection measures or implemented subsequent actions such that the personal data remains unintelligible, and the privacy risks are thereby not likely to materialise.

In the case of non-compliance with the notification obligation, data protection authorities (DPAs) are granted the power to impose an *administrative fine of €10.000.000 or 2% of the undertakings turnover, whichever is higher* (Art. 83, comma 4). The fine can be imposed when the data controller conceals a data breach or does not notify in due time.

To meet GDPR notification requirements, organisations dealing with European personal data need to introduce processes that enable them to react quickly to breaches. As the case of U.S. DBNLs demonstrates, organisations must consider how promptly notifying data subjects may be hard. Internal reporting processes should be in place to support communication to the authorities and data subjects. This includes processes that enable processors to notify controllers.

The requirements related to the notification's content must also be taken into account, such as the disclosed nature of the breach and the description of its likely consequences. In addition, controllers should consider the proper way to implement personal data security, because the notification obligation to data subjects does not take effect if appropriate protection measures have already been put into practice (e.g., encryption). Implementing such measures may demand changes in companies' current information systems.

7.2.2 Lessons from the U.S. data breach notification laws

The European landscape is very similar to the American one. Therefore lessons learned in the U.S. may be useful for Europe, even though the GDPR, contrary to the U.S. DBNLs, is not inspired by the sunlight-as-disinfectant principle (Ranger 2007).[155] We list below several valuable considerations based on the analysis performed in the U.S., which suggest that Europe can benefit from specific provisions at European and country level activating 'name and shame' reputational sanctions.

[155] As discussed earlier, this happens thanks to Attorneys General Offices publishing notification letters sent by breached organizations to consumers on their websites publicly.

Content of the letter

According to the GDPR, notifications *should describe the nature of the personal data breach as well as recommendations for the natural person concerned to mitigate potential adverse effects* (Recital 86). Some U.S. DBNLs require only that the notification is made, without providing any indication of its contents. Fifteen U.S. states provide guidance on what information to include in breach notifications (i.e., the 'content' of the notification).[156] The presence of mandatory elements to be included in the notification letter resulted fundamental to ensure that the notification goal of safeguarding breached individuals is achieved. The results of the first empirical study demonstrate that the choices that a company makes on the content are decisive in fostering a prompt customer reaction against identity theft and in shaping the relations between customers and the organisation. The results revealed interesting patterns related to included elements and the manner of expression of the letter related to the type of breach event. Among others, our analysis of indirect vs. direct arrangement patterns[157] demonstrated that a direct approach is used in the lowest percentages (53,33% vs. 47,67% indirect) in cases of hacking or malware. In such cases, there is likely no urgency to capture the attention of the reader in order to foster a reaction as the event has happened more than three months before the notification, according to the average timing. In cases of unintended disclosure, the use of a direct approach is consistently higher (69,31%). This finding suggests that companies may consciously decide to use a direct approach when they feel it is useful given the short detection time. At the same time, they may more frequently opt for an indirect approach when they are aware that it is already too late for consumers to protect themselves against the consequences of the breach.

Without a clear indication of the mandatory content of notification letters, Europe is likely to face a similar dynamic. The type of event will drive the type of formal response consumers receive from breached organisations, and a prompt consumer reaction regardless of the data breach

[156] See figure 4.2. for complete overview

[157] The first starts with an explanation, delivers the bad news and finally closes with an expression of goodwill. The latter presents the bad news, provides an explanation and also closes with a statement of goodwill. starts with an explanation, delivers the bad news and finally closes with an expression of goodwill.

cause will not be ensured. A stricter regulation of the content of notifications is necessary to prevent firms from minimising the actual risk through assuring full transparency of the incident description, providing clear recommendations to customers and fostering interaction with them (which represents an additional cost for companies). More specifically, the European rules should mandate a clear description of the breach incident, including specification of the breached personal data[158] and the number of residents affected by the breach in order to allow consumers to self-evaluate the size of the breach. These three elements were not always present in the analysed letters, but they are essential to enable customers to self-evaluate the severity of the breach. The rules should also require clear recommendations to the affected customers to perform the necessary actions to reduce breach-related risks, such as to review bank and credit card statements carefully. Finally, breached organisations should foster interaction with affected consumers in their letter by highlighting the possibility to receive customer support and clarifying any unclear aspects of the notification or breach.

Timing

The specification of the dates of the breach detection and of the breach itself is essential to foster consumers' reaction. The timing analysis (chapter 4) suggests that the first objective of DBNLs, the right to know, is not adequately fulfilled in the U.S., with an average of 133 days from the data breach to its notification. For certain sectors, the average is as high as 214 days (educational institutions), or even 258 days for insider breaches. The timing, therefore, poorly matches individuals' need to defend themselves against potential identity theft promptly. That the timing directly contradicts the aim of DBNL provisions highlights an aspect of possible criticality also for Europe. In the U.S., most states require notification without unreasonable delay. Eighteen states include a specific deadline for notifying affected individuals, two require a 30-day notice, two require a 60-day notice, one has a 90-day limit, one has a 15-day no-

[158] According to GDPR, Art.4 (1), personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

tice for medical information, one requires notice seven business days after law enforcement review and the remaining 11 states require a 45-day notice.

The GDPR requires the controller to notify the data subjects 'without undue delay'. As per Preamble Recitals 85–88 of the GDPR, *a notifiable breach must be reported to the relevant supervisory authority without undue delay and within 72 hours of discovery*. The GDPR recognises that it will often be impossible to investigate a breach fully within that time period and allows notifications to provide information in phases. If all the information cannot be provided within 72 hours, the reasons for the delay must be provided in the breach notification.

The regression analysis of notification time in chapter 5 indicates that companies from four U.S. states require 44 days on average to notify consumers. The European limit of three days to notify the central authority thus seems very challenging for organisations, if not outright unrealistic. A delay is indeed allowed for compelling reasons, but we suggest the inclusion at the Member-State level of additional time limits according to the different types of breach events in order to regulate motivated exceptions (i.e., where reasons for the delay accompany the late notification). More specifically, it could be possible to apply additional time limits to the motivated cases according to the breach cause.

From our analyses of timing and content, we conclude that it is essential for individuals to have knowledge of the actual magnitude of the breach and of the timing of breach detection. In fact, even if notification time is respected (72 hours in Europe, 45 days in most U.S. states where it is defined), there is a significant difference if the detection took place on the same day of the breach or four months later (as seems to happen in the U.S. according to our analysis). We, therefore, recommend that specification of the breach and detection dates be mandatory in the notifications made by breached organisations towards consumers and relevant authorities. Firstly, this information can enable an appropriate reaction from customers and authorities. Secondly, the analysis of such information enables the study of sectoral dynamics generated by the different types of events, in order to support better prevention of and response to data breaches. In the U.S., organisations belonging to certain sectors are

significantly slower in reacting after the breach discovery. Relevant differences in the breach detection capability in various industries and in various areas of Europe should, therefore, be taken into consideration to limit the effect of externalities.

Notified actors

Our research explored the question of underreporting and proposed a possible option to increase reporting based on the number of notified actors. In order to reinforce the role of information disclosure against misaligned incentives and information asymmetries, the visibility of notifications should be extended as much as possible. We concluded that states in which Attorneys General are already in the notification communication flow could contribute by making the notifications publicly available. Together with the provision of informing credit agencies, this requirement would also support the second goal of DBNLs: to provide sunlight as disinfectant.

European Supervisory Authorities could similarly foster the achievement of this second goal by providing public access to the list of companies that were obligated to report breaches. As of today, the degree of publicity for small breaches, which alone cannot reach the media, is strongly driven by the capacity and willingness of the individual whose data were breached to share such information after receiving the notification.

In order to understand the scale and the consequence of disclosure by the supervisory authorities, we refer to our third empirical study in chapter 5. There we examined the available breach statistics to model the impact of DBNL provisions on the number of known data breaches and breach notification times, while controlling for sector and state differences. We concluded that the data breaches that are publicly reported are just the tip of the iceberg. The dimensions of what is visible and what is hidden below the surface are dependent on how DBNLs are designed. Breaking down the iceberg structure, we estimated that 46% more notified breaches would be publicly reported as a result of implementing the *inform credit agency* provision and the provision *notification publication by informed Attorneys General*.

The lesson learned for Europe is therefore clear. In order to fully exploit the effect of the notification obligation to a supervisory authority, the received notifications should enter the public domain.

The European regulation does not provide for communication to credit agencies as in some states in the U.S.. More in general, another element which is missing in the GDPR is post-breach consumer protection. The GDPR clearly focuses on pre-emptive data protection, neglecting requirements such as credit-freeze or credit-monitoring services (both included in the U.S. DBNLs). Instead of a credit freeze and credit monitoring, individuals in the EU can expect data authorities to enforce 72-hour breach notifications and to impose hefty fines for data privacy breaches. If authorities are unable to enforce 72-hour notification, similar-credit monitoring services should be considered.

The European regulation creates an additional challenge in that the three-day limit exists for reporting to national authorities, not to affected individuals. As the national authorities do not (yet) make the notifications public, companies have the incentive to 'over-inform' the authority as a matter of caution, knowing that there will be little reputational damage for doing so. In order to illustrate this point, the number of breaches reported in the Netherlands alone in 2018 was 20.881 (Dutch DPA, 2019), which is higher than our upper-bound estimate for the entire U.S. and more than 16 times the data breaches recorded in the U.S. in the same year (1.244 according to the ITRC, 2018).

These extra notifications impose an administrative burden on the notifying firms and the regulator, without any clear security benefits. As a solution, we suggest that the national supervisory authorities publish breach notifications after a grace period, similar to some U.S. Attorneys Generals. In this way, the private administrative burden can at least generate some social benefits through supporting more informed decisions by individuals and organisations regarding suppliers and business partners. Such decisions can also take into account a company's past 'security tracks' (i.e., how well the company performed in terms of suffered data breaches).

Penalties

The GDPR provides for administrative fines for infringement of the regulation, and criminal penalties are also foreseen. Each supervisory authority has the power to impose administrative fines and identify the upper limits. In the U.S., each state law may allow for civil and criminal penalties as well, but they differ from state to state[159]. For example, in Alabama, any covered entity or third-party agent that knowingly fails to comply with notification requirements is liable for a penalty up to $2.000 per violation, not to exceed $500.000 per breach. A penalty not to exceed $5.000 per day may be imposed if the entity fails to take reasonable action to comply with the provisions.[160] In West Virginia the Attorney General has exclusive authority to bring action on behalf of residents. Civil action may only be pursued if it is found in court that the defendant has engaged in repeated and willful violations of this law. There is also a maximum penalty of $150.000 per security breach for civil action cases.[161]

An administrative fine has at least three theoretical advantages. Firstly, when the sanction is set at a deterrent level that forces all data controllers to comply, the sanction itself is costless, because it does not have to be executed. Secondly, the imposed fines are considered socially costless transfers of money.[162] Thirdly, higher sanctions allow for lower levels of enforcement to achieve the same level of deterrence. The high sanctions in Article 84(4) of the GDPR can consequently reduce enforcement costs. Nevertheless, it is essential that control bodies execute their monitoring and that fines are imposed to ensure constant vigilance. Our research in *chapter 5* indicates that companies in the U.S. evaluate whether to issue breach notifications based primarily on the reputational damage that results from the lapse in notification, and less on the tangible consequences of not notifying under the DBNL.[163]

---

[159] Digital Guardian (2018) The Definitive Guide to U.S. State Data Breach Laws.

[160] Section 8-19-11, Code of 23 Alabama 1975

[161] West Virginia Code § 46A-2A-101

[162] Contrary to other sanctions such as imprisonment.

[163] The direct financial consequences were checked via the private cause of action and penalty cap provision, which were both insignificant.

Relevant data collection

Finally, the analysis in chapter 6 illustrates that two core elements are missing for measuring the direct (fewer data breaches) and indirect (less identity theft) effects of the GDPR: a large-scale public collection of identity theft data, and a large-scale public collection of data breaches (both present in the U.S.) not only at the Member-States level, but also at the European level. The sources of identity theft data currently vary from country to country (e.g., police forces, associations), and there is no common framework to define them.

The situation for the three most populous European countries is as follows: in Germany, identity theft falls under internet crime. It includes phishing, fraud related to services and goods conducted via the Internet and malicious software. According to the German Institute for Economic Research, in 2015 the identity theft rate ranged between 1.265 and 4.135 per 100.000 inhabitants (based on analysed regions). In the UK, the number of incidences reported by the CIFAS[164] and recorded in the National Fraud Database was 169.592 for 2015, and 172.919 for 2016. However, this statistic only includes identity theft reported by the 277 CIFAS members. In France, a survey conducted by Fellowes / ObSoCo[165] in 2015 found that 200.000 incidences of identity theft take place yearly, in line with the figure reported by CREDOC[166] in 2009 (210.000). This overview demonstrates the need for a centralised public repository for such information. A number of European projects launched initiatives in this direction. For instance, EKSISTENZ[167] promoted the establishment of a European Observatory on Identity Theft.[168] This observatory brings together researchers across the EU to create a focal point and repository of knowledge for anti-identity theft projects. The observatory and its website inform citizens on methods, procedures and possibilities to recover their identities after theft; provide policy guidance to EU Member States and advance a common view for European identity protection. At pre-

[164] Credit Industry Fraud Avoidance System https://www.cifas.org.uk

[165] https://www.fellowes.com

[166] Centre de Recherche pour l'Étude et l'Observation des Conditions de Vie https://www.credoc.fr

[167] https://cordis.europa.eu/project/rcn/188570/reporting/en

[168] http://www.idtheftobservatory.eu

sent, 22 organisations participate in the observatory, including universities, research institutes, relevant Member State agencies, police forces and consultancy companies.

Additionally, as the GDPR, contrary to U.S. DBNL, is not inspired by the sunlight-as-disinfectant principle, most data breaches will not be revealed to the public. The GDPR Article 59 states:

> *Each supervisory authority shall draw up an annual report on its activities, which may include a list of types of infringement notified and types of measures taken in accordance with Article 58(2). Those reports shall be transmitted to the national parliament, the government and other authorities as designated by Member State law. They shall be made available to the public, to the Commission and to the Board.*

The inclusion of the list of breaches is, therefore, an option and not a duty. The incident summary submitted to the Directive on Network and Information Security cooperation group is similarly not accessible to the public.[169]

The absence of a central repository limits the impact of potential reputational damage resulting from a reported breach gone public. It, therefore, leads to weak application of the 'the sunlight as disinfectant' principle, which incentivises companies to invest more in cybersecurity and disinfects organisations of shoddy security practices (Ranger 2007). Proper enforcement by audit and detection will consequently be decisive to create greater incentives for organisations to take adequate steps

[169] Directive on Network and Information Security contains breach notification duties for operators of essential services (Energy, Transport, Banking, Financial Market infrastructures, Health sector and Drinking water supply and distribution) and for digital service providers. A Cooperation Group, composed of Member State representatives, the Commission, and ENISA, receives from each State's (single point of) contact a summary report on the number of notifications, the nature of notified incidents, and the actions taken in accordance with the Directive dispositions.

to secure the personal information they store. Without an appropriate enforcement mechanism[170] for notification (and time) compliance, the incentive for notification will be very low.

A preliminary comparison

However, one opportunity to investigate (with some major approximations) identity theft differences within Europe comes from the Commission's Eurobarometer reports, in particular two special reports (2017, 2018).[171] The reports asked EU citizens about whether they had been victims of identity theft, [172] and if they were, if they would contact the police[173]. We multiply these two numbers to have a figure that is comparable to the U.S. statistic that is the number of identity theft actually reported (collected by each state). The results, presented in Table 7.1, show a large difference between the U.S., where 0,11% of the population actually reported identity theft, and the EU were between 0,58% (Greece) and 4,11% (Belgium) of those surveyed said they that suffered an identity theft (and would have reported it to the police). The EU numbers are, in our opinion, should be seen as an upper limit of actual identity theft, since they come from a survey rather than actual reported cases.[174]

---

[170] The GDPR does not give further instruction on how to enforce the obligation, apart from the statement that enforcement should be 'strong' according to Recital 7.

[171] These are special reports 464a (2017) and 480 (2018) "Europeans' attitudes towards cyber security". The Standard Eurobarometer was established in 1974. Each survey consists of approximately 1000 face-to-face interviews per country. Reports are published twice yearly. Special Eurobarometer reports are based on in-depth thematic studies carried out for various services of the European Commission or other EU Institutions and integrated in the Standard Eurobarometer's polling waves.

[172] We divided that value by three as the question QD10 is structured as follows "In the last three years, how often have you personally experienced or being victim of identity theft" (page 27).

[173] QB13.1 "If you experienced or were a victim of identity thefts, who would you contact?" (page 92).

[174] Part of the difference might be that survey participants inflate the numbers because of "telescoping" effects (where incidents occurring outside the reference period are inflated when reported to the interviewer). Also, if the cases would be reported, the police might not deem them all to be legally significant to investigate or even count as an identity theft.

In the same Table, we also compare data breaches statistics for Europe (for 2019) based on a DLA Piper report that has collected available aggregate statistics across the EU[175]. To make the numbers comparable with the U.S., we divide the total breaches by the number of firms in each country.[176] This difference is stark and revealing: while the number of breaches per 100k firms in the U.S. is 20,5[177], in most EU countries this metric is over 100, and in Denmark, Ireland, and the Netherlands there have been more than 5.000 breaches reported per 100.000 firms. This large difference reflects a difference in the notification regime, with European organisations not fearing the reputational effect related to notification to supervising authorities as dictated by GDPR.

[175] DLA Piper GDPR Data Breach Survey: January 2020.

[176] As per U.S. we excluded firms with no employee. Source: EUROSTAT business demography by size class (from 2004 onwards, NACE Rev. 2) [bd_9bd_sz_cl_r2]

[177] The U.S. statistics is for 2018, the latest ITRC number available at the moment of this publication.

Table 7.1 Breaches for 100k and firms, identity theft rate

Country	Estimated ID Theft %	Breaches p100k persons	Breaches p100kFirm	Population	Employer firms
Netherlands	0,90%	147,20	10.544,49	17.081.507	238.456
Ireland	2,16%	132,52	5.712,05	4.784.383	110.998
Denmark	0,93%	115,43	5.544,44	5.748.769	119.684
Finland	1,23%	71,11	2.881,20	5.503.297	135.825
Germany	1,42%	31,12	1.722,02	82.521.653	1.491.314
Sweden	2,17%	48,14	1.684,10	9.995.153	285.712
Luxembourg	2,64%	56,97	1.671,73	590.667	20.129
Slovenia	1,45%	52,55	1.600,44	2.065.895	67.833
Malta	1,40%	31,00	1.073,03	460.297	13.298
Poland	1,87%	13,74	694,13	37.972.964	751.657
Austria	2,27%	12,10	544,61	8.772.865	194.913
UK	2,61%	17,79	524,16	65.808.573	2.233.560
Belgium	4,11%	7,88	469,14	11.351.727	190.672
Estonia	1,37%	9,74	235,89	1.315.634	54.322
Czech Republic	1,52%	4,03	188,39	10.578.820	226.304
France	2,64%	3,20	188,37	66.989.083	1.138.011
Latvia	1,56%	6,13	169,17	1.950.116	70.662
Lithuania	0,89%	4,18	158,56	2.847.904	75.075
Hungary	3,01%	4,87	129,90	9.797.561	367.328
Cyprus	2,61%	4,80	121,21	854.802	33.852
Romania	2,88%	1,90	100,21	19.644.350	372.471
Italy	1,94%	2,05	90,46	60.589.445	1.373.008
Spain	1,84%	2,08	74,12	46.527.039	1.305.705
Greece	0,58%	1,50	39,25	10.768.193	411.555
USA	**0,11%**	**0,38**	**20,48**	**325.025.206**	**6.073.017**

Longitudinal data breach data is even harder to find for Europe, even at aggregated level. The Irish Data Protection Commission is among the few regulators that have released this information, from 2009 (prior to the implementation of the Irish Personal Data Security Breach Code of Practice) to 2019. We plot this data in Figure 7.2, with a steep growth after the GDPR. The implementation of the GDPR in Ireland leads to an approximate 102% increase in the number of reported breaches (i.e., comparing a year before and after the GDPR). This effect can be compared with the effect of enacting a DBNL (+11%) and also informing the AG (+28%) in

the U.S.. The GDPR effect is much stronger and relates back to the over-reporting point in the previous paragraph.[178]

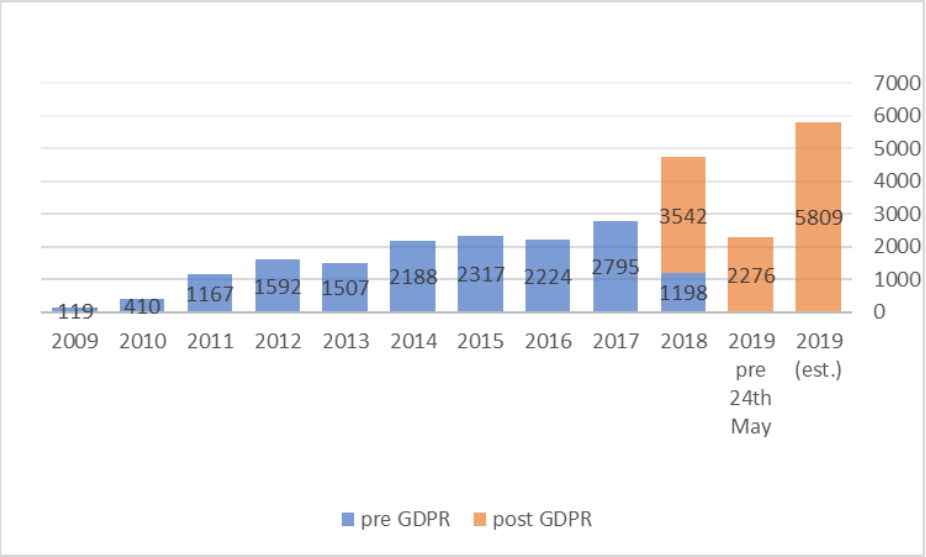


Figure 7.2 - Data breaches by year for Ireland (2009-2019)

Source: Irish Data Protection Officer Annual Reports[179]. The numbers for 2019 are extrapolated from the available period for this year (Jan to May).

[178] With the obvious caveat that the effect is only for one country, Ireland. An additional reason for the difference maybe that in the many U.S. states there are minimum thresholds (in terms of affected records or possible harms) before there is a duty to notify of a breach.

[179] https://www.dataprotection.ie/sites/default/files/uploads/2019-10/Info%20Note_Data%20Breach%20Trends%202018-19_Oct19.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2019-03/DPC%20Annual%20Report%2025%20May%20-%2031%20December%202018.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-11/DPC%20annual%20Report%202018_0.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-11/Annual%20Report%202017.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-11/Annual%20Report%202016.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-11/Annual%20Report%202015.pdf

Surely additional data that will become public will support a more in-depth analysis of the European landscape related to the current regulation addressing data breaches and identity theft.

Final considerations

In view of the greater availability of data for comparison, our research can help EU Member States to increase the positive effects of the data protection reform package in several additional ways. One lesson is to include more actors in the notification flow to support better visibility of the breaches. Another lesson is to adopt a sectoral approach to help balance the discrepancies in detection timing across sectors. Soft-law initiatives such as codes of conduct and codes of practice, implemented at the Member-State level, could also support the management of these discrepancies and may foster the appointment of sectoral bodies as industry reference points to collect and analyse information on notified data breaches and advise on existing security risks and available detection measures. Such initiatives could supplement and support the implementation of the GDPR.

Finally, public breach disclosure may facilitate faster development of the market for risk-rating services and support the improvement of cyber-insurance in Europe. Public disclosure can contribute to satisfying the need for risk-based and economic approaches in managing cybersecurity issues, especially in favour of business organisations and insurance companies with ad hoc products and services, similar to the ones brought to the market in the U.S. by companies such as Quadmetrics[180] and Bitsight.[181]

https://www.dataprotection.ie/sites/default/files/uploads/2018-11/Annual%20Report%202014.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-12/Annual%20Report%202013.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-12/Annual_Report_2012.pdf

https://www.dataprotection.ie/sites/default/files/uploads/2018-12/AnnualReport2011.pdf

[180] https://www.quadmetrics.com/

[181] https://www.bitsighttech.com/

7.3 Main limitations and future study

As with all studies of real-world application and implementation, we must also reflect on the potential impact of possible measurement errors in our data and other limitations. Broadly stated, this research can be extended in two ways.

The first is to improve the quality of the data, both concerning measurements that capture the phenomena of interest, and the independent variables for sectoral characteristics and environment. The first type of data could include additional data breach-related variables such as the number of affected individuals and the PII contained in the breached data. Independent variables could capture, for example, the security maturity level and security investment by sector. Such information would require a novel data collection that is not performed in the different U.S. states.

The second approach is to employ other statistical instruments. There are other statistical instruments and empirical methods that could produce better results, specifically with regards to determining causality. For example, the use of Bayesian multi-level modelling (chapter 6) could be extended to the analysis performed in previous chapters.

The dissertation consists of several standalone studies. Standalone studies have drawbacks regarding overall generalisability, notwithstanding the general issue of generalisability in social sciences (Little 1993, Bernstein et al. 2000). This dissertation studied only one geographical area. Other areas, especially the EU, are well suited for follow-up research along the pathway outlined by this thesis.

7.4 Enhancing security via training and technology

Through this research, we first broadly affirmed that information availability is key to contributing to a safer cybersecurity environment. We then studied to what extent U.S. DBNLs are achieving their goals (sunlight as disinfectant, right to know and identity theft reduction) and what effects they actually generate. Even if the guiding objectives are appropriately addressed by DBNLs, laws alone cannot sustain a safer environment. A full understanding of the required competencies, at individual and organisational level, is a necessary condition to increase the overall security level. In order to gather this understanding, we need to examine

the reasons behind the events generating data breaches. In fact, the data breach drivers are strictly related to the operational routine of organisations where they occur. Namely, three groupings of factors deriving from organisational, business process, and technological attributes are fundamental for managing information security risk across an organization (NIST 2011).

Kamoun and Nicho (2014) identified six organisational factors as possible root causes of breaches: security culture, practices, policies and procedures for handling security, ongoing employee security training, vendor selection, and strong risk management processes. Additionally, organisations must consider both business processes (and the level of exposure these processes incur) and technological assets (such as hardware and software) that play a relevant role in reducing vulnerabilities. Above all, organisations require a continuous and structured update of technologies and technological skills. Thus, investments in technological evolution and security training programmes are essential for organizations to be able to prevent, to face and to react to both cyber-crime attacks and data breaches, especially where it hurts the most: databases.

Nevertheless, leveraging on the single factors is still insufficient. In order to create an efficient and sustainable context for effectively better data protection, in particular of personal data, we need an essential cultural change in order to ensure that each employee, more in general each individual, is aware of the data value as well as of their direct responsibility for the right data treatment.

We shall consider personal data as currency. Typically, currency funds are kept safe in the authorised and screened organisations, each transaction needs a key to identify only those individuals who are authorised to proceed with a specific operation. Such a change is required from organisations, but particularly from individuals who shall start to perceive their personal data as valuable assets. In fact, various companies have taken advantage of individuals' who underestimate this element. People too often 'donate' their own personal data without any interest in their treatment. It is still not common to realise that the breadcrumb trails left

online and offline have real monetary and therefore commercial value,[182] that once transferred should be kept safe by the ones who receive or have access to it. Only the collective awareness of individuals, developed through training and technological engagement, shall generate forceful impacts in terms of more conscious choices. In such a context, 'name and shame' principle can then correctly function as an organisational rating tool for security behaviour.

The crucial point in the discussion is that, at the current status of evolution, data are managed by humans. And humans make mistakes. The open question is consequently to which extent training can effectively avoid or limit such mistakes or whether technology should be placed at the centre of the issue as it should eliminate unintentional human errors. This open question shall be the core of a fascinating future debate that will examine the relation between humans, technology and related processes. One-fourth of the notifications are triggered by data breaches generated by unintended disclosure. Therefore, a large number of data breaches could be potentially avoided by focusing on proper processes, training programmes and implementation of technical solutions to exclude human errors (moving therefore from ex post to ex ante approach).

[182] To emphasise that point, cybersecurity firm Kaspersky Lab run a pop-up shop in London called The Data Dollar Store in September 2017. Inside, one found exclusive t-shirts, mugs and screen prints by street artist Ben Eine. Customers could only buy them by giving up some personal data. https://getcodify.com/kaspersky-store-accepts-personal-data-currency/

References

Abel A.B., & Eberly J.C. (1999). The impact of uncertainty on capital accumulation. Journal of Monetary Economics, Vol. 44, pp. 330-377.

Ablon, L., Heaton P., Lavery D., & Romanosky S. (2016). Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information. Rand.

Acquisti A., Friedman A., & Telang R. (2006). Is There a Cost to Privacy Breaches? An Event Study. Paper presented at the fifth workshop on the Economics of Information Security, University of Cambridge, England, June 2006.

Alred G. J., Brusaw C. T., & Oliu W. E. (2011). The Business Writer's Handbook, Boston: Bedford/St. Martin's.

Anderson R., & Moore T. (2006). The Economics of Information Security. Science, 314, pp. 610-613.

Angrist J. D., & Pischke J. S. (2014). Mastering 'Metrics: The Path from Cause to Effect. Princeton, NJ: Princeton University Press.

Baker Hostetler (2014). State Data Breach Statute Form. http://www.bakerlaw.com/files/Uploads/Documents/Data%20Breach%20documents/State_Data_Breach_Statute_Form.pdf.

Benoit W. L., & Shirley D. (1997). Appropriateness and Effectiveness of Image Repair Strategies. Communication Reports 10: 153–63.

Bernstein, S., Lebow R. N., Gross Stein J, & Weber S. (2000). God Gave Physics the Easy Problems: Adapting Social Science to an Unpredictable World. European Journal of International Relations 6 (1): 43–76. doi:10.1177/1354066100006001003.

Bies R. J. (2012). The 10 Commandments for Delivering Bad News. Forbes Leadership forum.

Bies R. J. (2013). The Delivery of Bad News in Organizations: A Framework for Analysis. Journal of Management Vol. 39 No. 1, January 2013 136-162.

Bies R. J., & Shapiro D. L. (1987). Interactional fairness judgments: The influence of causal accounts. Social Justice Research, 1: 199-218.

Bisogni F. (2013). Evaluating Data Breach Notification Laws - What Do the Numbers Tell Us? TPRC 41: The 41st Research Conference on Communication, Information and Internet Policy, September 2013.

Bisogni F. (2015). Data Breaches and the Dilemmas in Notifying Customers. WEIS 2015: 14th Workshop on the Economics of Information Security.

Bisogni F. (2016). Proving Limits of State Data Breach Notification Laws: Is a Federal Law the Most Adequate Solution? Journal of Information Policy 6 (2016): 154-205, Penn State University Press.

Bisogni F., Asghari H., & Van Eeten M. (2017). Estimating the size of the iceberg from its tip. An investigation into unreported data breach notifications. WEIS 2017 - 16th annual Workshop on the Economics of Information Security, At La Jolla, US, June 2017.

Bovée C. L., & Thill J. V. (2012). Writing Negative Messages. In Business Communication Today. 11th ed., 180–208. Upper Saddle River, NJ: Prentice Hall, 2012.

Bradford J. L., & Garrett, D. E. (1995). The effectiveness of corporate communicative responses to accusations of unethical behaviour. Journal of Business Ethics, 14, 875–892.

Bruck T., Karaisi M., & Schneider F. (2006). A survey of the economics of security. NEAT Economics of Security working paper 1.

Bug M., Kroh M., Meier K., Rieckmann J., van Um E., & Wald N. (2015). WISIND-Datensätze: Kriminalitätsbefragung. Berechnungen des DIW Berlin.

Burdon M. (2011). The conceptual and operational compatibility of data breach notification and information privacy laws. Dissertation, Queensland University of Technology.

Caballero R. J. (1991). On the sign of the investment-uncertainty relationship. American Economic Review, Vol. 81, No. 1, pp. 279-288.

California Civil Code § 1729.98(a).

Campbell K., Gordon L. A., Loeb. M. P., & Zhou L. (2003). The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market. Journal of Computer Security 11: 431–48.

Carter, C. (2012). Negative Messages. In Keys to Business Communication: Success in College, Career, and Life, 208–35. Upper Saddle River, NJ: Prentice Hall.

Cavusoglu H., Mishra B., & Raghunathan S. (2004). The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers. International Journal of Electronic Commerce 9: 70–104.

Census Brief (2010). Population Distribution and Change: 2000 to 2010.

Chlotia P. G., & Ncube M. (2011). A longitudinal analysis of data breaches. Information Management & Computer Security, Vol. 19 Issue: 4: 216-230.

Choi S., & Johnson M. E. (2017). Do Hospital Data Breaches Reduce Patient Care Quality? WEIS 2017 - 16th annual Workshop on the Economics of Information Security, At La Jolla, US, June 2017.

Claburn T. (2008). Most Security Breaches Go Unreported. Dark Reading. http://www.darkreading.com/attacks-and-breaches/most-security-breaches-go-unreported/d/d-id/1070576.

CLLA (2012). Data Breach Notification Laws by State. http://www.clla.org/documents/breach.xls.

Cohen J. R. (1999). Advising clients to apologize. Southern California Law Review, 72, 1009–1073.

Commercial Law League of America (2012). Data Breach Notification Laws by State.

Conlon D. E., & Murray N. M. (1996). Customer perceptions of corporate response to product complains: the role of explanations. Academy of Management Journal, Vol.39, No. 4, 1040-1056.

Coombs W. T. (1995). Choosing the right words. Management Communication Quarterly, 8, 447-477.

Coombs W. T. (1999). Ongoing crisis communication: Planning, managing and responding. Thousand Oaks, CA: Sage.

Coombs W. T., & Holladay S. J. (2008). Comparing Apology to Equivalent Crisis Response Strategies: Clarifying Apology's Role and Value in Crisis Communication. Public Relations Review, 34, 252-257

Creelman V. (2012). The Case for "Living" Models. Business Communication Quarterly 75 (2) 192-207.

Data breaches and identity theft (2005). Prepared statement of the Federal Trade Commission before the Committee on Commerce, Science and Transportation. US Senate 109th Congress.

Dean D. W. (2004). Consumer reaction to negative publicity: Effects of corporate reputation, response, and responsibility for a crisis event. Journal of Business Communication, 41, 192–211.

DeKay S. H. (2012). Where is the research on Negative Messages. Business Communication Quarterly 75 (2) 173-175.

Di Ciccio F. (2014). Comparison of identity theft in different countries. https://courses.cs.ut.ee/MTAT.07.022/2014_fall/uploads/Main/francesco-report-f14.pdf.

Digital Guardian (2018). The Definitive Guide to U.S. State Data Breach Laws.

Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016.

Dixit A.K., & Pindyck R.S. (1994). Investment under Uncertainty. Princeton University Press.

DLA Piper (2020) GDPR Data Breach Survey, January 2020.

Dolezel, D., & McLeod, A. (2019). Managing Security Risk. Modeling the Root Causes of Data Breaches. The Health Care Manager: October/December 2019 - Volume 38 - Issue 4 - p 322–330.

Draper A. (2006). Identity theft: Plugging the massive data leaks with a stricter nationwide breach notification law. Journal Marshall & Law Review, 40, 681–703.

Dutch DPA (2019). Overzicht meldingen datalekken eerste kwartaal 2017. Available online at: https://www.autoriteitpersoonsgegevens.nl/nl/nieuws/ap-ontvangt-bijna-21000-datalekken-2018.

Economic Census (U.S. Government) (2012). Online at: http://www.census.gov/econ/census/. Accessed 01-Feb-2017.

Edwards B., Hofmeyr S., & Forrest S. (2015). Hype and Heavy Tails: A Closer Look at Data Breaches. WEIS 2015: 14th Workshop on the Economics of Information Security, June 2015.

ENISA (2011). Data Breach Notifications in Europe.

European Commission. Special Eurobarometer report 464a (2017). Europeans' attitudes towards cyber security.

European Commission. Special Eurobarometer report 480 (2018). Europeans' attitudes towards cyber security.

Faulkner B. (2007). Hacking into Data Breach Notification Laws. 59(5) Florida Law Review. 1097, 1104.

Federal S.177 - Data Security and Breach Notification Act of 2015.

Fuchs-Burnett T. (2002). Mass Public Corporate Apology. Dispute Resolution Journal 57, no. 3: 26–32.

Galor E., & Ghose A. (2004). The Economic Consequences of Sharing Security Information. Advances in Information Security, Vol. 12. 124 No. 81, 1st Q. 2011.

GAO, United States Government Accountability Office (2007). Report to congressional requesters. Personal Information. Data Breaches Are Frequent, but Evidence of Resulting Identity Theft Is Limited; However, the Full Extent Is Unknown. July 2007.

Garg A., Curtis J., & Halper H. (2003). Quantifying the financial impact of IT security breaches. Information Management & Computer Security, 11/2: 74-83.

Garrison C. P., & Ncube M. (2011). A longitudinal analysis of data breaches. Information Management & Computer Security, Vol. 19 Issue: 4, pp.216-230, https://doi.org/10.1108/09685221111173049.

Gavison R. (1980). Privacy and the Limits of Law. 89 YALE L.J. 421, 423.

Gerber B., & Teske P. (2000). Regulatory Policymaking in the American States: A review of theories and evidence. Political research quarterly 53 (4): 849-886.

Gordon G. R., Rebovich D. J., Choo K., & Gordon J. B. (2007). Identity fraud trends and patterns: Building a data-based foundation for proactive enforcement. Utica, NY: Center for Identity Management and Information Protection (CIMIP), Utica College.

Gordon L. A., & Loeb M. P. (2002). The Economics of Information Security Investment. Advances in information security. Vol. 12.

Gordon L. A., Loeb M. P., & Lucyshyn W. (2003). Information Security Expenditures and Real Options: A Wait-and-See Approach. (May 31, 2003). Computer Security Journal, Vol. XIX, No. 2, Spring, 2003.

Granade P. (2008). Sec: Best practices lost or stolen data: minimizing fallout. Inside Counsel, May, 7 pp.

Greenberg J. (1990). Looking fair vs. being fair: Managing impressions of organizational justice. In B. M. Staw & L. L. Cummings (Eds.), Research in organizational behavior, vol. 12:111-157. Greenwich, CT: JAI Press.

Greene, E. (1990). Media effects on jurors. In Law and Human Behavior, Volume 14, Issue 5, pp 439–450.

Greene W.H. (2007). Econometric Analysis. Prentice Hall.

Guffey M. E., & Lowey D. (2011). Business Communication: Process and Product. 7th ed. Mason, OH: South-Western/Cengage Learning.

Hans, V. P., & J. L. Dee (1991). Media coverage of law: Its impact on juries and the public. American Behavioral Scientist, 35 (2): 136- 149.

Heath R.L., Jaesub L., & Lan N. (2009). Crisis and risk approaches to emergency management planning and communication: the role of similarity and sensitivity. J Pub Relat Res 2(2):123–141.

Hilbe J. M. (2011). Negative Binomial Regression. Cambridge University Press, 2nd Edition.

Huang Y., & Su S. (2009). Determinants of Consistent, Timely, and Active Responses in Corporate Crises. Public Relations Review, 35, 7-17.

Hynes G. E. (2008). Routine Messages. In Managerial Communication: Strategies and Applications. 4th ed., 99–125. Columbus, OH: McGraw-Hill.

Identity Theft Resource Center (ITRC) (2014). 2014 Data Breach Reports. https://www.idtheftcenter.org/images/breach/DataBreachReports_2014.pdf.

Identity Theft Resource Center (ITRC) (2016). Data Breaches. https://www.idtheftcenter.org/2016databreaches/. Accessed 28-Feb-2017.

Identity Theft Resource Center (ITRC). (2017). Data Breaches. https://www.idtheftcenter.org/data-breaches/. Accessed 28-Feb-2017.

Identity Theft Resource Center (ITRC). (2018). Data Breaches. https://www.idtheftcenter.org/data-breaches/. Accessed 29-Nov-2019.

Internet Crime Compliant Center. (2016). 2015 Internet Crime Report. US Department of Justice, Federal Bureau of Investigation.

Irish Data Protection Commission, Information Note: Data Breach Trends from the First Year of the GDPR, October 2019.

Jansen F., & Janssen D. (2011). Explanations First: A Case for Presenting Explanations Before the Decision in Dutch Bad-News Messages. Journal of Business and Technical Communication 25, no. 1: 36–67.

Javelin Strategy & Research (2009). 2009 Identity Fraud Survey Report.

Javelin Strategy & Research (2011). 2011 Identity Fraud Survey Report.

Javelin Strategy & Research (2018). 2018 Identity Fraud Survey Report.

Joerling J. (2010). Data Breach Notification Laws: An Argument for a Comprehensive Federal Law to Protect Consumer Data. 32 Wash. U. J. L. & Pol'y 467. http://openscholarship.wustl.edu/law_journal_law_policy/vol32/iss1/14.

Kamoun F., & Nicho M. (2014). Human and organizational factors of healthcare data breaches: the Swiss Cheese Model of data breach causation and prevention. Int J Healthcare Info Syst Informatics. 2014;9:42-60.

Kang J. (1998). Information Privacy in Cyberspace Transactions. 50 STAN. L. Rev. 1193, 1203.

Ko M., & Dorantes C. (2006). The Impact of Information Security Breaches on Financial Performance of the Breached Firms: An Empirical Investigation. Journal of Information Technology Management 17: 13–22.

Kolin P. C. (2007). Successful Writing at Work. Boston: Houghton Mifflin.

Kraft M. E., Stephan M., & Abel T. (2011). Coming clean, Cambridge, MA: MIT Press.

Kreps D. (1990). A Course in Microeconomic Theory. Princeton University Press.

Kruschke, J. K. (2015). Doing Bayesian data analysis: A tutorial with R and BUGS. New York, NY: Academic Press.

Kwon J., & Johnson E. (2015). The Market Effect of Healthcare Security: Do Patients Care about Data Breaches? WEIS 2015: 14th Workshop on the Economics of Information Security.

Laube S., & Böhme R. (2015). The Economics of Mandatory Security Breach Reporting to Authorities. WEIS 2015: 14th Workshop on the Economics of Information Security, June 2015.

Lehman C. M., & DuFrene D. M. (2012). Delivering Bad-News Messages. In BCOM. 3rd ed., 110–29. Mason, OH: South-Western/Cengage Learning.

Lenard T. M., & Rubin P. H. (2005). An Economic Analysis of Notification Requirements for Data Security Breaches. Technology Policy Institute.

Limpert E., Stahel W. A., & Abbt M. (2001). Log-normal Distributions across the Sciences: Keys and Clues: On the charms of statistics, and how mechanical models resembling gambling machines offer a link to a handy way to characterize log-normal distributions, which can provide deeper insight into variability and probability—normal or log-normal: That is the question. BioScience, Volume 51, Issue 5, May 2001, Pages 341–352.

Little, D. (1993). On the Scope and Limits of Generalizations in the Social Sciences. Synthese 97 (2): 183–207. doi:10.1007/BF01064114.

Locker K. O. (1999): Factors in reader responses to negative letters: Experimental evidence for changing what we teach. Journal of Business and Technical Communication, 13, 5-48.

Locker K. O., & Kienzler D. S. (2010). Delivering Negative Messages. In Business and Administrative Communication. 9th ed., 286–321. New York: McGraw-Hill/Irwin, 2010.

Loewenstein G., John L., & Volpp K. (2012). Using decision errors to help people help themselves. In E. Shafir (Ed.), The behavioral foundations of policy. Princeton, NJ: Princeton University Press.

Lyon L., & Cameron G. T. (1998). Fess up or stonewall? An experimental test of prior reputation and response style in the face of negative news coverage. Web Journal of Mass Communication Research, 1(4). Retrieved May 1, 2015 from http://www.scripps.ohiou.edu/wjmcr/vol01/1-4a.htm.

Mas-Colell D., Winston M. & Green J. (1995). Microeconomic Theory. Oxford University Press.

McElreath, R. (2016). Statistical Rethinking: A Bayesian Course with Examples in R and Stan. CRC Press.

McKinsey Global Institute (2015). Digital America: a tale of the haves and have-mores. December 2015.

Microsoft (2018). Security Intelligence Report. Volume 23. https://info.microsoft.com/rs/157-GQE-382/images/EN-US_CNTNT-eBook-SIR-volume-23_March2018.pdf.

Mintz Levin (2012). State Data Security Breach Notification Laws. http://www.mintz.com/newsletter/2007/PrivSec-DataBreachLaws-02-07/state_data_breach_matrix.pdf.

Moor J. H. (2010). Towards a Theory of Privacy in the Information Age. 27 COMP. & Soc. 27, 31.

Moore T., Dynes S., & Chang F. (2016). Identifying how firms manage cybersecurity investment. WEIS 2016: 15th Workshop on the Economics of Information Security.

Nieuwesteeg B., & Faure M. (2018). An analysis of the effectiveness of the EU data breach notification obligation. Computer Law & Security Review, 34(6), 1232-1246.

NIST Special Publication 800-39 (2011). Managing Information Security Risk Organization, Mission, and Information System View. March 2011.

Oliu W. E., Brusaw C. T., & Alred G. J. (2009). Writing Business Correspondence. In Writing that Works: Communicating Effectively on the Job. 9th ed., 320–52. Boston: Bedford/St. Martins.

Patel A., & Reinsch L. (2003). Companies Can Apologize: Corporate Apologies and Legal Liability. Business Communication Quarterly 66.1, 9-25.

Perkins (2014). Security Breach Notification Chart. http://www.perkinscoie.com/files/upload/LIT_09_07_SecurityBreachExhibits2.pdf.

Personal Data Notification & Protection Act. http://www.whitehouse.gov/sites/default/files/omb/legislative/letters/updated-data-breach-notification.pdf.

Personal Data Privacy and Security Act of 2014, S. 1897, 113th Cong. (Sen. Leahy).

Personal Data Protection and Breach Accountability Act of 2014, S. 1995, 113th Cong. (Sen. Blumenthal).

Pike G. H. (2008). Legal Issues: Data Breaches Top the Agenda at RSA Conference. Information Today 25, no. 6: 19.

Ponemon Institute LLC (2012). 2012 Consumer Study on Data Breach Notification.

Ponemon Institute, LLC (2014). 2014 Cost of Data Breach Study: Global Analysis.

Ranger S. (2007). Data Breach Laws Make Companies Serious about Security. September 3, 2007. Silicon.com. http://management.silicon.com/itdirector/0,39024673,39168303,00. htm?r=1.

Ray S. J. (1999). Strategic communication in crisis management lessons from the airline industry. Westport, CT: Quorum.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

Reynolds B., & Seeger M. (2005). Crisis and emergency risk communication as an integrative model. J Health Commun Int Perspect 10(1):43–55.

Roberds W., & Schreft S. L. (2008). Data breaches and identity theft. Working Paper, Federal Reserve Bank of Atlanta, No. 2008-22.

Romanosky S., & Acquisti A. (2009). Privacy Costs and Personal Data Protection: Economic and Legal Perspectives. Berkeley Technology Law Journal, 24 2009, Nr. 3.

Romanosky S., Telang R., & Acquisti A. (2011). Do Data Breach Disclosure Laws Reduce Identity Theft? Journal of Policy Analysis and Management 30, no. 2 (2011): 256–86.

Romanosky S., Hoffman D., & Acquisti A. (2014). Empirical Analysis of Data Breach Litigation. Journal of Empirical Legal Studies, 11 (2014): 74–104.

Sasso, B. (2015). Why Businesses Love Obama's Push for Security Regulation. National Journal (January 2015). http://www.nationaljournal.com/tech/why-businesses-love-obama-spush-for-security-regulation-20150112.

Schafer B. (2016). Compelling truth: legal protection of the infosphere against big data spills. Phil. Trans. R. Soc. A.37420160114 rans. R. Soc. A.37420160114 http://doi.org/10.1098/rsta.2016.0114.

Schafer B. (2017). Speaking Truth to/as Victims – A Jurisprudential Analysis of Data Breach Notification Laws. In: Taddeo M., Floridi L. (eds) The Responsibilities of Online Service Providers. Law, Governance and Technology Series, vol 31. Springer, Cham.

Schwartz P. M., & Janger E. (2007). Notification of Data Security Breaches. Michigan Law Review 105, no. 913.

Schwartz P. M., & Solove D. J. (2014). Reconciling Personal Information in the United States and European Union. 102 Cal. L. Rev. 877.

Security Scorecard (2016). 2016 Financial Industry Cybersecurity Report. Security Scorecard R&D Department. August 2016.

Seeger M. W. (2006). Best Practices in Crisis Communication: An Expert Panel Process. Journal of Applied Communication Research, 34:3, 232-244.

Seeger M. W., Sellnow T. L., & Ulmer R. R. (2003). Communication, organization and crisis. West port, CT: Quorum.

Seeger M. W., Sellnow T. L., & Ulmer R. R. (1998). Communication, organization, and crisis. In M. E. Roloff (Ed.), Communication yearbook (Vol. 21, pp.231-275). Thousand Oaks, CA:SAGE.

Shinar D., & McKnight, A. J. (1986). The effects of enforcement and public information on compliance. In L. Evans, & R.C. Schwing (Eds.), Human behaviour and traffic safety. New York: Plenum Press.

Shwom B. G., & Snyder L. G. (2012). Communicating Bad-News Messages. In Business Communication: Polishing Your Professional Presence, 212–45. Upper Saddle River, NJ: Prentice Hall.

Simitian J. (2009). How a bill becomes a law, really. UCB security breach notification symposium March 6, 2009, Berkeley Technology Law Journal, 24, 1009–1018.

Sinanaj G., & Zafar H. (2016). Who wins in a Data Breach? - A comparative study on the Intangible Costs of Data Breach Incidents. PACIS 2016 Proceedings. Paper 60.

Skinner T. H. (2003). California's database breach notification security act: The first state breach notification law is not yet a suitable template for national identity theft legislation. Richmond Journal Law & Technology, 10, 1–40.

Steelman, T.A., & McCaffrey, S. (2013). Best practices in risk and crisis communication: Implications for natural hazards management. Nat Hazards 65, 683–705 (2013). https://doi.org/10.1007/s11069-012-0386-z.

Stephens K. K., Malone P. C., & Bailey C.M. (2005). Communicating with Stakeholders during a Crisis, Evaluating Message Strategies. Journal of Business Communication, Volume 42, Number 4, October 2005 390-419.

Steptoe (2015). Data Breach Notification Chart 2015. Comparison of US State and Federal Security Breach Notification Laws. Current through May 26, 2015.

Sunstein C. R. (1999). Informational Regulation and Informational Standing: Akins and Beyond. University of Pennsylvania Law Review147, 613-675.

Taddeo M., & Floridi L. (2017). The Moral Responsibilities of Online Service Providers. In: Taddeo M., Floridi L. (eds) The Responsibilities of Online Service Providers. Law, Governance and Technology Series, vol 31. Springer, Cham

Telang R., & Wattal S. (2007). An Empirical Analysis of the Impact of Software Vulnerability Announcements on Firm Stock Price. IEEE Transactions on Software Engineering 33: 544–57.

ThreatTrack Security (2013). Malware Analysts Have the Tools to Defend Against Cyber-Attacks, But Challenges Remain. White Paper. November 2013.

Towle H. K. (2003). Identity theft: Myths, methods, and new law. Rutgers Computer & Technology Law Journal, 30, 237–326.

Twerski A. D., & Cohen N. B. (1999). The Second Revolution in Informed Consent: Comparing Physicians to Each Other. 94 Nw U L Rev 1, 9.

US Senate 109th Congress. 2005. Data Breaches and Identity Theft. Prepared statement of the Federal Trade Commission before the Committee on Commerce, Science and Transportation.

Veltsos J. R. (2012). An Analysis of Data Breach Notifications as Negative News. Business Communication Quarterly 75, no. 2: 192–207.

Verizon (2014). 2014 Data Breach Investigations Report.

Vogel, D. (1995). Trading Up: Consumer and Environmental Regulation in a Global Economy.

Vogel, D., & Kagan, R. (2004). Introduction. In: Vogel, D., Kagan, R.A. (Eds.), Dynamics of Regulatory Change: How Globalization Affects National Regulatory Policies.

Walsh E. (2014). A Company That Does Background Checks For The US Government Was Victim Of 'State-Sponsored' Cyber Attack. Reuters, August 7. http://www.reuters.com/article/2014/08/07/us-usa-security-contractor-idUSKBN0G62N420140807.

White House, Economic Report of the President Together with The Annual Report of the Council of Economic Advisers, March 2019.

Williams D. E., & Olaniran, B. A. (1998). Expanding the crisis planning function: Introducing elements of risk communication to crisis communication practice. Public Relations Review, 24 (3), 387-402.

Winn J. K. (2009). Are "Better" security breach notification laws possible? 2–3. Berkley Technology Law Journal, 24, 1133–1165.

Wolff J. (2018). You will see this message when it is too late. The legal and economic aftermath of cybersecurity breaches. the MIT Press

Attorney general websites accessed for notification downloads:

https://oag.ca.gov/ecrime/databreach/list

http://www.oag.state.md.us/idtheft/businessGL.htm

http://doj.nh.gov/consumer/security-breaches/

http://www.atg.state.vt.us/issues/consumer-protection/privacy-and-data-security/vermont-security-breaches.php

http://www.maine.gov/ag/consumer/identity_theft/

Letters downloaded for statistics:

1. East West Bank—02 January 2014

2. Erie Insurance—02 January 2014

3. T-Mobile—02 January 2014

4. Unicef letter to Consumers re Security Breach—06 January 2014

5. Customer Notice Final Generic version—06 January 2014

6. AHS letter to Consumers re Security Breach—06 January 2014

7. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—07 January 2014

8. Experian—07 January 2014

9. Lafarge West, Inc.—07 January 2014

10. Straight Dope LLC—09 January 2014

11. Barry University letter to Consumers re Security Breach—10 January 2014

12. Edgepark letter to Consumers re Security Breach—13 January 2014

13. Update Legal—13 January 2014

14. Apex Systems, Inc.—14 January 2014

15. Genworth—15 January 2014

16. Easton Bell Sports letter to Consumers re Security Breach—16 January 2014

17. Burlington letter to Consumers re Security Breach—16 January 2014

18. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—16 January 2014

19. TD Bank—16 January 2014

20. Vermont Health Connect—17 January 2014

21. Neiman Marcus letter to Consumers re Security Breach—17 January 2014

22. Dartmouth Hitchcock letter to Consumers re Security Breach—20 January 2014

23. Complete Medical Homecare—21 January 2014

24. PCC Structurals—21 January 2014

25. Discover letter to Consumers re Security Breach—22 January 2014

26. Sidney Regional Medical Center—22 January 2014

27. MilCo Enterprises, Inc. DBA EasyDraft—22 January 2014

28. Focus on Surety LLC DBA Suretegrity—22 January 2014

29. Coca Cola letter to Consumers re Security Breach—23 January 2014

30. W. J. Bradley Mortgage Capital, LLC—23 January 2014

31. TD Bank letter to Consumers re Security Breach—24 January 2014

32. State Industrial letter to Consumers re Security Breach—27 January 2014

33. Michaels letter to Customers re Security Breach—27 January 2014

34. Bring it To Me, LLC—29 January 2014

35. Tribeca Film Institute—30 January 2014

36. Intuit—30 January 2014

37. Beebe Healthcare—31 January 2014

38. Neilsen letter to Consumers re Security Breach—03 February 2014

39. University of California Davis Medical Center—03 February 2014

40. Greenleaf Book Group, LLC—03 February 2014

41. Bank of the West—05 February 2014

42. K. Min Yi, M.D. General Surgery—05 February 2014

43. St. Joseph Health System—05 February 2014

44. Mimeo.com—05 February 2014

45. San Francisco Airport letter to Consumers re Security Breach 1—07 February 2014

46. Easter Seal Society of Superior California—07 February 2014

47. Catamaran—07 February 2014

48. Farmers and Merchants Trust Company of Chambersburg—07 February 2014

49. Mymatrixx—07 February 2014

50. Home Depot letter to Consumers re Security Breach—10 February 2014

51. The Freeman Company—10 February 2014

52. 80s Tees letter to Consumer re Security Breach—11 February 2014

53. Embassy suites—11 February 2014

54. Fresenius Medical Care—11 February 2014

55. TD Bank—11 February 2014

56. Zevin Asset Mgmt letter to Consumer re Security Breach—13 February 2014

57. MSPCC letter to Consumers re Security Breach—13 February 2014

58. Carmike Cinemas, Inc.—13 February 2014

59. Experian letter to Consumers re Security Breach—14 February 2014

60. Rubin Lublin, LLC—14 February 2014

61. TD Bank Security Breach Notice—18 February 2014

62. Blue Shield of California—18 February 2014

63. John Hancock Life & Health Insurance Company—18 February 2014

64. Department of Resources Recycling and Recovery—20 February 2014

65. Discover Financial Services—21 February 2014

66. Alaska Communications letter to Consumer re Security Breach—24 February 2014

67. Merrill Lynch Wealth Management—24 February 2014

68. DST Systems, Inc.—24 February 2014

69. eScreen, Inc.—25 February 2014

70. The Variable Annuity Life Insurance Company—26 February 2014

71. Mkenna Long & Aldridge—26 February 2014

72. Smucker letter to Consumers re Security Breach—27 February 2014

73. L.A. Care Health Plan—27 February 2014

74. ProAssurance Mid-Continent Underwriters, Inc.—27 February 2014

75. Sands Casino letter to Consumers re Security Breach—28 February 2014

76. AppleCare Insurance Services, Inc.—28 February 2014

77. Digia USA, Inc.—28 February 2014

78. ThermoFisher—28 February 2014

79. Capital One letter to Consumers re Security Breach—03 March 2014

80. Timken Co letter to Consumers re Security Breach—03 March 2014

81. Assisted Living Concepts LLC Security Breach Notice—03 March 2014

82. St. Joseph Health—03 March 2014

83. Equifax—03 March 2014

84. EMC—03 March 2014

85. Eureka Internal Medicine—04 March 2014

86. Assisted Living Concepts Notice—05 March 2014

87. Oak letter to Consumers re Security Breach—06 March 2014

88. OANDA letter to Consumers re Security Breach—12 March 2014

89. UCSF Family Medicine Center at Lakeshore—12 March 2014

90. Silversage Advisors—13 March 2014

91. USAA letter to Consumers re Security Breach—17 March 2014

92. Arcadia Health Services, Inc. d/b/a Arcadia Home Care & Staffing—17 March 2014

93. Shelburne Country Store Notice to Consumers—18 March 2014

94. Auburn University letter to Consumers re Security Breach—19 March 2014

95. Discover letter to Consumers re Security Breach—20 March 2014

96. Marian Regional Medical Center—20 March 2014

97. Sorenson letter to Consumers re Security Breach—21 March 2014

98. Castle Creek Properties, Inc., dba Rosenthal the Malibu Estates—21 March 2014

99. Human Resource Advantage—21 March 2014

100. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—25 March 2014

101. RBS—25 March 2014

102. Palomar Health—28 March 2014

103. ITHAKA—31 March 2014

104. RK Internet—31 March 2014

105. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—01 April 2014

106. Susquehanna Health—01 April 2014

107. Kaiser Permanente Northern CA Department of Research—02 April 2014

108. California Department of Corrections and Rehabilitation—02 April 2014

109. American Health Information Management Association (AHIMA)—02 April 2014

110. Citibank, N.A.—02 April 2014

111. Cole Taylor Bank—03 April 2014

112. Sutherland Healthcare Solutions—03 April 2014

113. Logos Management Software, LLC—03 April 2014

114. Parallon—03 April 2014

115. Deltek letter to Consumer re Security Breach—07 April 2014

116. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—07 April 2014

117. City of Crossville, Tennessee—07 April 2014

118. FujiFilm—07 April 2014

119. CRL letter to Consumer re Security Breach—08 April 2014

120. StumbleUpon, Inc.—08 April 2014

121. LaCie USA—11 April 2014

122. Society for Science & the Public—11 April 2014

123. Wilshire Mutual Funds letter to Consumers re Security Breach—14 April 2014

124. Mid Atlantic Professionals, Inc. DBA SSI—14 April 2014

125. Blue Cross and Blue Shield of Kansas City, Inc.—16 April 2014

126. Discover letter to Consumers re Security Breach—17 April 2014

127. Michaels press release re Security Breach—17 April 2014

128. VFW letter to Consumers re Security Breach—21 April 2014

129. NCO FinancialRevSpring, Inc. letter to Consumers re Security Breach—22 April 2014

130. Snelling letter to Consumers re Security Breach—22 April 2014

131. Johns Hopkins University (Identity Theft)—22 April 2014

132. Seattle University—22 April 2014

133. Larsen Dental Care—22 April 2014

134. L Brands, Inc.—23 April 2014

135. JCM Partners letter to Consumer re Security Breach—24 April 2014

136. Westlife Distribution USA, LLC—24 April 2014

137. CCC letter to Consumer re Security Breach—25 April 2014

138. Willis North America letter to Consumers re Security Breach—25 April 2014

139. Central City Concern—25 April 2014

140. Federal Home Loan Mortgage Corporation (Freddie Mac)—25 April 2014

141. Seterus—29 April 2014

142. Boomerang Tags—30 April 2014

143. UMass Memorial MC ltrt Consumer (Redacted) re Security Breach—05 May 2014

144. ground(ctrl)—05 May 2014

145. Maschino, Hudelson & Associates—05 May 2014

146. Department of Child Support Services—06 May 2014

147. 2014 Gingerbread Shed letter to Consumer re Security Breach—07 May 2014

148. Green's Accounting—07 May 2014

149. Mercer HR Services, LLC—07 May 2014

150. Entercom Portland, LLC—07 May 2014

151. PREIT—08 May 2014

152. Lowes letter to Consumer re Security Breach—12 May 2014

153. Santander Bank, N. A.—12 May 2014

154. Hubbard-Bert, Inc.—13 May 2014

155. University of California Irvine—14 May 2014

156. Precision Planting LLC—14 May 2014

157. Discover letter to Consumers re Security Breach—16 May 2014

158. Affinity Gaming—19 May 2014

159. Paytime Harrisburg, Inc. d/b/a Paytime, Inc.—21 May 2014

160. Hanover Foods Corporation—21 May 2014

161. CoreLogic Saferent—21 May 2014

162. Experian letter to Consumer re Security Breach—22 May 2014

163. San Diego State University—22 May 2014

164. CenturyLink—22 May 2014

165. Ebay—22 May 2014

166. Power Equipment Direct Security Breach Notice to Consumers—23 May 2014

167. The Home Depot, Inc.—23 May 2014

168. AutoNation (Ford White Bear Lake) letter to Consumers re Security Breach—26 May 2014

169. Placemark Investments, Inc.—27 May 2014

170. Walgreen Co.—27 May 2014

171. Service Alternatives, Inc.—27 May 2014

172. SHARPER FUTURE—28 May 2014

173. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—29 May 2014

174. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—02 June 2014

175. Kimpton—02 June 2014

176. Gordon Feinblatt LLC—02 June 2014

177. Rowan Companies, Inc.—02 June 2014

178. Craftsman Book Company—03 June 2014

179. National Credit Adjusters letter to Consumers re Security Breach—05 June 2014

180. College of the Desert—09 June 2014

181. AT&T Mobility, LLC—10 June 2014

182. Stanford Federal Credit Union—11 June 2014

183. Santa Rosa Memorial Hospital—12 June 2014

184. The Union Labor Life Insurance Company—12 June 2014

185. Ullico, Inc.—12 June 2014

186. AirBorn letter to Consumers (Redacted) re Security Breach—13 June 2014

187. Riverside Community College District—13 June 2014

188. Fidelity National Financial, Inc.—13 June 2014

189. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—16 June 2014

190. David Stanley Dodge—16 June 2014

191. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—17 June 2014

192. Specialized Eye Care—17 June 2014

193. The Metropolitan Companies, Inc. letter to Consumers re Security Breach—18 June 2014

194. Bell Nursery USA, LLC—18 June 2014

195. Papa John's USA, Inc.—19 June 2014

196. Excelitas—19 June 2014

197. Rady Children's Hospital-San Diego—20 June 2014

198. University of California, Washington Center (UCDC)—20 June 2014

199. Primerica—20 June 2014

200. Montana Department of Public Health Human Services letter to Consumers re Security Breach—23 June 2014

201. Safety First—Non MA Notice Template with data elements—23 June 2014

202. MileOne letter to Consumers re Security Breach—23 June 2014

203. Giant Eagle letter to Consumer re Security Breach—23 June 2014

204. Riverside County Regional Medical Center—24 June 2014

205. Butler University letter to Consumers re Security Breach—26 June 2014

206. Sterne, Agee & Leach, Inc.—26 June 2014

207. Legal Sea Foods letter to Consumers re Security Breach—27 June 2014

208. Benjamin F Edwards letter to Consumer re Security Breach—27 June 2014

209. Record Assist letter to Consumers—27 June 2014

210. Invest Financial Corporation—27 June 2014

211. Baltimore School of Massage Therapy—27 June 2014

212. Seterus—27 June 2014

213. Dennis East International, LLC—30 June 2014

214. P.F. Chang's—01 July 2014

215. Thomson Reuters—01 July 2014

216. Wayneburg University—02 July 2014

217. Black Mountain Software—03 July 2014

218. Montana Department of Public Health and Human Services—03 July 2014

219. Watermark Retirement Communities, Inc.—03 July 2014

220. Jiffy Lube—07 July 2014

221. ABM Parking Services, Inc.—08 July 2014

222. AECOM Technology Corporation—08 July 2014

223. Heartland Automotive Services Inc.—08 July 2014

224. TotalBank letter to Consumer re Security Breach—09 July 2014

225. Park Hill School District—10 July 2014

226. Department of Managed Health Care—11 July 2014

227. Davidson Hotel Company LLC d/b/a Davidson Hotels & Resorts—14 July 2014

228. City of Encinitas 7 San Dieguito Water District—15 July 2014

229. Freshology, Inc.—15 July 2014

230. Bank of the West—16 July 2014

231. Bay Area Pain Medical Associates—16 July 2014

232. United Air Temp Conditioning & Heating, Inc.—16 July 2014

233. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—17 July 2014

234. Bank of America—17 July 2014

235. Seattle University—17 July 2014

236. Archdiocese of Portland Ltrt Consumer re Security Breach—18 July 2014

237. Blue Cross Blue Shield of Michigan—18 July 2014

238. Experian letter to Consumer re Security Breach—21 July 2014

239. NRG Assets LLC—21 July 2014

240. Vermont Office of Professional Responsibility Ltrt Consumer—22 July 2014

241. Discover letter One to Consumers re Security Breach—23 July 2014

242. Washington National Insurance Company—23 July 2014

243. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—25 July 2014

244. Managed Med, A Psychological Corporation—25 July 2014

245. NorthShore University Healthsystem—25 July 2014

246. Self Regional Healthcare—25 July 2014

247. Backcountry Gear—28 July 2014

248. Seattle University—28 July 2014

249. Northern Trust—29 July 2014

250. Dreslyn—30 July 2014

251. Lasko Group, Inc.—30 July 2014

252. Oppenheimer Funds letter to Consumers re Security Breach—30 July 2014

253. Reading Partners—30 July 2014

254. The Houstonian Hotel, Club, and Spa—30 July 2014

255. Chicago Yacht Club—31 July 2014

256. Recreational Equipment, Inc.—31 July 2014

257. Signal Outdoor Advertising, LLC—01 August 2014

258. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—04 August 2014

259. Crothall Services Group—04 August 2014

260. Test Effects, LLC—04 August 2014

261. Vibram USA, Inc.—05 August 2014

262. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—06 August 2014

263. Jersey City Medical Center letter to Consumer re Security Breach—06 August 2014

264. Polish Falcons of America—06 August 2014

265. The Dreslyn letter to Consumer re Security Breach—06 August 2014

266. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—07 August 2014

267. Anderson & Murison—07 August 2014

268. Harry Barker letter to Consumers re Security Breach—07 August 2014

269. San Mateo Medical Center—07 August 2014

270. Diatherix Laboratories—08 August 2014

271. St. Francis College letter to Consumers re Security Breach—08 August 2014

272. Freedom Management Group, LLC dba The Natural—12 August 2014

273. Kleiner Perkins Caufield & Byers—12 August 2014

274. The Natural letter to Consumers re Security Breach—14 August 2014

275. Hatchwise.com or eLogoContest.com letter to Consumer re Security Breach—18 August 2014

276. MeeTMe, Inc.—18 August 2014

277. Community Health Systems Professional Services Corporation—20 August 2014

278. M&T Bank—20 August 2014

279. The UPS Store, Inc. on behalf of 51 franchised center locations—20 August 2014

280. Ascensus, Inc.—21 August 2014

281. George Mason letter to Consumer (Redacted) re Security Breach—22 August 2014

282. Liberty Tax—22 August 2014

283. Bimbo Bakeries USA letter to Consumers re Security Breach—26 August 2014

284. Geekface LLC—26 August 2014

285. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—27 August 2014

286. ClamCase LLC letter to Consumer re Security Breach—28 August 2014

287. Xerox State Healthcare, LLC—28 August 2014

288. AB Acquisition LLC (Shaw's)—29 August 2014

289. AltaMed Health Services Corporation—29 August 2014

290. Bartell Hotels—29 August 2014

291. Department of Social Services—29 August 2014

292. LPL Financial LLC—29 August 2014

293. Goodwill Industries International—02 September 2014

294. Goodwill Industries of Sacramento Valley and Northern Nevada, Inc.—02 September 2014

295. LPL Financial LLC—02 September 2014

296. Nationstar Mortgage LLC—02 September 2014

297. Aventura Hospital and Valesco Ventures letter to Consumer re Security Breach—05 September 2014

298. California State University East Bay letter to Consumers re Security Breach—05 September 2014

299. J.P. Morgan Corporate Challenge—05 September 2014

300. Republic Bank & Trust Company—05 September 2014

301. Intuit—06 September 2014

302. Holy Cross Hospital—08 September 2014

303. Yandy.com—08 September 2014

304. Ameriprise Financial Services, Inc.—09 September 2014

305. Cedars-Sinai Health System—10 September 2014

306. County of Napa, Health and Human Services Agency, Comprehensive Services for Older Adults—12 September 2014

307. Tim McCoy & Associates (DBA NEAT Management Group)—15 September 2014

308. CareCentrix, Inc.—18 September 2014

309. Discover letter 1 to Consumers re Security Breach—19 September 2014

310. SELF Loan—19 September 2014

311. Viator letter to Consumer re Security Breach—19 September 2014

312. North American Title Company—22 September 2014

313. Rentrak Corporation—23 September 2014

314. American Express Travel Related Services Company, Inc. and/or its Affiliates (“AXP”)—24 September 2014

315. Jimmy John's Franchises LLC—24 September 2014

316. Pacific Biosciences of California, Inc.—25 September 2014

317. Advantage Funding Company—26 September 2014

318. Bay Area Bioscience Association—26 September 2014

319. Experian—26 September 2014

320. Fidelity Investments—26 September 2014

321. USAA letter to Consumers re UPS Security Breach—26 September 2014

322. Albertson's LLC—29 September 2014

323. Imhoff and Associates, P.C.—29 September 2014

324. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—01 October 2014

325. AT&T letter to Consumers re Security Breach—01 October 2014

326. DHLS letter to Consumers re Security Breach—01 October 2014

327. Flinn Scientific, Inc.—01 October 2014

328. Community Technology Alliance—02 October 2014

329. East West Bank—02 October 2014

330. East West Bank—02 October 2014

331. Touchstone Medical Imaging LLC letter to Consumers 2 re Security Breach—03 October 2014

332. Advanced Data Processing, Inc.—08 October 2014

333. International Dairy Queen, Inc. ("IDQ") on behalf of 9 Dairy Queen franchise locations in California listed in the attached addendum—09 October 2014

334. Penn Highlands Brookville—09 October 2014

335. National Domestic Workers—10 October 2014

336. SAUSALITO YACHT CLUB—10 October 2014

337. University of California Davis Medical Center—13 October 2014

338. GovMint Com letter to Consumers re Security Breach—14 October 2014

339. Pulte Mortgage LLC—14 October 2014

340. Gold's Gym—15 October 2014

341. National Domestic Workers Alliance letter to Consumers re Security Breach—16 October 2014

342. Primerica—16 October 2014

343. Backcountry Gear—17 October 2014

344. Sourcebooks letter to Consumers re Security Breach—17 October 2014

345. Columbia Southern University—20 October 2014

346. Experian—20 October 2014

347. Experian letter To Consumers re Security Breach—22 October 2014

348. The Sinclair Institute letter to Consumers re Security Breach—22 October 2014

349. Alliance Workplace Solutions, LLC—23 October 2014

350. American Soccer Company, Inc.—23 October 2014

351. Reeves International, Inc.—23 October 2014

352. Benefit Express Services—24 October 2014

353. c3controls—24 October 2014

354. Duluth Pack—24 October 2014

355. Fidelity National Financial, Inc.—24 October 2014

356. Capital One letter to Consumers re Security Breach—27 October 2014

357. Direct Learning Systems, Inc., d/b/a 123ce.com—27 October 2014

358. East West Bank-CA Impacted Customers-Kmart Data Breach—27 October 2014

359. Green Energy Training Academy—27 October 2014

360. Modern Gun School—27 October 2014

361. Modern Gun School—27 October 2014

362. The Evolution Store letter to Consumers re Security Breach—27 October 2014

363. Arizona State Retirement System—28 October 2014

364. Cape May-Lewes Ferry—30 October 2014

365. Delaware River & Bay Authority—30 October 2014

366. US Investigations Services, LLC letter Consumer re Security Breach—30 October 2014

367. Anderson & Murison, Inc.—31 October 2014

368. Nationstar Mortgage, LLC d/b/a Champion Mortgage—31 October 2014

369. M&T Bank (Identity Theft)—02 November 2014

370. Camp Bow Wow Franchising, Inc.—03 November 2014

371. Experian—03 November 2014

372. One Love Organics, Inc.—03 November 2014

373. Palm Springs Federal Credit Union—03 November 2014

374. West Publishing Corporation—03 November 2014

375. Nova Southeastern University—06 November 2014

376. Nova Southeastern University—06 November 2014

377. Aarow Equipment & Services, Inc.—07 November 2014

378. Evolution Nature Corp. d/b/a The Evolution Store—07 November 2014

379. Weill Cornell Medical College—07 November 2014

380. EZ Prints, Inc. letter to Consumer re Security Breach—10 November 2014

381. Easter Seals New Hampshire, Inc.—12 November 2014

382. Citibank, N.A.—13 November 2014

383. Visionworks 1st letter to Consumer re Security Breach—13 November 2014

384. REEVE-WOODS EYE CENTER—14 November 2014

385. AHS letter to Consumer re Security Breach—18 November 2014

386. MemberClicks, Inc. d/b/a Moolah Payments—18 November 2014

387. Amgen, Inc. letter to Consumer re Security Breach—19 November 2014

388. Discover letter to Consumers re Security Breach—19 November 2014

389. AlliedBarton Security Services LLC—21 November 2014

390. APi Group, Inc.—21 November 2014

391. Experian—21 November 2014

392. Blue Zebra Sports—24 November 2014

393. Cultivian Ventures, LLC—24 November 2014

394. Fairway Independent Mortgage Corporation—24 November 2014

395. Visionworks 2nd letter to Consumer re Security Breach—24 November 2014

396. Form—25 November 2014

397. New Hampshire Employment Security—25 November 2014

398. Simms Fishing Products letter to Consumers re Security Breach—25 November 2014

399. State Compensation Insurance Fund—25 November 2014

400. Calypso St. Barth letter to Consumer re Security Breach—26 November 2014

401. Highlands-Cashiers Hospital—26 November 2014

402. Shutterfly, Inc.—26 November 2014

403. Holiday Motel letter to Consumer re Security Breach—28 November 2014

404. American Residuals and Talent, Inc. (ART) letter to Consumer re Security Breach—01 December 2014

405. Big East Conference—01 December 2014

406. Blue Mountain Community Foundation—01 December 2014

407. Godiva Chocolatier, Inc.—01 December 2014

408. Highlands-Cashiers Hospital—01 December 2014

409. Bebe Stores, Inc.—05 December 2014

410. Econolight501 General Proofs—05 December 2014

411. Sands Casino Resort Bethlehem—05 December 2014

412. AHS letter to Consumers re Security Breach—09 December 2014

413. Seterus—09 December 2014

414. EMCOR Services Mesa Energy Systems—11 December 2014

415. ABM Parking Services—12 December 2014

416. Acosta, Inc. and its subsidiaries, including Mosaic Sales Solutions US Operating Co. LLC—12 December 2014

417. Clay County Hospital—12 December 2014

418. University of California, Berkeley—12 December 2014

419. Apple Leisure Group and AMResorts—15 December 2014

420. Point Loma Nazarene University—15 December 2014

421. Valplast Supply Services, Inc. letter to Consumer re Security Breach—16 December 2014

422. Ascena Retail Group, Inc.—17 December 2014

423. Harmonic Inc.—18 December 2014

424. American Express Travel Related Services Company, Inc. and/or its Affiliates ("AXP")—19 December 2014

425. Mercy Medical Center Redding Oncology Clinic—19 December 2014

426. Presidian Hotels & Resorts—19 December 2014

427. Quest Diagnostics—19 December 2014

428. Staples, Inc.—19 December 2014

429. BolderImage SBN to Consumers—20 December 2014

430. Azusa Pacific University—22 December 2014

431. ID Parts LLC letter to Consumers—22 December 2014

432. Nvidia Corporation—22 December 2014

433. DutchWear—23 December 2014

434. Public Architecture—23 December 2014

435. Rob Kirby, CPA—23 December 2014

436. Transamerica Premier Life Insurance Company—23 December 2014

437. Corday Productions, Inc.—24 December 2014

438. Lokai Holdings LLC—24 December 2014

439. Allianz Life Insurance Company of North America—26 December 2014

440. Empi, Inc./DJO, LLC—26 December 2014

441. Physicians Skin and Weight Centers, Inc.—26 December 2014

442. Six Red Marbles—26 December 2014

443. Stagecoach Transportation, Inc. SBN to Consumer—December 26, 2014

444. Fast Forward Academy, LLC—30 December 2014

445. La Jolla Group—31 December 2014

Appendixes

Chapter 5

Appendix I. Data sources

ITRC current data sources (as of 28th February 2017)

California Attorney General's Office	*letters already available in 2014*
Maryland Attorney General's Office	*letters already available in 2014*
New Hampshire Department of Justice	*letters already available in 2014*
Vermont Attorney General's Office	*letters already available in 2014*
Health & Human Services (HHS.gov)	*sectoral DB*
HIPAA Journal	*sectoral DB*
www.databreaches.net	*multisectoral DB*
Maine Attorney General's Office	*No letters available only list of breaches*
Indiana Attorney General's Office	*No letters available only list of breaches*
Montana Attorney General's Office	*from mid-2015 letters available*
Oregon Attorney General's Office	*from 2016 letters available*
Washington Attorney General's Office	*from mid-2015 letters available*

Data breach databases websites

http://veriscommunity.net/vcdb.html

http://www.idtheftcenter.org

https://www.privacyrights.org

Attorney General websites accessed for notification downloads

https://oag.ca.gov/ecrime/databreach/list

http://www.oag.state.md.us/idtheft/businessGL.htm

http://doj.nh.gov/consumer/security-breaches/

http://www.atg.state.vt.us/issues/consumer-protection/privacy-and-data-security/vermont-security-breaches.php

Appendix II. Regression Diagnostics

The residuals versus predicted, and observed versus predicted plots for the breach-count model are as follows. (As a reminder, the McFadden pseudo R-square is 0.23).

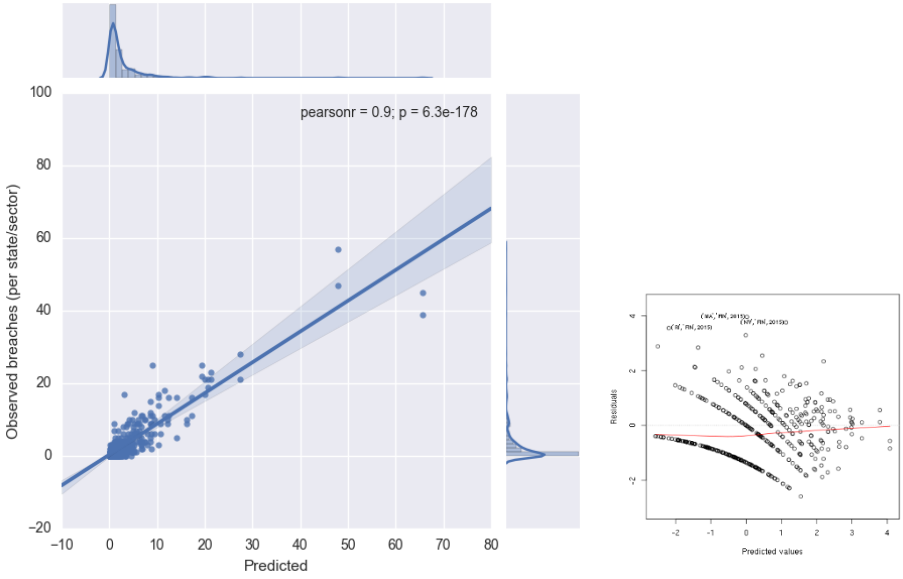


Figure A.1 Breach count model diagnostic plots

The variance inflation factor is between 1 and 2 for all the variables, showing no multicollinearity.

Outliers and datasets with reduced sectors. The top outliers are from the finance sector. This reflects the fact that financial firms are mostly headquartered in specific U.S. states due to tax laws and such, and that ITRC records the place of breach for large companies to the headquarters. This is a systematic bias, not a measurement error that might be corrected by removing outliers. We can however run the regression on a reduced dataset that excludes the multistate firms—e.g., without the

financial sector, and the business sector. The results are presented below. The direction of the coefficients remains the same, indicating robust results.

	Dependent variable:		
	breaches		
	(1)	(2)	(3)
inform_agp	0.361**	0.292**	0.247
	(0.149)	(0.142)	(0.164)
inform_agnp	0.044	-0.040	-0.124
	(0.101)	(0.098)	(0.115)
inform_credit	0.289***	0.250***	0.215**
	(0.094)	(0.091)	(0.107)
penalty_cap	-0.053	-0.063	-0.077
	(0.085)	(0.082)	(0.097)
priv_cause	0.119	0.174*	0.165
	(0.094)	(0.091)	(0.108)
risk_harm	-0.231**	-0.234**	0.017
	(0.104)	(0.099)	(0.117)
fin	-8.587***		
	(0.161)		
med	-7.777***	-7.726***	-7.842***
	(0.134)	(0.126)	(0.142)
edu	-7.481***	-7.439***	-7.536***
	(0.162)	(0.154)	(0.167)
gov	-9.218***	-9.179***	-9.287***
	(0.154)	(0.146)	(0.159)
bso	-9.963***	-9.901***	
	(0.135)	(0.127)	
Observations	478	382	286
Log Likelihood	-802.491	-677.914	-463.575
theta	5.171*** (1.196)	8.250*** (2.417)	12.175*** (6.001)
Akaike Inf. Crit.	1,626.981	1,375.828	945.150

Note: Negative binomial regression, with sector size as offset, in datasets with all sectors, and all excluding finance, and business.
AICs cannot be compared as the datasets differ.
*p<0.1; **p<0.05; ***p<0.01

Appendix III. Alternative breach count models

We present two alternative model specifications: adding a dummy variable for the breach year, and adding an interaction term. Comparing the Akaike Information Criteria of these models against our simple model, both are slightly worse, suggesting that the year dummy or the interaction terms do not add to the preferred parsimonious model.

	Dependent variable:		
	breaches		
	(1)	(2)	(3)
inform_agp	0.361**	0.360**	0.371
	(0.149)	(0.149)	(0.257)
inform_agnp	0.044	0.044	0.102
	(0.101)	(0.101)	(0.174)
inform_credit	0.289***	0.289***	0.316*
	(0.094)	(0.094)	(0.161)
penalty_cap	-0.053	-0.054	-0.001
	(0.085)	(0.085)	(0.145)
priv_cause	0.119	0.119	0.193
	(0.094)	(0.094)	(0.157)
risk_harm	-0.231**	-0.232**	-0.675***
	(0.104)	(0.104)	(0.175)
fin	1.376***	1.376***	-9.118***
	(0.136)	(0.136)	(0.386)
med	2.186***	2.187***	-7.793***
	(0.098)	(0.098)	(0.201)
edu	2.482***	2.483***	-7.148***
	(0.135)	(0.135)	(0.342)
gov	0.745***	0.746***	-9.651***
	(0.126)	(0.126)	(0.313)
y2015		0.027	
		(0.078)	
bso			-9.747***
			(0.203)
inform_agp:fin			0.466
			(0.528)
inform_agp:med			-0.121
			(0.363)
inform_agp:edu			-0.054
			(0.476)
inform_agp:gov			-0.129
			(0.437)
inform_agp:bso			
inform_agnp:fin			0.559
			(0.350)
inform_agnp:med			-0.136
			(0.239)
inform_agnp:edu			-0.207

			(0.344)
inform_agnp:gov			-0.447
			(0.315)
inform_agnp:bso			
inform_credit:fin			0.346
			(0.363)
inform_credit:med			-0.091
			(0.221)
inform_credit:edu			-0.461
			(0.313)
inform_credit:gov			0.133
			(0.299)
inform_credit:bso			
penalty_cap:fin			0.027
			(0.297)
penalty_cap:med			-0.029
			(0.201)
penalty_cap:edu			-0.082
			(0.286)
penalty_cap:gov			-0.239
			(0.266)
penalty_cap:bso			
priv_cause:fin			-0.418
			(0.329)
priv_cause:med			-0.158
			(0.221)
priv_cause:edu			-0.186
			(0.324)
priv_cause:gov			0.359
			(0.290)
priv_cause:bso			
risk_harm:fin			0.427
			(0.365)
risk_harm:med			0.600**
			(0.246)
risk_harm:edu			0.588*
			(0.344)
risk_harm:gov			0.988***
			(0.318)
risk_harm:bso			
Constant	-9.963***	-9.976***	
	(0.135)	(0.140)	
Observations	478	478	478
Log Likelihood	-802.491	-802.433	-785.665
theta	5.171*** (1.196)	5.179*** (1.197)	6.639*** (1.734)
Akaike Inf. Crit.	1,626.981	1,628.867	1,641.330

Note: Negative binomial regression, with sector size as offset.
*p<0.1; **p<0.05; ***p<0.01

Appendix IV. Models with additional state-level controls

We have tested a number of alternative models that include state-level controls. The idea behind them was to control for differences among states that might affect the number of breaches reported (other than the DBNL provisions and sector sizes that we included). We tested the following:

- *Household income* (from the U.S. Census) as a proxy for states' wealth. Most directly, richer individuals might be more interesting targets for identity theft. Additionally, household income is highly correlated with gdp-per-capita, which also reflects wealthier companies, that might have more resources to invest in cybersecurity, and generally, better overall infrastructure.
- We look at *crime rates* (from the Internet Crime Compliant Center report) in the categories of crimes that could also be causes of data-breaches, and all categories of crimes. This could reflect the prevalence of crime in a state, leading to insider breaches or physical theft. It can alternatively also reflect how often citizens report crimes in a state.
- We look at the *centralization of sectors* in various states, e.g., the number of banks per capita, by dividing the number of firms in each sector by the state's population (from the U.S. Census).

We did not find general attributes on cybersecurity investment in states, or business attitudes to risk across U.S. states, that could be interesting controls. Another common choice is to add one dummy variable for each state, but adding 48 dummies does not make sense in a dataset with 480 observations.

The results are provided in the table below. Overall, they offer little improvement over our base model in terms of AIC, and also do not change the sign of the coefficients, indicating our existing models are robust.

```
================================================================================
                                        Dependent variable:
                    ---------------------------------------------------
                                             breaches
                        (1)             (2)             (3)             (4)
```

inform_agp	0.128	0.543***	0.528***	0.335**
	(0.160)	(0.184)	(0.174)	(0.149)
inform_agnp	-0.062	0.123	0.110	0.019
	(0.104)	(0.112)	(0.107)	(0.102)
inform_credit	0.258***	0.262***	0.270***	0.304***
	(0.093)	(0.094)	(0.093)	(0.094)
penalty_cap	-0.042	-0.044	-0.032	-0.055
	(0.084)	(0.084)	(0.085)	(0.085)
priv_cause	0.108	0.072	0.063	0.133
	(0.094)	(0.098)	(0.098)	(0.094)
risk_harm	-0.191*	-0.345***	-0.367***	-0.236**
	(0.103)	(0.125)	(0.128)	(0.103)
fin	1.370***	1.390***	1.392***	2.021***
	(0.135)	(0.135)	(0.135)	(0.398)
med	2.193***	2.185***	2.186***	2.779***
	(0.097)	(0.097)	(0.097)	(0.359)
edu	2.471***	2.489***	2.490***	3.147***
	(0.134)	(0.134)	(0.134)	(0.410)
gov	0.754***	0.753***	0.755***	1.337***
	(0.125)	(0.126)	(0.126)	(0.368)
house_income	0.00002***			
	(0.00001)			
victim_cause		-0.0001*		
		(0.0001)		
pop2012			-0.000*	
			(0.000)	
orgs_p100				0.406*
				(0.235)
Constant	-10.934***	-9.812***	-9.779***	-10.639***
	(0.299)	(0.162)	(0.167)	(0.419)
Observations	478	478	478	478

```
Log Likelihood        -796.066         -801.131         -800.855         -
801.033
theta                5.441*** (1.264) 5.302*** (1.247) 5.301*** (1.245)
5.226*** (1.206)
Akaike Inf. Crit.      1,616.133           1,626.263           1,625.709
1,626.067
=========================================================================
==============
Note: Negative binomial regression, with sector size (organiza-
tions) as offset.
      AIC of model without controls: 1627.0. *p<0.1; **p<0.05; ***p<0.01
```

Appendix V. Alternate time models

```
=========================================================================
==============
                                    Dependent variable:
                    ---------------------------------------------------
--------------
                                    notification_time
                        (1)             (2)             (3)             (4)
-------------------------------------------------------------------------
--------------
hacking                   0.130                                     0.241*
0.254*
                        (0.126)                                   (0.134)
(0.134)                                                           physical
0.068                                   0.020           0.017
                        (0.158)                                   (0.162)
(0.161)
insider                   0.124                                     0.116
0.137
                        (0.194)                                   (0.196)
(0.198)
unintended

bso                                             0.058               0.006
0.127
                                        (0.230)                   (0.238)
(0.241)
fin                                            -0.082              -0.080
0.022
                                        (0.263)                   (0.262)
(0.262)
edu                                             0.115               0.087
0.146
                                        (0.295)                   (0.298)
(0.296)
med                                             0.419*              0.467*
0.448*
                                        (0.243)                   (0.248)
(0.245)
gov

CA                                                                        -
0.070
```

```
(0.125)
MD                                                                  -
0.374**

(0.149)
NH                                                                  -
0.308*

(0.173)
VT                                                                  -
0.342**

(0.144)
multi                                                           0.017
0.552**
                                                          (0.113)
(0.234)
Constant                  3.701***           3.649***           3.529***
3.728***
                         (0.101)             (0.219)             (0.223)
(0.240)
-------------------------------------------------------------------------
--------------
Observations               260                260                260
260
Log Likelihood          -1,237.540         -1,232.446         -1,230.251
-1,225.652
theta                   1.498*** (0.126)   1.552*** (0.131)   1.575*** (0.133)
1.626*** (0.138)
Akaike Inf. Crit.        2,483.081          2,474.891          2,478.502
2,477.305
=========================================================================
==============
Note: Negative binomial regression. AIC/LL of null model is 2478.2/-
1237.1.
      p<0.1; **p<0.05; ***p<0.01
```

```
================================================================================
                                        Dependent variable:
                    ---------------------------------------------------
                                     uninformed_exposure_time
                          (1)             (2)             (3)             (4)
--------------------------------------------------------------------------------
hacking                  0.412**                                         0.537***
0.623***
                        (0.170)                                         (0.183)
(0.183)
physical               -0.660***                       -0.588***          -
0.430**
                        (0.201)                                         (0.207)
(0.207)
insider                  1.689***                                        1.594***
1.792***
                        (0.247)                                         (0.252)
(0.251)
unintended

bso                                      0.238           -0.342           -
0.552
                                         (0.357)                         (0.340)
(0.343)
fin                                      -0.570          -0.747**         -
0.855**
                                         (0.392)                         (0.352)
(0.353)
edu                                      0.216           -0.268           -
0.493
                                         (0.524)                         (0.467)
(0.467)
med                                                0.709*                  0.318
0.189
                                         (0.380)                         (0.351)
(0.349)
gov

CA                                                                        0.277*

(0.167)
MD                                                                         0.278

(0.190)
NH                                                                       0.587***

(0.200)
VT                                                                          -
0.019

(0.187)
multi                                                    0.258*           -
0.230
                                                          (0.142)
(0.277)
```

```
Constant                  4.595***           4.795***           4.629***
4.381***
                         (0.137)            (0.343)            (0.298)
(0.322)
------------------------------------------------------------------------
--------------
Observations               257                257                257
257
Log Likelihood          -1,496.118         -1,531.416         -1,483.143
-1,478.444
theta                  0.877*** (0.068)   0.713*** (0.054)   0.948*** (0.074)
0.976*** (0.077)
Akaike Inf. Crit.        3,000.235          3,072.833          2,984.286
2,982.888
========================================================================
==============
Note: Negative binomial regression. AIC/LL of null model is 3088/-1542.
      *p<0.1; **p<0.05; ***p<0.01
```

Chapter 6

Appendix I. Identity Theft vs. Breached Records

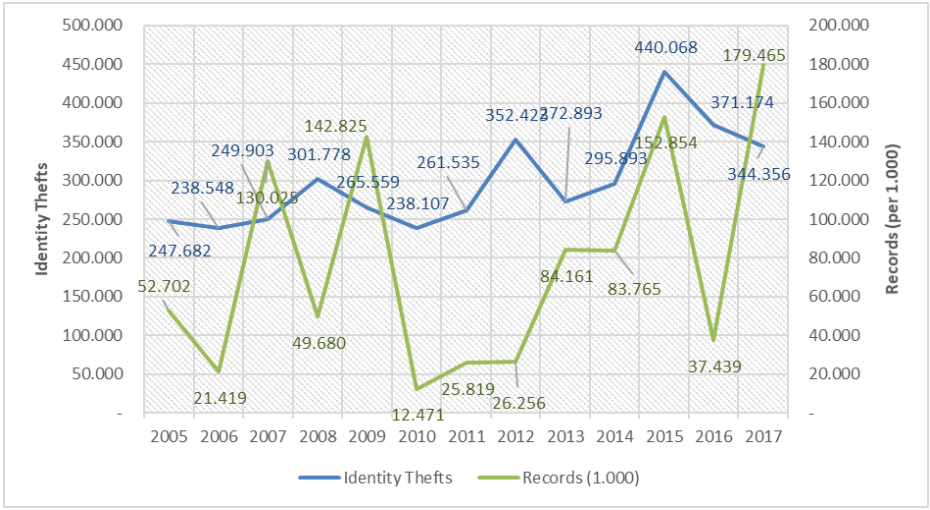


Figure A.2 Identity theft and breached records 2005-2017

In the above Figure we compare identity thefts with the number of available records breached in data breaches.

Appendix II. Difference in Differences Summary

Generalized Linear Model Regression Results

Dep. Variable:	**breaches**	No. Observations:	650
Model:	GLM	Df Residuals:	587
Model Family:	NegativeBinomial	Df Model:	62
Link Function:	log	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1929.6
Date:	Tue, 10 Sep 2019	Deviance:	159.72
Time:	13:05:18	Pearson chi2:	126.
No. Iterations:	9		
Covariance Type:	nonrobust		

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.0892	0.408	-2.672	0.008	-1.888	-0.290
st[T.AL]	1.3936	0.496	2.808	0.005	0.421	2.367
st[T.AR]	0.7319	0.489	1.497	0.134	-0.226	1.690
st[T.AZ]	1.6743	0.470	3.562	0.000	0.753	2.596
st[T.CA]	3.9716	0.462	8.601	0.000	3.067	4.877
st[T.CO]	2.2532	0.466	4.835	0.000	1.340	3.167
st[T.CT]	1.9302	0.469	4.120	0.000	1.012	2.849
st[T.DE]	0.2864	0.503	0.569	0.569	-0.700	1.272
st[T.FL]	2.9928	0.464	6.453	0.000	2.084	3.902
st[T.GA]	2.6037	0.465	5.595	0.000	1.692	3.516
st[T.HI]	0.3784	0.498	0.760	0.447	-0.598	1.354
st[T.IA]	1.2670	0.475	2.667	0.008	0.336	2.198
st[T.ID]	0.2149	0.505	0.425	0.671	-0.775	1.205
st[T.IL]	2.7511	0.463	5.939	0.000	1.843	3.659
st[T.IN]	2.1382	0.467	4.581	0.000	1.223	3.053
st[T.KS]	0.8172	0.485	1.685	0.092	-0.133	1.768
st[T.KY]	1.6015	0.477	3.354	0.001	0.666	2.537
st[T.LA]	1.0198	0.481	2.118	0.034	0.076	1.963
st[T.MA]	2.5885	0.462	5.600	0.000	1.683	3.495
st[T.MD]	2.1207	0.465	4.559	0.000	1.209	3.032
st[T.ME]	0.6912	0.489	1.413	0.158	-0.268	1.650
st[T.MI]	1.9587	0.467	4.193	0.000	1.043	2.874
st[T.MN]	1.9144	0.469	4.085	0.000	0.996	2.833
st[T.MO]	1.5857	0.470	3.371	0.001	0.664	2.508
st[T.MS]	0.2054	0.505	0.407	0.684	-0.784	1.195
st[T.MT]	0.7114	0.489	1.456	0.145	-0.246	1.669
st[T.NC]	2.4072	0.465	5.177	0.000	1.496	3.319
st[T.ND]	-0.6612	0.558	-1.185	0.236	-1.755	0.433
st[T.NE]	0.8085	0.486	1.663	0.096	-0.144	1.761
st[T.NH]	1.1493	0.478	2.405	0.016	0.213	2.086
st[T.NJ]	2.1932	0.466	4.702	0.000	1.279	3.107
st[T.NM]	0.8111	0.502	1.614	0.106	-0.174	1.796
st[T.NV]	1.1871	0.478	2.482	0.013	0.250	2.124
st[T.NY]	3.4484	0.461	7.479	0.000	2.545	4.352
st[T.OH]	2.6571	0.464	5.730	0.000	1.748	3.566
st[T.OK]	1.1637	0.477	2.440	0.015	0.229	2.098
st[T.OR]	1.8130	0.468	3.874	0.000	0.896	2.730
st[T.PA]	2.5388	0.464	5.468	0.000	1.629	3.449
st[T.RI]	0.9043	0.484	1.869	0.062	-0.044	1.853
st[T.SC]	1.2044	0.476	2.530	0.011	0.271	2.138
st[T.SD]	-0.3078	0.554	-0.556	0.578	-1.393	0.777
st[T.TN]	2.0570	0.469	4.389	0.000	1.138	2.976
st[T.TX]	3.1904	0.460	6.938	0.000	2.289	4.092
st[T.UT]	1.2570	0.476	2.641	0.008	0.324	2.190
st[T.VA]	2.4042	0.463	5.190	0.000	1.496	3.312
st[T.VT]	0.7576	0.486	1.558	0.119	-0.196	1.711
st[T.WA]	2.1432	0.468	4.579	0.000	1.226	3.061
st[T.WI]	1.5795	0.472	3.344	0.001	0.654	2.505
st[T.WV]	0.0718	0.510	0.141	0.888	-0.928	1.071
st[T.WY]	-0.5431	0.548	-0.992	0.321	-1.616	0.530
ys[T.2006]	1.2275	0.257	4.785	0.000	0.725	1.730
ys[T.2007]	1.3148	0.266	4.938	0.000	0.793	1.837
ys[T.2008]	1.0217	0.280	3.649	0.000	0.473	1.570
ys[T.2009]	0.7473	0.293	2.552	0.011	0.173	1.321
ys[T.2010]	1.5495	0.286	5.416	0.000	0.989	2.110

ys[T.2011]	1.5587	0.288	5.417	0.000	0.995	2.123
ys[T.2012]	1.5507	0.288	5.388	0.000	0.987	2.115
ys[T.2013]	1.4918	0.288	5.177	0.000	0.927	2.057
ys[T.2014]	1.8052	0.289	6.249	0.000	1.239	2.371
ys[T.2015]	1.7391	0.289	6.014	0.000	1.172	2.306
ys[T.2016]	2.1162	0.288	7.358	0.000	1.552	2.680
ys[T.2017]	2.5281	0.289	8.759	0.000	1.962	3.094
hasdbnl	**0.0173**	**0.203**	0.085	0.932	-0.381	0.415

Appendix III. Stan Code & Convergence Details

Multilevel Poisson Model with Varying Intercepts Per State/Year and Offset (for Data Breaches):

```
data {
    int<lower=1> nY;
    int<lower=1> nS;
    int<lower=1> nP;  // number of (individual) predictors
    matrix[nY*nS, nP] X; // predictors (e.g., dbnl enactment, revisions)
    vector[nY*nS] offset;  // a rate, has the coef set to 1
    int yy[nY*nS];
    int ss[nY*nS];
    int<lower=0> Y[nY*nS];  // outcome/observations
}
transformed data {
    int N = nY * nS;
}
parameters {
    real alpha; // overall intercept
    vector[nY] a_yy;  // unique intercept (poisson level) per year
    vector[nS] a_ss;  // unique intercept per state
    real<lower=0> sigma_y; // pool unique YY intercepts
    real<lower=0> sigma_s; // pool unique SS intercepts
    vector[nP] beta;  // beta for all predictors
}
transformed parameters {}
model {
  vector[N] mu;
  // priors
  target += normal_lpdf(alpha | 0, 10);
  target += normal_lpdf(beta | 0, 1);
  target += normal_lpdf(a_yy | 0, sigma_y);
  target += normal_lpdf(a_ss | 0, sigma_s);
  target += cauchy_lpdf(sigma_y | 0 , 1);
  target += cauchy_lpdf(sigma_s | 0 , 1);
  // linear model
  for ( i in 1:N )
      mu[i] = alpha + a_yy[yy[i]] + a_ss[ss[i]] + X[i] * beta + offset[i];
  target += poisson_log_lpmf(Y | mu);
}
generated quantities {
    vector[N] yhat;
```

```
    vector[N] log_lik;
    for ( i in 1:N ) {
        real mu;
        mu = alpha + a_yy[yy[i]] + a_ss[ss[i]] + X[i] * beta + offset[i];
        mu = fmin(mu, 20.7944);  // max for poisson;
        yhat[i] = poisson_log_rng(mu);
        log_lik[i] = poisson_log_lpmf(Y[i] | mu);
    }
}
```

Multilevel Log-Normal Model with Varying Intercepts Per State/Year and Offset (for Identity Theft):

```
data {
    int nY;
    int nS;
    int nP;  // number of (individual) predictors
    matrix[nY*nS, nP] X; // predictors, e.g., laws, etc.
    vector[nY*nS] offset;  // a rate, has the coef set to 1 (should be
logged)
    int yy[nY*nS];
    int ss[nY*nS];
    real<lower=0> Y[nY*nS];  // outcome/observations
}
transformed data {
    int N = nY * nS;
}
parameters {
    real alpha; // overall intercept
    vector[nY] a_yy;  // unique intercept per year
    vector[nS] a_ss;  // unique intercept per state
    real<lower=0> sigma_y; // pool unique YY intercepts
    real<lower=0> sigma_s; // pool unique SS intercepts
    vector[nP] beta;  // beta for all predictors
    vector<lower=0>[nS] sigma_l; // log normal sigma (per state).
}
transformed parameters {}
model {
  vector[N] mu;
  vector[N] sigma;
  // priors
  target += normal_lpdf(alpha | 0, 10);
  target += normal_lpdf(beta | 0, 10);
  target += normal_lpdf(a_yy | 0, sigma_y);
  target += normal_lpdf(a_ss | 0, sigma_s);
  target += cauchy_lpdf(sigma_y | 0, 1);
  target += cauchy_lpdf(sigma_s | 0, 1);
  target += exponential_lpdf(sigma_l | 2);  // tighter (re lognorm)
  // linear model
  for ( i in 1:N ) {
      mu[i] = alpha + a_yy[yy[i]] + a_ss[ss[i]] + X[i] * beta + offset[i];
```

```
      sigma[i] = sigma_l[ss[i]];
  }
  target += lognormal_lpdf(Y | mu, sigma);
}
generated quantities {
    vector[N] yhat;
    vector[N] log_lik;
    for ( i in 1:N ) {
        real mu;
        real sigma;
        mu = alpha + a_yy[yy[i]] + a_ss[ss[i]] + X[i] * beta + offset[i];
        sigma = sigma_l[ss[i]];
        yhat[i] = lognormal_rng(mu, sigma);
        log_lik[i] = lognormal_lpdf(Y[i] | mu, sigma);
    }
}
```

The Bayesian chains converge well: the Gelman-Rubin statistic (***rhats* are equal to 1±0.005**) and Stan gives no serious warnings. A complementary posterior predictive plot is shown below (next to the one in the text). The observed *ys* fall within the light blue posterior predictive band, indicating a reasonable fit.

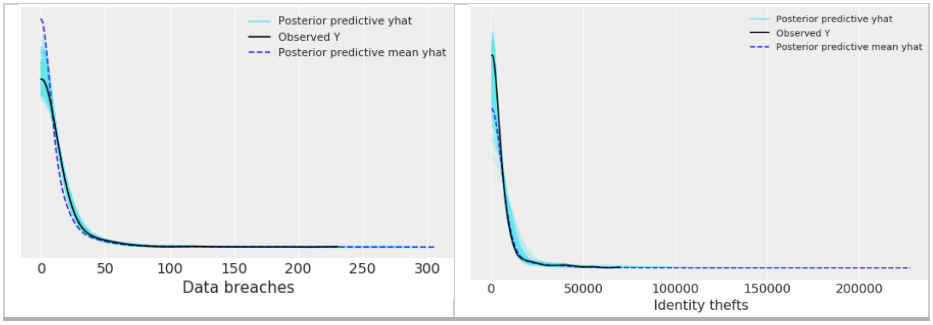


Figure A.3 Posterior predictive plots based on the y/y_hat distribution. Left: data breach model; right: identity theft model.

Appendix IV. Unique Year & State Intercepts

Year Intercepts. The model estimates unique intercepts for each state and year. When a unique intercept's credible interval is around zero, it can be interpreted as random noise. In the data breach model, the year intercepts for 2005 and 2009 are below zero, and for 2006, 2007, and 2010 above zero. These intercepts are what remains after detrending (via y_trend), and they point to unknown influences on breach levels in those years.

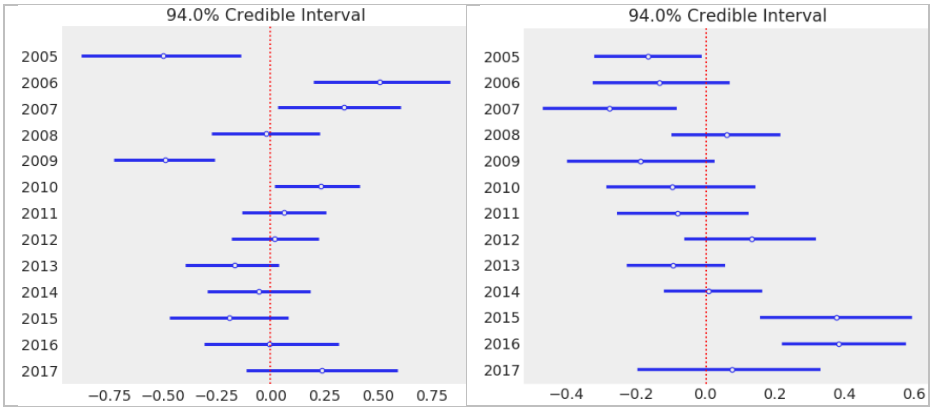


Figure A.4 Unique year intercepts. Left: data breach model. Right: identity theft model.

As before, the unique effect be estimated using $e^{mid\text{-}point}$.

State Intercepts. In both multi-level models, the state intercepts for a number of states differ from the baseline. This reflects differences that remain after controlling for the state size and regulatory and control predictors.

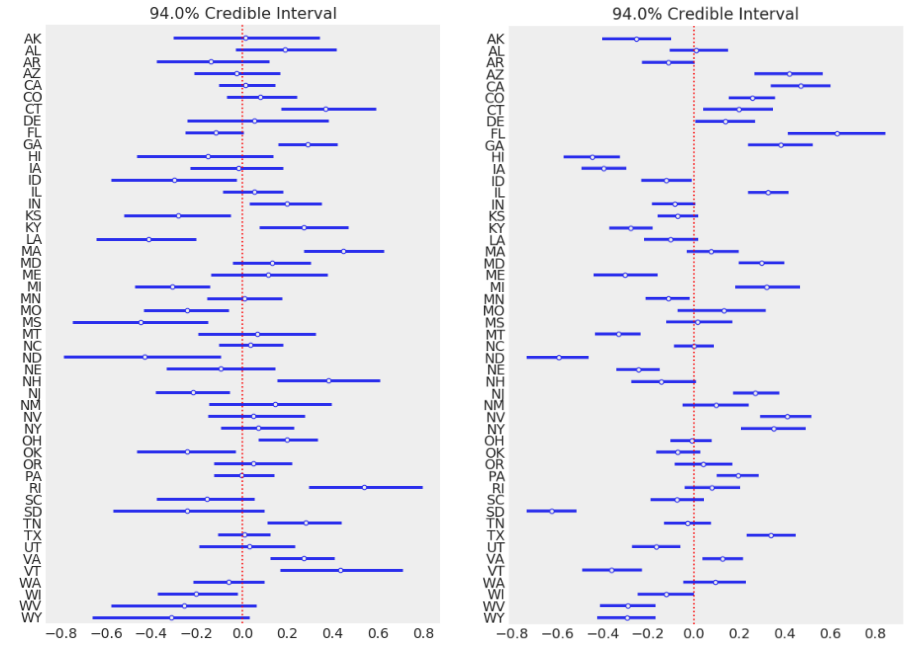


Figure A.5 Unique state Intercepts. Left: data breach model. Right: identity theft model.

List of Publications

Bisogni, F., Cavallini S. & Trocchio S. (2011). Cybersecurity at European Level: The Role of Information Availability. Communications & Strategies, Number 81 - The Economics of Cybersecurity, pp.105-123, March 2011.

Bisogni, F. (2013). Evaluating Data Breach Notification Laws - What Do the Numbers Tell Us? TPRC 2013, Virginia, September 2013.

Bisogni F. (2015). Data Breaches and the Dilemmas in Notifying Customers. WEIS 2015 - 14th annual Workshop on the Economics of Information Security, At Delft University of Technology, the Netherland, June 2015.

Bisogni F. (2016). Proving Limits of State Data Breach Notification Laws: Is a Federal Law the Most Adequate Solution? Journal of Information Policy Vol. 6 (2016), pp. 154-205 - Penn State University PressConference.

Bisogni F., Asghari H., & Van Eeten M. (2017). Estimating the size of the iceberg from its tip. An investigation into unreported data breach notifications. WEIS 2017 - 16th annual Workshop on the Economics of Information Security, At La Jolla, US, June 2017.

Bisogni F., & Asghari H. (2020). More than a suspect. An investigation into the connection between data breaches, identity thefts and data breach notification laws. Journal of Information Policy Vol. 10 (2020), pp.45-82 - Penn State University PressConference.

ENISA (2012). Economics of Security: Facing the Challenges. A multidisciplinary assessment. Sections: Economic Incentives for Security: The role of information asymmetry and lack of information; Economic impact of intervention policies.

ENISA (2013). Position Paper of the EP3R Task Forces on Trusted Information Sharing. October 2013.

Felici M., Wainwright N., Bisogni F., & Cavallini S. (2015). What's New in the Economics of Cybersecurity?: Observational and Empirical Studies. Editorial, IEEE Security & Privacy, 2015.

Felici M., Wainwright N., Cavallini S., & Bisogni F. (2016). What's New in the Economics of Cybersecurity?: Observational and Empirical Studies. Editorial,-IEEE Security & Privacy, 2016.

Authorship contribution

The five core chapters of this thesis are based on peer-reviewed publications which are the result of collaborative work with a variety of co-authors. I am the sole author of two studies and the lead author on three. I was fortunate to receive valuable contributions from several colleagues. Below, I summarise their main contributions to each paper.

In the communications and strategies study, the overall methodology is indebted to ideas generated by Simona Cavallini, who provided significant help regarding the approach, visualising the graphs and drafting the text. Most of the co-authors contributed to the approach and analysis, as well as improving the text of the manuscript.

The WEIS paper on estimating the size of the iceberg benefited from the contribution of Hadi Asghari and Michel van Eeten. The estimation of hidden data breaches was carried out and written by Hadi, and Michel revised the final text. All co-authors contributed significantly to the approach and quality of the paper.

For the JIP paper, Hadi Asghari contributed important ideas on fleshing out the methodology needed to answer the central question, as well as extensively helping to clarify the overall argument and drafting and improving the text. He also conducted the analysis on the impact of DBNLs on data breaches and identity theft.

Finally, for the remaining two studies that I wrote alone, Michel van Eeten significantly contributed in conceptualising the main ideas during my visits in Delft.

About the Author

Fabio Bisogni was born in Rome, Italy, in 1980. Before beginning his doctoral studies at the Delft University of Technology, he completed a Master's Degree in Economics and a Professional Master's Degree in Governance of Information Systems from ROMA TRE University.

Bisogni previously gained research experience at the Fraunhofer Gesellschaft, in consultancy at Ernst&Young and in the industrial sector in the FIAT group (now FCA). He is currently the President of FORMIT – Fondazione per la Ricerca sulla Migrazione e Integrazione delle Tecnologie – (Italy), where he has carried out more than 20 studies and projects in the fields of critical infrastructure, cybersecurity, economics of information security and policing over the last 10 years. Since 2018, he has served as Vice President and Delegate for Educational Offering and Communication at Università degli Studi Internazionali di Roma (UNINT). He is a chartered certified accountant and financial auditor, and is also currently an expert for the European Commission.

Bisogni was appointed Member of the Italian Presidency of the Council of Ministers working group on cyberspace security ('Gruppo di studio per la sicurezza dell'utilizzo dello spazio cibernetico'). He acted as an external expert for the European Union Agency for Network and Information Security (ENISA) Working Group on Economics of Security, and for the European Public-Private Partnership for Resilience (EP3R) Task Force on Trusted Information Sharing (TF-TIS). He was also a guest co-editor of two special issues of the IEEE Security & Privacy on Economics of Cybersecurity.

He has significant international experience, having worked and lived in Italy, Germany, the UK, the Russian Federation, the Middle East and

Southeast Asia while conducting activities within European Commission and United Nations projects.

Outside of work, he enjoys following AS Roma ups and downs and kitesurfing. He is an International Kiteboarding Organization (IKO) instructor.

Non ragioniam di lor,
ma guarda e passa.

Let us not talk of them,
but look and pass.

Dante Alighieri
Divina Commedia, Inferno, Canto III

Data breach notification laws require any business that suffers a data breach, or believes that it suffered a data breach, to notify customers. Such laws offer incentives to the party who owes the notification duty to minimise the number of triggering events and also enable the affected third parties to diminish the consequences, namely identity theft, and to make prudent choices in the future.

Public policy that seeks to improve the effects of data breach notification legislation must be informed by comprehensive understanding of the behaviour and incentives of the organisations and individuals involved in the notification flow. Thus, this book answers the following research question:
"What are the effects of the provisions of data breach notification laws on (1) communications issued by breached organisations to their customers; (2) the timing of breach detection and reaction; (3) the number of data breaches reported; and (4) the volume of identity theft stemming from data breaches?"

This book reflects thefore on the role of disclosure policies in the information security arena and on the implications, given the results of the conducted studies, for European data breach notification policies.