Information Lost

(Apologies to Milton)

Will the 'paradise' that information promises, to both consumer and firm, be 'lost' on account of data breaches? The epic is playing out.

Catherine L. Mann
Rosenberg Professor of Global Finance, Brandeis University
CLMann@Brandeis.edu
www.CLMann.com

NBER Conference
Economics of Digitization: An Agenda
June 6-7, 2013

This version: May 31, 2013

Abstract

As businesses and consumers search, communicate, and transact on-line, firms gather more and more personal and financial information. On the one hand, all this information can enhance market efficiency and consumer surplus as firms tailor products to buyers. On the other hand, there is increased risk of information loss, either by accident or through theft. What issues should be on the digital agenda with regard to information loss, and what data are available to underpin both business response and any policy approach? This paper reviews the situation and points out where we need more thought and more data. Topics include: (1) Frameworks for analysis : How should we model the information marketplace, particularly with regard to the benefits and costs of information collection, retention, and aggregation? (2) Quantification and data: What is the evidence on the prevalence and nature of information loss, what are the costs of information loss and how valuable is this information in the marketplace? (3) Market and Policy Response: What do we know about the efficacy of market vs. other approaches to disciplining market participants either to avoid loss or remediate after information loss? Throughout, of particular interest is the international dimension of information loss. What issues arise when countries differ in their attitudes and policies toward information acquisition, aggregation, retention, and, importantly, in disclosure of information lost?

Excellent research assistance from Alok Mistry, who experienced his own data breach (stolen laptop) during the course of this project.

I. Introduction

The expanding scope of Internet use yields a widening array of firms with access to ever expanding databases of information on individuals' search, transactions, and preferences. This information translates into buyers' greater simplicity and ease of transacting, receiving suggestions on complementary purchases, targeted news and advertising, and other directed products and information, all of which increase customer value—but also raise the potential for compromised privacy and information loss. Similarly, firms have unprecedented windows into customer behavior and preferences, with which they can improve products, segment markets, and therefore enhance profits—but also raise the risk of losing or abusing customer information. A key issue for the Digitization Agenda is how to balance these benefits and risks, particularly in the context of increasingly global flows of information and transactions.

The structure of the information marketplace affects the balance between the benefits and risks. Market structure issues include externalities in the acquisition, retention, and aggregation of information, probability and characteristics of information loss, relationships between market participants in terms of market power, and potential for global arbitrage. Does the information marketplace exemplify the classic type of market imperfection—the deviation between social and private costs and/or benefits—that may elicit policy intervention, and if so, by whom? Do differences in culture, including as expressed in a policy framework, play a role in creating market imperfections?

Disclosure of data breaches is a watershed for this study. Without disclosure, it is impossible to investigate risks and potential costs of information loss. Disclosure also exposes to firms who have not had a data breach the potential cost of incurring one, and thus may create an incentive for them to evaluate their own systems. However, disclosure can be along a spectrum from every incident being announced to everyone vs. only critical incidents being communicated to a few, who might be policymakers external to the firm or might be security specialists internal to the firm or market hierarchy. There is no global standard approach to disclosure, nor even to the notion of disclosure at all; should there, or even can there be such a global approach?

Moreover, whether or not a firm will take action to reduce the probability or type of data loss depends not only on whether the market punishes the firm, but also on how and who bears the burdens of lost information. If the market response to disclosure is sufficient to ensure equilibrium between public and private costs and benefits, then, in principle, no more specific policy intervention is needed. Information on the nature of data breaches, on market response, and on evidence of the efficacy of policy intervention should help prioritize the Digitization Agenda.

This paper proceeds along the following path. The next Section reviews various frameworks with which we can analyze the economic structure of the information marketplace. Section III presents evidence on the extent and nature of information lost. What are the trends: size of loss, sector of loss, source of loss, cost of loss, market value of information, and so on, including in the global context. Section IV addresses market and policy responses to information loss, as well as reviews legislative and legal

strategies that could complement market discipline. Particular attention is given to the challenges of cross-border information flows, differences in attitudes and priorities toward data protection, issues of extra-jurisdictional enforcement, and the associated potential for forum shopping. Section V concludes with priorities for the Digitization Agenda.

II. Frameworks for analyzing the information marketplace and data breaches

That consumers gain from using the Internet is clear from increased competition and reduced prices (Morton, 2006), greater variety (Goolsbee and Klenow, 2006) and faster access to a wider range of public information (Greenstein and McDevitt, 2009; Yan, Jeon, Kim, 2012), even if it is hard to pin down to exactly how large the increase in consumer surplus might be.

Modeling and evaluating the costs and benefits of the information marketplace is somewhat distinct from considering how using the Internet affects consumers and businesses, although the process of use of the Internet generates the data that is the basic building blocks of the marketplace for information. The information marketplace can be decomposed into originators of personal data (consumers); participants that collect and aggregate this original data (intermediaries); and final users of the aggregated data (firms). How should we model the interactions between the data and the participants? Is the data atomistic or is aggregation a key characteristic? Do participants differ in market power and economic relationships? In the marketplace, are social and private costs and benefits aligned or are there market imperfections?

Numerous authors have taken up these questions—usually in the context of the privacy of personal information. The several papers reviewed below offer a concise discussion of the issues, although they do not reach a final assessment on the existence of market imperfections nor what to do about the problem, if indeed such imperfections exist.

The Section concludes with a general framework that focuses on the fact that the market structure includes multiple players, that the type of information collected, aggregated, and lost may vary substantially in its value (including across geographies and cultures) and that the probability function of a data breach may have important attributes that factor into the calculation of costs and benefits in the information marketplace.

Complete markets: The starting point

The purpose of outlining the characteristics of the perfectly competitive marketplace—the Adam Smith marketplace—is to provide a benchmark against which the environment of global information activities can be assessed. If the environment for undertaking information-rich activities is characterized by perfect competition, then Adam Smith's invisible hand—whereby each acting in his own self-interest achieves the highest economic wellbeing for all players—leaves little room for policy intervention to improve the functioning of that marketplace.

In Adam Smith's market, one-off, unrelated transactions take place in a face-to-face setting with unique and uncorrelated prices for each transaction. In this classic marketplace, there is no history of transactions that link buyers, intermediaries, and sellers; no aggregations of purchases (such as databases with the history of a buyer's transactions and personal characteristics) that change the value of future transactions in the marketplace; and no there are no networks that transmit the transactions and allow remote purchases.

An extension of Adam Smith allows for transactions across time, proximity, jurisdiction, and currency. In a 'complete' market (so-called Arrow-Debreu market, Arrow and Debreu, 1954), unique prices, transactions, and economic instruments exist for all possible transactions that the set of market participants can undertake with each other. A 'complete' market accommodates all dimensions of a transaction: including simple examples such as transactions through time and proximity (such as non-cash and consumption smoothing transactions with credit cards) and over geography and regulatory jurisdiction (such as international purchase of goods and services); and accommodating heterogeneity of tastes and preferences (such as credit card vs. Paypal for on-line transactions). A 'complete' market is one where Adam Smith's one-off and face-to-face transactions can be replicated regardless of the proximity, temporality, size, or preferences of transactors (or their representatives, such as governments).

In a complete markets framework, the socially optimal outcome, with private and social interests aligned, can be achieved because there is a perfect (complete) match between specific products and atomistic market participants over all possible states-of-nature and time. The price of each match is unique and efficient; transactions are frictionless.

Considering the complete markets framework in the context of the information marketplace would require atomistic actors and data, full disclosure of all information, and a matched instrument set between information use and instrument that protects information from misuse. In a number of ways the information marketplace violates key assumptions of the complete markets framework, which opens up for consideration the issues of market imperfections and of the second-best.

The first violation is the assumption that information is atomistic. The value of information in aggregate (database) is greater than the sum of the parts (individual behaviors and transactions). Even if each unique piece of information had a uniquely matched price, which would efficiently price use (and therefore matched instrument to protect against misuse) of that information, there would be a tension between the value of that morsel of information by itself and its value in a database. The perfect mapping between state of nature (probability of unauthorized use) and product fails because there is an externality inherent in the aggregation of information in this marketplace.

The second violation of the complete-markets assumptions is that actors are not atomistic. In many cases, individuals do not transact with other individuals. Rather information intermediaries and aggregators stand between the originators of unique data and users of the aggregated data. These information intermediaries and aggregators are

not atomistic, but rather have some market power, which will affect the price and value of information and its associated protection.

These violations of the complete markets framework offer jumping-off points for more specific analyses of the information marketplace that trace through the nature of the gap between private and social costs and benefits. [1] It is not surprising that authors do not agree on whether or what kind of intervention with regard to acquiring, aggregating, disclosing, and protecting information might be required to bring private and social costs and benefits into closer alignment. In part this is because with incomplete markets marginal improvements among the second-best cannot be ranked. However, any consideration of approaches to closing the gap between social and private costs and benefits is also hampered by the probabilistic nature of information loss and the specification of that probability.

Applying the 'Pollution' model to information externalities

Hirsch (2006) starts with the presumption that information is not atomistic and that collecting and aggregating personal information generates negative externalities. 'There is a growing sense that the digital age is causing unprecedented damage to privacy…. digital economy businesses often do not bear the cost of the harms that they inflict'. Just as pollution, as an externality, is an outcome of production, so too is harm to privacy an externality of the information 'production' activity itself. In the pollution-model of the information marketplace, information loss is not necessary to generate harm; aggregation alone generates a deviation between private and social benefits/costs.

Hirsh continues with the environmental analogy and reviews the evolution of policy strategy from 'command and control' compliance to 'second-generation' or 'outcome oriented' policy whereby the regulated entities find their own cost-effective strategy to achieve the legislated goal. Tang, Hu, and Smith (2007) nicely model this kind of regulation in the information marketplace, and find that mandatory regulation raises consumer prices and reduces firm profits, just as would be expected. Whether such internalization of the losses to privacy improves social or private well-being is not resolved.

Moreover, while environmental economics offers a model for the information marketplace, the analogy is stretched because consumers and produces do gain from information aggregation, whereas it is hard to imagine anyone actually gaining from downstream pollution.

Information aggregation and trade-offs with limits to rationality

Acquisti (2010) argues that the information marketplace is all about trade-offs. "In

[1] For other approaches to modeling privacy, see U.S. Dept of Commerce, NTIA chapter compendium of articles. Roberds and Schreft (2009), Anderson (2006), and references therein.

choosing the balance between sharing or hiding one's personal information (and in choosing the balance between exploiting or protecting individuals' data), both individuals and organizations face complex, sometimes intangibles, and often ambiguous trade-offs. … But trade-offs are the natural realm of economics."

In this view, what opens the door to incomplete markets, and perhaps regulation of some sort, is the limit to consumer rationality and transactions costs, which might affect the price of information, and thus the distribution of benefits and costs. If consumers don't know the value of their information, they cannot calculate the trade-offs, nor price their actions appropriately so as to move the overall marketplace toward the joint private and social optimum.

The problem goes beyond limited rationality of individuals. Even if each piece of information was appropriately priced (perfect rationality and perfect markets), the aggregation of individual pieces of data is greater than their sum … but by how much? Moreover, how should the economic gains to aggregation be apportioned? Individualistic pricing of data may not be a valid starting point.

Researchers have attempted to calculate the value of the aggregation of one's own information. Conjoint analysis by Hann, et al (2002) finds that consumers value their privacy at about $40-$50—that is, they trade their information for about $40-$50 of product value. Convenience is often cited as a rationale for allowing the aggregation of one's own personal information, as in on-line banking. (Lichtenstein and Williamson, 2006). But research also finds that that the aggregation of personal information incurs a negative externality on account of distrust of the aggregator. (IAB McKinsey). And, the cost to firms of the inability to use individual and aggregate personal information to target advertising has been documented. (Goldfarb and Tucker, 2010)

Lack of transparency and ownership and control over what is being collected, retained, and aggregated puts the limited rationality problem, for both the individual and individual's data, at the center of the gap between social and private outcomes. Policy interpretations (for example, the EU Privacy Directive) about how to bridge this gap--including, in the extreme, disallowing the collection and retention of personal information--is at the heart of differences in regulatory response across countries, discussed in more detail in Section IV. But, it is also the case that market response to demand for ownership and control is becoming a distinguishing niche for firms. (NYTimes: If My Data Is an Open Book, Why Can't I Read It? http://nyti.ms/12UHryv).

Multi-actor markets and the role for disclosure

Most of the literature discussing externalities in the information marketplace uses a two-actor framework– so-called data subjects (such as customers that 'provide' the information) and so-called data holders (such as a firm that aggregates customer data to create customized products). Often there is a third player in the information marketplace through which the data 'transits' or 'rests', such as payment processors. How these three

players interact creates opportunities for market imperfections and pricing that deviate from the perfect markets solution. In addition, there can be asymmetric externalities (costs vs benefits) depending on the type of information gathered, aggregated, and potentially lost. Individuals may get disproportionate benefits in some examples of information aggregation (free mobile phone apps), but bear disproportionate costs in cases of certain kinds of data loss (such as financial or medical information).

Romanosky and Acquisti (2009) use a systems control strategy to map three legislative approaches to reducing harm to privacy, which in their case happens when information is lost, not just (as in Hirsch's case) when information is collected. Two of the three approaches draw from accident legislation: First, ex ante 'safety regulation' (think seat belts) in the context of the information marketplace would include promulgation and adherence by information collectors to, say, Payment Card Industry Standards. But, they argue that ex ante standards focus on inputs (encryption) rather than outcomes (harm); so are not efficient. Second, ex poste liability law (think law suits) would include negligence in the treatment of personal information. But, effectiveness of ex-post liability is reduced because courts have been unwilling to award damages based on the probability of some future harm coming as a consequence of a data breach (but see the evolving legal landscape in Section IV).

A third mechanism is information disclosure, such as the California data breach disclosure law. Information disclosure offers great promise to close the gap between private and social outcomes, but consumer cognitive bias (misperception of risk) and transactions costs suggest that the gap cannot be completely closed. Moreover, it is not clear what disclosure actually means – disclosing what to whom?

Romanosky and Acqusiti use their framework to outline an example of where cognitive bias and transactions costs problems have become less apparent and where disclosure and market structure (specifically concentration) has been key to that happening: The relationship between credit-card issuing institutions and firms that hold (and lose) credit-card data. They argue that information disclosure has promoted the internalization of the costs of remediation by the data holders (and losers), which increases the incentives for the adequate protection of personal information even when the individual provider of that information cannot demand such protection.

Why does disclosure help align private interests to move closer to closing the gap between private and socially optimal outcomes? First, a sufficient number of data breaches have occurred such that these costs have begun to be quantified (to be discussed in Section III below). Second, the number of affected intermediaries (card issuers in this case) is sufficiently small that they have market power to demand remediation (or impose punishment) from the other concentrated actor, the data aggregators/holders. Third, the chain of causation between information loss and required remediation is known because of data-breach disclosure laws. The disclosure laws along with quantification of costs as well as the small number of economic actors promotes the transfer of remediation costs from the card issuers to the database aggregators, those who actually lost the information. Thus, at least some of the externality is internalized.

However, the costs of information loss borne by individual card holders is not transferred to those firms where the data breach occurred, and the market power of individuals is small and, in a transactions-sense, distant from the data aggregators/holders since the individuals' data transits before it rests. Individuals can change card issuers, but they have no power to affect a change in the relationship between their card issuer and what firm aggregates the transactions of that card. Thus, the individuals' costs are not internalized, so there remains a gap between private and social outcomes.

Modeling frameworks in the international context

Information-rich activities are becoming increasingly fragmented. Just as the production of physical products has a global value chain, information-rich products have a globalized value chain, thus exposing information to multiple jurisdictions of protection. When different jurisdictions have different rules it may not be clear whose rules apply.

For example, the FDIC (2004) considered the implications of 'offshoring' of financial activities to third parties in foreign countries. It noted that, while the Gramm-Leach-Bliley Act affirmed that U.S. data-protection rules covered personal information regardless of its geographic location, it also noted that it can be difficult in practice to ensure the extra-territorial application of U.S. rules. In particular, fragmentation and global information flow mean that U.S. firms may not have (or may choose not to have) full transparency over their global value chain, and therefore U.S. rules may not be applied at all points. Information loss from a third party can yield large and broad based thefts: (NYTimes: Cyberthieves Looted A.T.M.'s of $45 Million in Just Hours http://nyti.ms/ZKTw5H.)

Another example, the European Union has a set of Directives related to personal information. (Discussed more in Section IV.) The interplay of the mandate-approach of the EU Directives with the U.S. more market-oriented disciplines immediately led to the negotiation of the Safe Harbor Agreement (2000), where by U.S. firms' transmission of information to their subsidiaries in the EU, and/or engaged in transactions with EU consumers would not have to treat the information differently depending on where it was in terms of geography. The Safe Harbor solved the firm's problem, but it did not address the treatment of data loss, nor did it examine whether this agreement undermined the EU approach in favor of the US approach to data protection.[2]

Finally, the issue of cross-border regulation of information flows comes up in global trade negotiations. The World Trade Organization General Agreement on Trade in Services (WTO GATS) is a 'positive' list approach to trade negotiations. This is as

[2] See Mann, C.L. (2001) "International Internet Governance: Oh, What A Tangled Web We Could Weave!," *Georgetown Journal of International Affairs*, Summer/Fall and Mann and Orejas (2003), "Can the NAFTA Partners Forge a Global Approach to Internet Governance?" in North-American Linkages, Richard G. Harris, ed. Ottawa: Industry Canada, 2003.

opposed to the 'negative list' approach whereby trade flows between countries are assumed to be unburdened by regulations, tariffs, and quotas, except for specific derogations (the negative list). (This is the so-called Most Favored Nation principle, embraced in the WTO-precursor of the General Agreement on Tariffs and Trade, GATT). The positive list approach implies that regulatory and tax treatment of bilateral data flows must be individually negotiated, thus creating the potential for a complex web of jursidictions.

Modeling the probability of data breaches

Increasingly, investment and risk modeling and analysis is being applied to the decision to invest in information technology security, both in the theoretical domain and by practitioners. (Aurora, Hall, Pinto, Ramsey, Telang, 2004; Gordon and Loeb, 2006; Bojanc and Jerman-Blazic, 2007;Carty, Pimont, Schmid, 2012.) A topic that should receive more attention in modeling and analysis frameworks is how the probability of information loss affects the equilibrium of costs and benefits. Suppose the problems of limited rationality and complex market structure were solved. And, that the market was complete in the sense that there was a match between personal information and protection. Still unknown, and important for a cost-benefit calculation, is the probability of a data breach. The actual vs. the perceived probability distribution is an important ingredient to the numerical calculation of the cost-benefit tradeoff.

An analogy comes from the market for foreign exchange. Suppose a firm wants to put a floor on the value in the home currency of the revenue stream earned abroad in that currency. In a perfect markets world, the firm could by an option that will pay-off when the home currency reaches a particular value, and that option would be priced exactly so as to make the firm indifferent to buying the option or not.

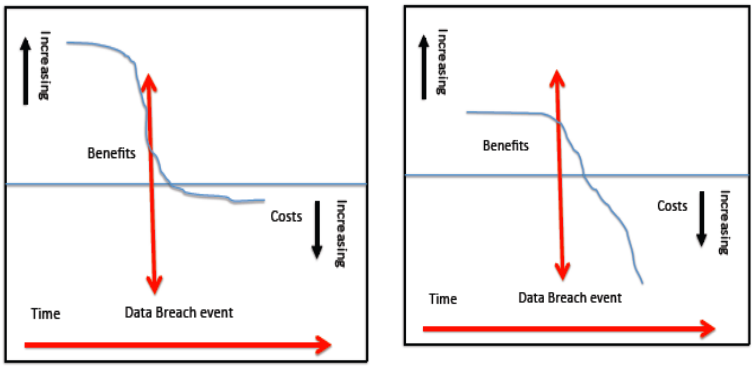
However, the firm must make the decision of whether to buy this for-ex option based on its estimate of the probability of the depreciation of the home currency. The probability of a home-currency depreciation can be calculated based on historical experience, but the choice of functional form for these historical data matters. If the historical data is assumed to follow a normal distribution, but the true distribution has 'fatter tails', then the probability of the depreciation will be underestimated, the price of the option will be over-priced, and the matching between the instrument and its option will not be perfect, in the complete markets sense. The firm will choose not to buy the option, and it will experience an uncompensated loss.

In the information market place there is a similar problem of estimating the probability of a data breach. Both consumers and firms face the challenge. Moreover, unlike the for-ex market where there are millions of data points, disclosure of data breaches is relatively recent and incomplete, so data on the probability of a data breach is limited. How to incorporate the probability of a data breach into the cost-benefit calculations of data breach and security investment warrants continued assessment and investigation.

A general framework: Externalities, asymmetries, and probabilities

Suppose we could calculate costs and benefits. What would the function look like around the time of a data breach? Would it look like left panel below: A lot of benefit all around to consumers and firms and not much loss to anyone in the case of a data breach, or like right panel below: Benefits yes, but an increasing loss after the data breach event.

It is reasonable to believe that the cost-benefit functions look different depending on the nature of the information (restaurant and book preferences vs. financial and medical information). It is possible that these cost-benefit functions look different depending on culture and history of the population (German vs U.S. nationals). These considerations might point to how to structure an empirical strategy (discussed in Section IV) to evaluate market and policy responses to potential imperfections in the information marketplace.

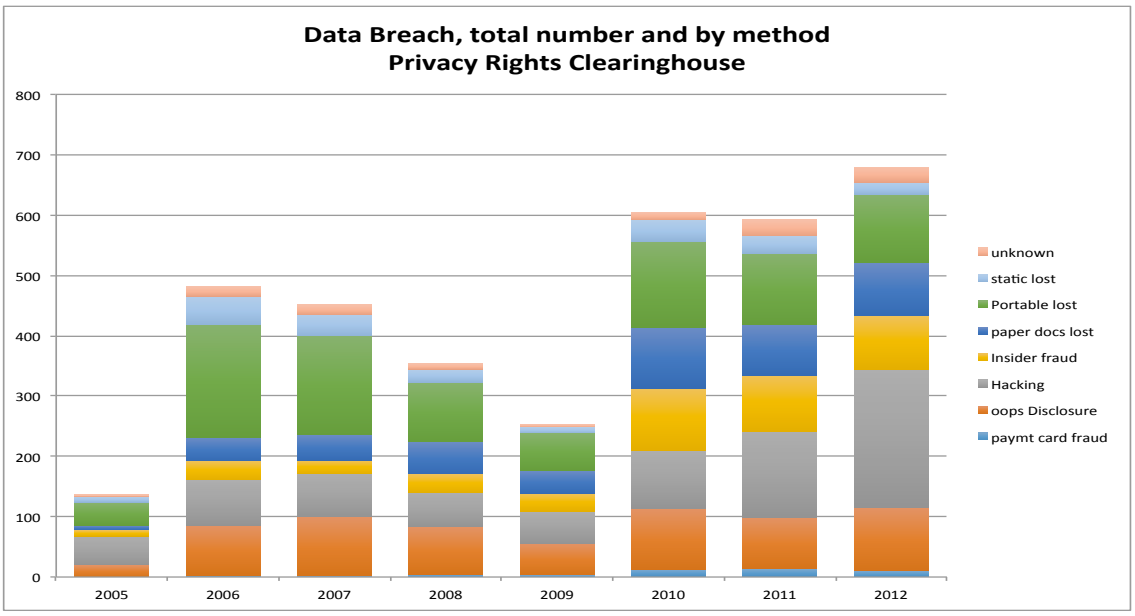


III. Trends in Information Lost

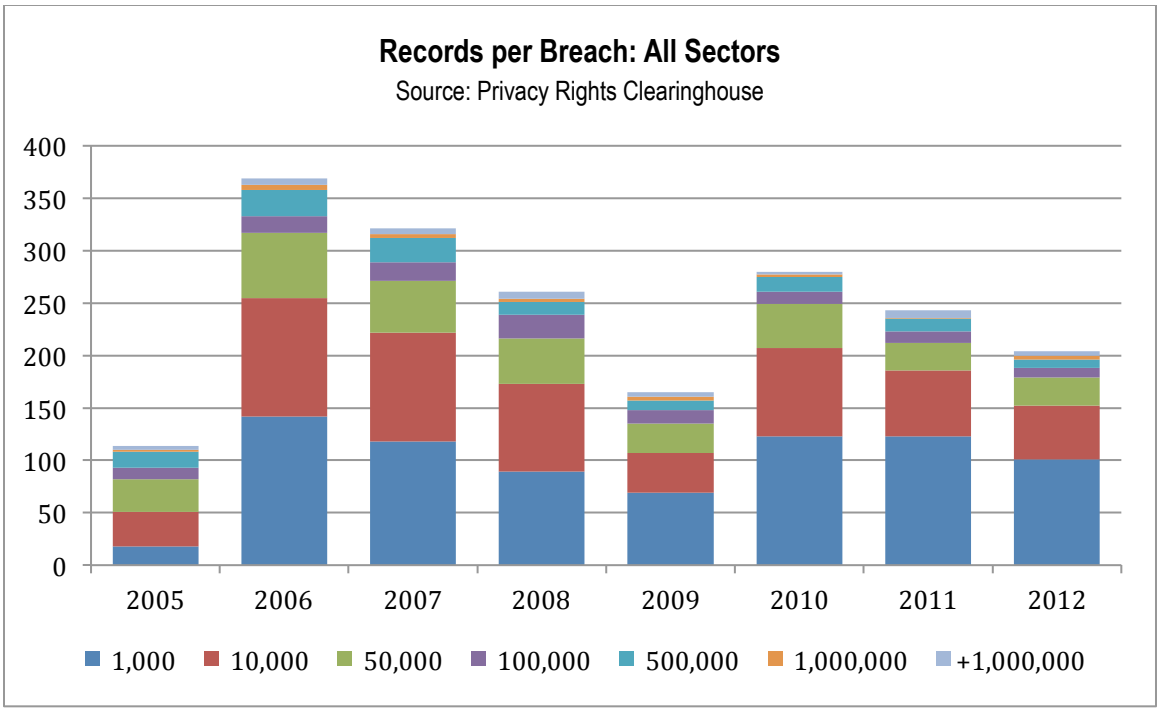
Presented below are a variety of trends in information lost. The raw data come from several sources including: Privacy Rights Clearinghouse, which draws from public news sources; Poneman Institute, survey-based; Symantec, based on survey; Verizon, which uses industry survey; KPMG Europe, also uses survey; Federal Trade Commission, draws on consumer fraud on-line report database.

How much information is lost? And by what means?

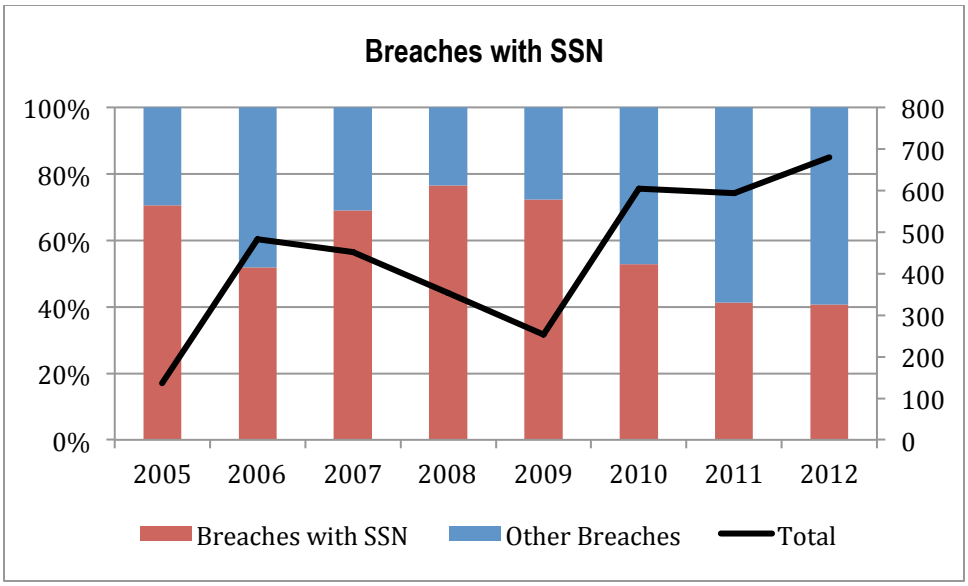
The Privacy Rights Clearinghouse data for 2005 to 2012 show that after a notable drop in data breaches in 2009 (do people use the Internet less in the depths of a recession?), data breaches are on the increase again. (The number of records lost in each breach, which is a different measure of information lost, will be discussed below.) Hacking and insider fraud are the increasingly important source of data breaches for most sectors, but a surprising number of data breaches still take place the 'old-fashioned way' by losing paper documents or laptops and through unintended disclosure (for example, cc vs. bcc in the e-mail list).

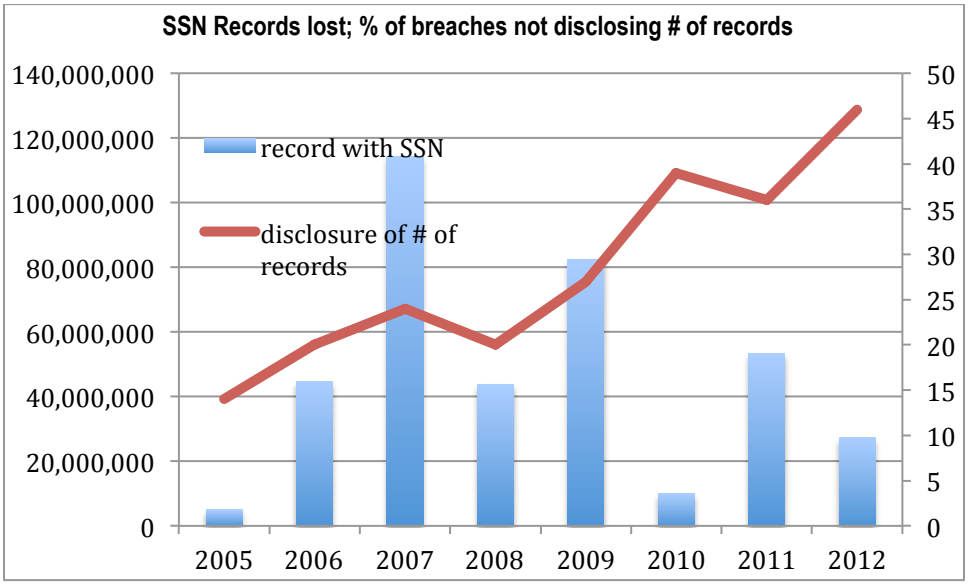


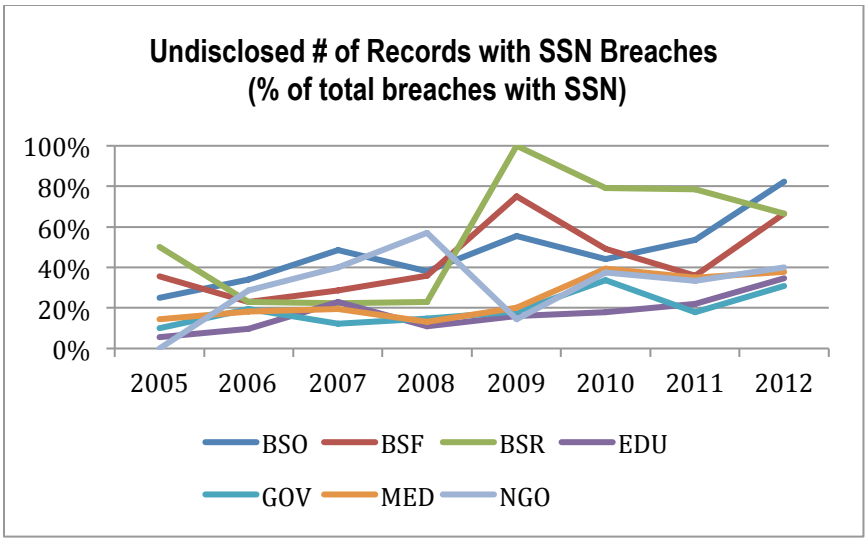
Whereas the announcement of a breach indicates that information has been compromised, the actual number of records involved in each breach could be a better measure of potential cost in that a record represents granular information. Not all breach disclosures reveal how many records were lost in the breach. In fact only about half of the announcements include that information. (See more discussion of SSN-related breaches and records lost below.) For the breach disclosures that reveal the number of records lost, over the 2005-2012 period, the histogram of records lost per breach shows some reduction in breaches with medium sized losses (100,000 – 500,000 records lost), but little progress in stemming breaches of either small or huge size. Small and moderately sized breaches (1,000-10,000 records lost) still dominate the breach dynamic, and have not declined over time. Huge breaches (1,000,000 and up) have not been controlled either.



Revealing a social security number (SSN) during a data breach generates far greater concern and potential for costly information loss compared to a data breach that compromises other types of personal information (see the next section on the SSN and equity market and legal discipline). Over the time period, the number of breaches that reveal SNN has increased; but as a share of all data breaches, those that reveal SSN has declined. Importantly, the number of reported records where the SSN was compromised has declined from a peak in 2007, although not in trend fashion. Finally, recalling that not all breaches announce the number of records lost, it is the case that within breaches that compromise SSN, the share of those breaches that have not disclosed the number of SSN-related records lost has increased over time… with the Business-Other being the largest sector not disclosing. Sectors that perhaps are under greater scrutiny, such as Medical, Financial, and Retail appear to disclose more information. Thus, to the extent that SSN does promote market discipline, the role for disclosure may have, on the one hand, led to fewer SSN-related breaches, but on the other, has prompted less transparency in public report. But, if disclosure is an adequate and appropriate disciplining device, there is much more to reveal.


Breaches with SSN
100%
80%
60%
40%
20%
0%
800
700
600
500
400
300
200
100
0
2005
2006
2007
2008
2009
2010
2011
2012
Breaches with SSN
Other Breaches
Total

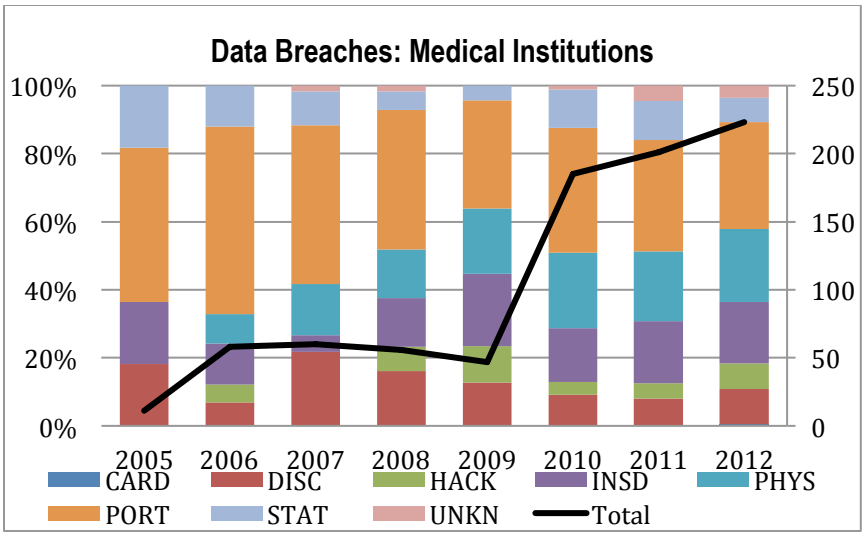

SSN Records lost; % of breaches not disclosing # of records
140,000,000
120,000,000
100,000,000
80,000,000
60,000,000
40,000,000
20,000,000
0
50
45
40
35
30
25
20
15
10
5
0
record with SSN
disclosure of # of records
2005
2006
2007
2008
2009
2010
2011
2012

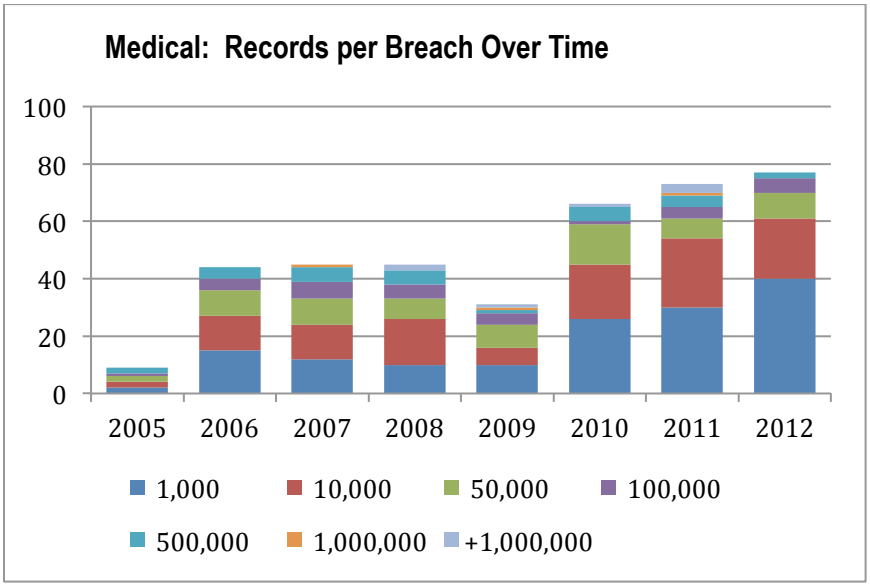


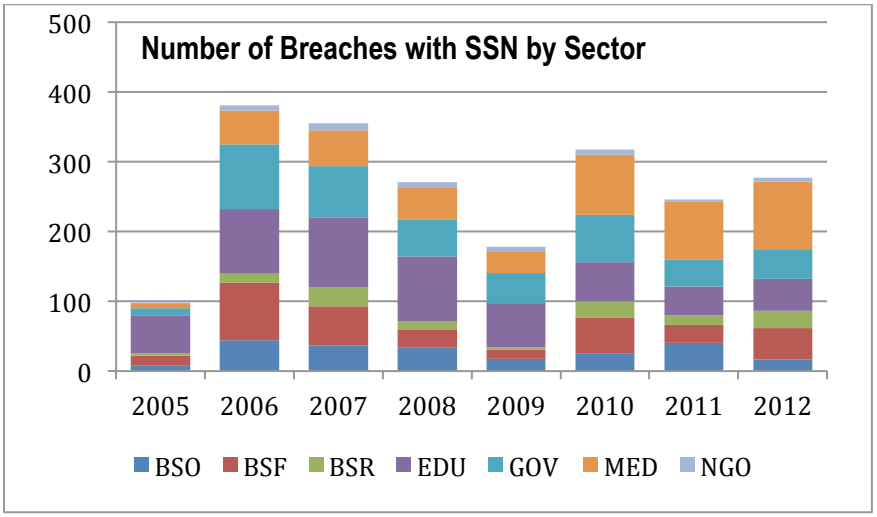
Is there differentiation by sector?

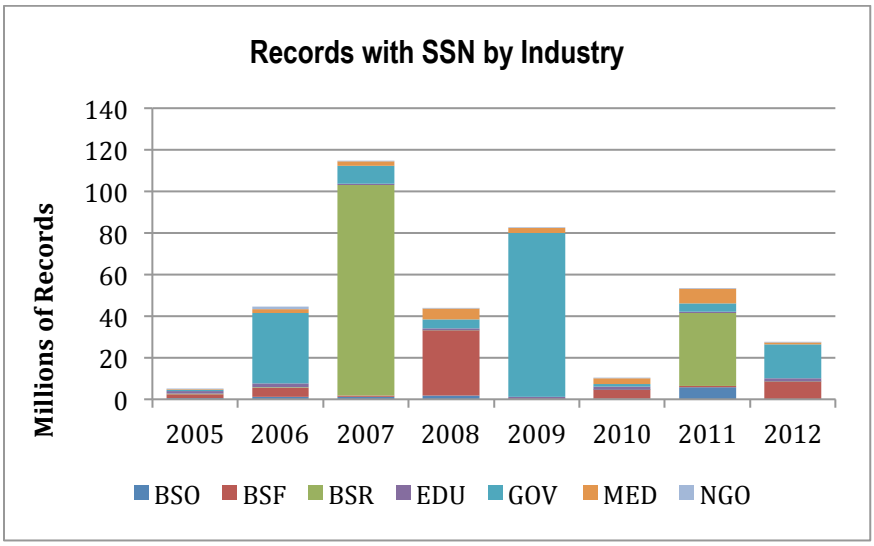
Looking below the average over all breach events, are there differences by sector? Which sectors are the most prone to data breaches, by what means, and does the size of breach and information revealed differ by sector?

Data breaches in the medical sector are about double any other sector, with a huge increase in the last couple of years. (This could be a fact, a function of disclosure, or a function of disclosure and reporting.) In contrast to the overall picture, the main source of data breach is lost paper documents and lost laptops, with a rising role for insider fraud (as opposed to the overall picture where hacking appears the greatest threat). The vast majority of data breaches for medical institutions are small breaches – 1,000 to 10,000 records lost—but a lot of these data breaches reveal SSN. However, when the number of records lost with SSN is considered relative to other sectors, medical is not the largest problem sector.

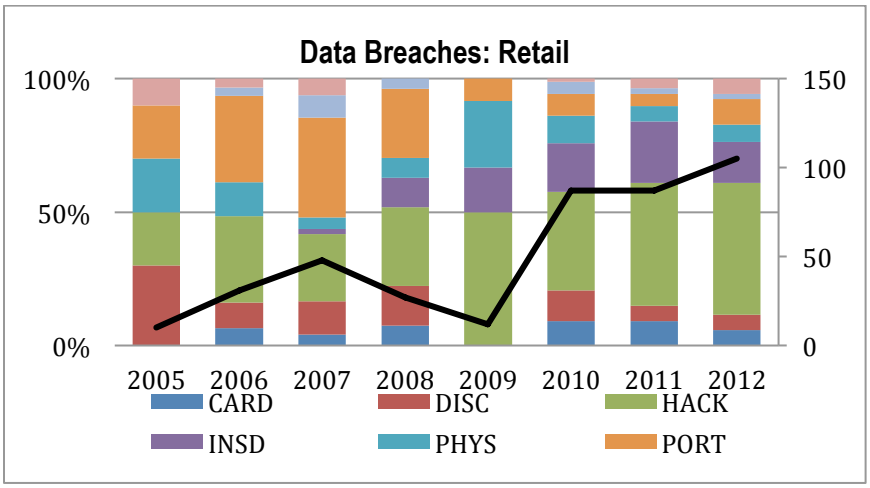

Data Breaches: Medical Institutions
100%
80%
60%
40%
20%
0%
250
200
150
100
50
0
2005 2006 2007 2008 2009 2010 2011 2012
CARD DISC HACK INSD PHYS
PORT STAT UNKN Total

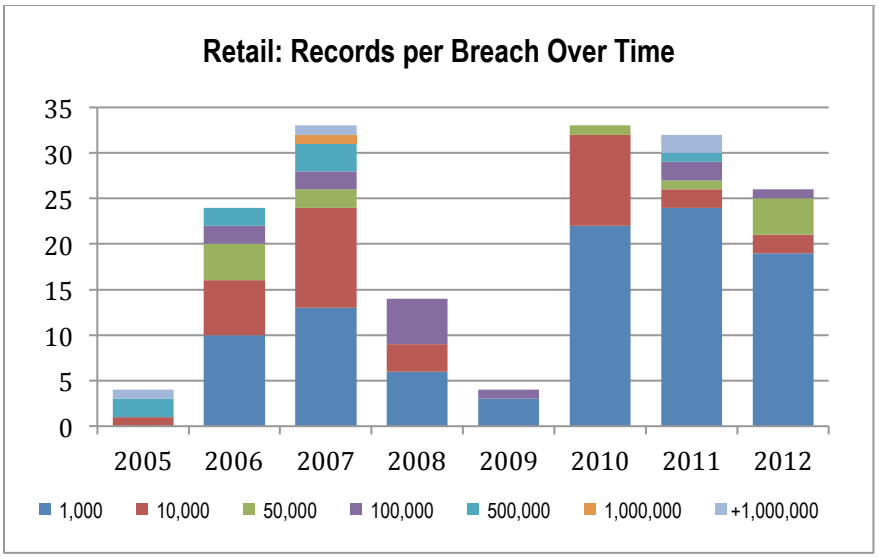

Medical: Records per Breach Over Time
100
80
60
40
20
0
2005 2006 2007 2008 2009 2010 2011 2012
1,000 10,000 50,000 100,000
500,000 1,000,000 +1,000,000


Number of Breaches with SSN by Sector
500
400
300
200
100
0
2005 2006 2007 2008 2009 2010 2011 2012
BSO BSF BSR EDU GOV MED NGO

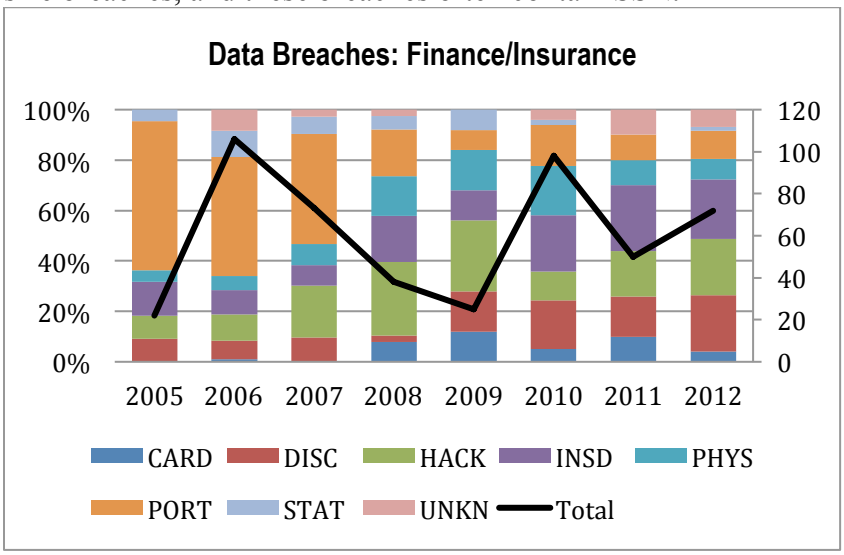


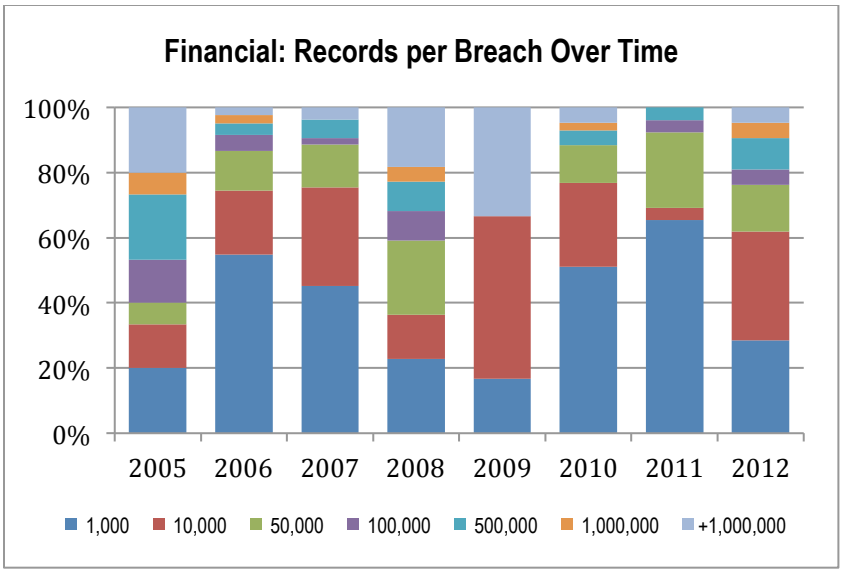
The chart above on records lost that compromise SSN reveals that retail is another sector that has a lot of data breaches. As shown below, the vast majority of data breaches in retail are by hackers. The number of records lost per breach is very small, generally, and the number of breaches that reveal SSN is quite small generally. But, when retail sector experiences a big exposure (2007 and 2011), the loss of records with SSN is enormous. The chart also reveals that 2009, which was the low point for overall breaches, was low because of the low number of small retail breaches. The Great Recession hit consumer spending and small business retailing relatively hard. So, the relationship between overall economic activity and data breaches, once again, appears a relevant hypothesis to test.



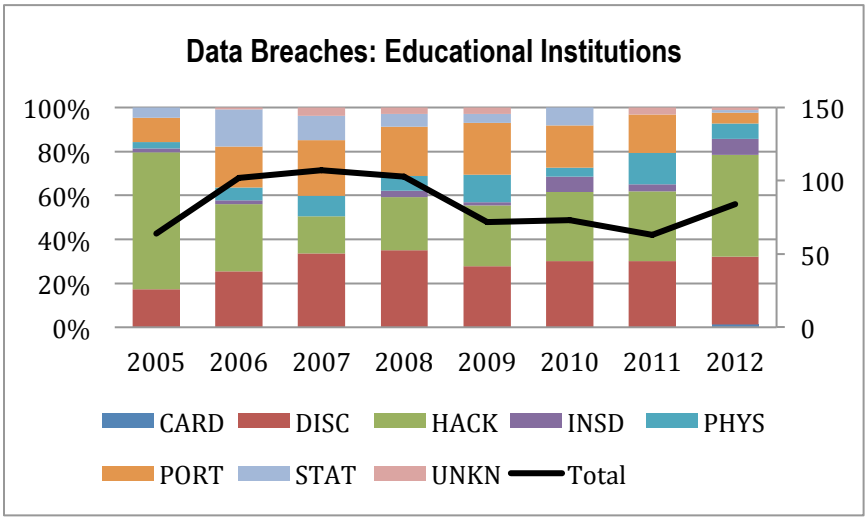


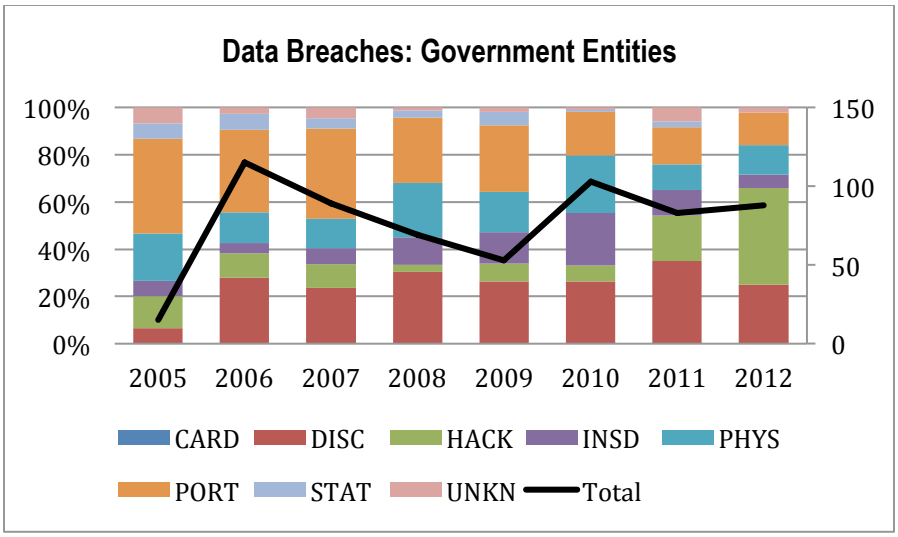
A third sector of particular interest is financial and insurance institutions. The number of data breaches appears to be under control. However the origin of the breach through insiders is significantly greater share than other sectors, and both hackers and unintended disclosures also are large. Very large breaches occur nearly every year, along with mid-size breaches, and these breaches often contain SSN.

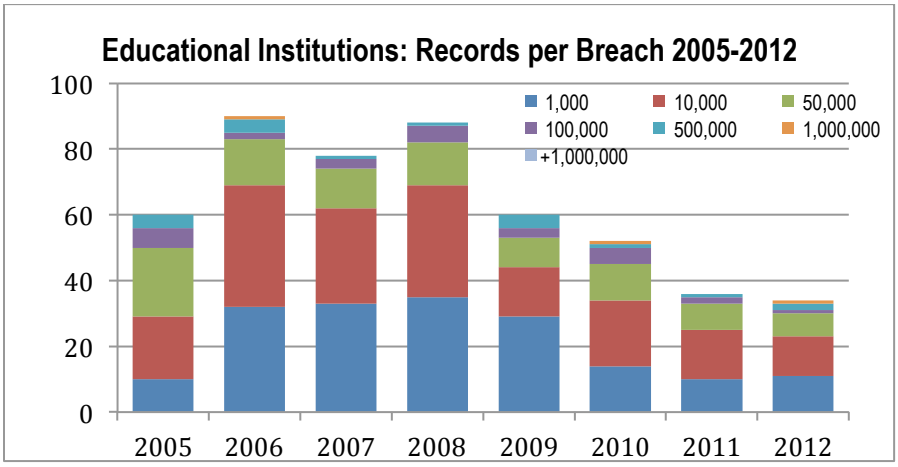


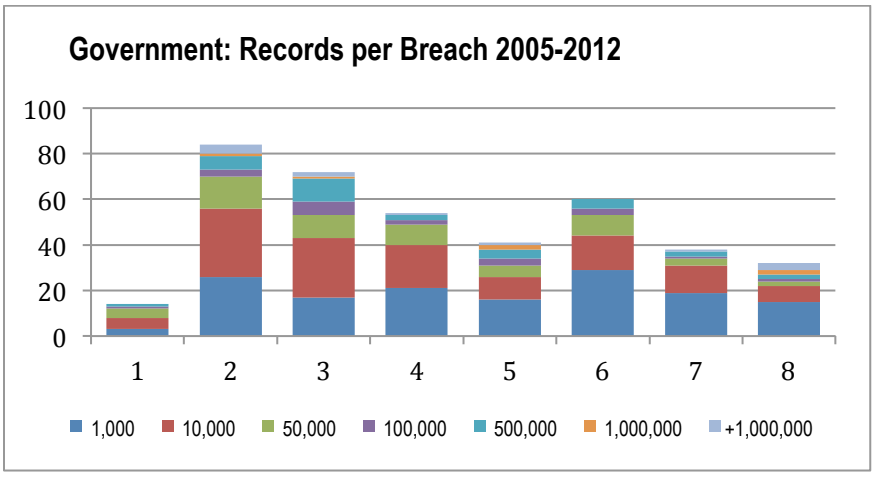


Government and educational institutions lose data both from hacking and from unintended disclosure. The bulk of the losses in the education sector are small, but the government has experienced some very large losses, and with a large number of records containing the SSN.









The two dimensions of international breaches

Cross-border data breaches have two dimensions: A U.S. institution or consumer may lose information to foreign perpetrators. Or, a U.S. institution, when it incurs a data breach, may expose the personal information of a foreign person or firm. What are the characteristics of these cross-border breaches? The picture is quite murky since public and reported disclosure is, at present, only required of U.S. firms. When US firms incur a data breach event, further analysis of the cross-border aspects of those incidents is by surveying of the firm, albeit including some relatively hard data such as IP addresses. Data on consumer's information exposed is self-reported to the US. Federal Trade Commission or by other survey and therefore must be considered spotty and incomplete.

With regard to information at U.S. firms exposed to external threats (e.g. what are called hackers by Privacy Rights Clearing House), Verizon reports that about 20% of incidents are U.S. hackers compromising the data of U.S. firms. However Verizon reports a significant rise in the Central-Eastern European countries as origin of compromise, which it reports as 'organized crime' targeting smaller U.S. firms using primarily point-of-sale or other skimming-type devices (this is consistent with the prevalence in the PRCH data on small breaches in the retail sector). Note however, the very large share of incidents where the origin cannot be determined.

Geographical origin of external information lost,
% of incidents
Source: Verizon, DBIR

	2007	2008	2009	2010	2011		2012
Americas-North	23	15	19	19	20	US	16
Americas-South	3	6	na	<1		Colombia	1
						Brazil	1
Asia-East	12	18	18	3	2	China	2*
Asia-North/central	9	nr/	nr	0	nr		
Asia-South/Southeast	14	3	2	6	1		
Europe-East, Russia, Turkey	24	22	21	65	67	Romania	28
						Russia	5
						Armenia	1
						Bulgaria	7
Europe-West/South/North	9	3	10	2	4	Germany	1
						Netherland	1
Middle-East	5	na	5	na	na		
Africa	1	1	2	4	1		
unknown	nr	nr	31	nr	10		

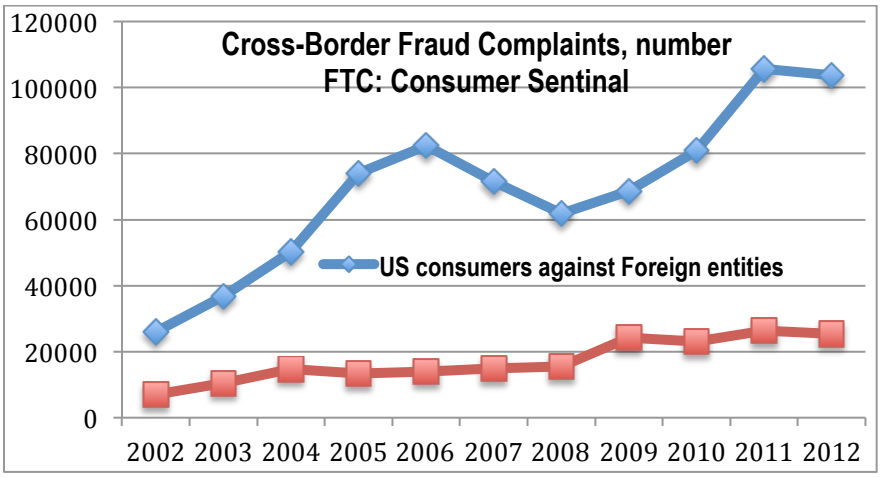
* external threats from China: 30% of threats; only 2% focused on financial (as opposed to industrial) information… the focus of this section of the paper is not on industrial espionage.

nr/ not reported

With regard to specific countries in CEE, Bulgaria and Romania especially (singled out in the 2012 report) joined the European Union in 2007, and at that point should have been brought under the umbrella of the EU Directives on Privacy (1998), Privacy and Electronic Communications (2003) and Data Retention (2006). So it is perhaps a surprise that so much threat activity emanates from these countries. However, another point to consider is that, because the EU has always essentially disallowed data retention, its approach to data protection has focused on data-in-transit, not data-at-rest. In contrast, the U.S. tends to focus on data-at-rest for security emphasis. It is possible that the different security focus exposes weakness in the U.S. data-in-transit protocols that can be exploited by forum shoppers who have different knowledge sets. (See more discussion of Directives below.)

KPMG reporting on data breaches in the U.K. presents a quite different picture. The U.S. remains the largest source of global incidents (about 50% of the incidents in January-June 2012), but that is down from 75% of the incidents taking the KPMG data for 2008-June 2012. KPMG does not even report separately the CEE region or any of its countries. The U.K. originates about 10% of incidents, which is not a country that Verizon separates out. Whether the origin of breaches indeed is so different depending on who is surveyed, or whether the reporting is so uneven across countries is an obvious question.

Considering just consumers, rather than a survey of firms from the previous sources, the FTC reports that cross border consumer fraud continues to be dominated by U.S. consumers reporting to the U.S. FTC. Cross-border consumer complaints (e.g. foreign consumers complaining about U.S. firms) account for about 13% of all fraud complaints and that percentage has not changed over time.



In sum, in the United States, much information is lost the 'old fashioned way' (e.g. lost laptops and paper work, and unintended disclosure). But, information lost via cross-border hacking is increasingly important in all sectors, with insider-originated losses as well particularly notable in finance. Many more data breaches occur with small numbers of records lost, but in any given year, the largest breaches with a huge number of records, and high proportion of SSN can occur in any sector. So, should the focus be on the

numerous small breaches or the very few disastrous breaches? Finally, sectoral variation and variation the size of the breach in how information is lost may be relevant when considering the role for a domestic focus vs. a global approach to information security.

IV. Market discipline vs non-market regulatory and legal discipline

This Section reviews evidence on strategies to discipline market actors to internalize the costs of data breaches and close the gap between the social and private benefits of information. If market discipline is sufficiently robust, alternative approaches of government regulatory intervention, or private legal action may not be necessary.

Given the analysis on the nature of data breaches (above) it may be that private sector response (Payment Card Industry Standard, for example) will focus on the numerous small breaches, whereas policy intervention is needed to when trying to prevent or ameliorate the infrequent disastrous events. On the other hand, if information security is expensive, it may be inefficiently costly relative to societal benefits for firms with small, albeit numerous, breaches, to protect their data. Would standardizing security products or data help reduce the implementation cost? Moreover, the increasingly fragmented and globalized information flows challenge the application of market discipline or regulatory discipline alike. Data on the costs of breaches is an integral part of the analysis—but how to measure costs, and costs born by whom? Further what is the stolen information worth in the marketplace?

First, how might market actors respond when information is lost? If the company is customer-facing, such as a retail firm, sales might drop as customers buy from competitors. If the company is a financial intermediary, such as a payment processor, it may be shunned or fined by other parts of the payment chain. If the company is a technology firm, corporate governance of its own activities may be questioned. If the company is in the health-care sector, its reputation may suffer. How costly are these market responses to the announcement of a data breach? And, how costly to remediate or prevent a data breach?

A key problem, noted in the frameworks sections, is that there are multiple actors. Whether or not a firm will take action to reduce the probability or type of data breach depends not only on whether the market punishes the 'right' firm, but also on how and who bears the burdens of lost information. For example, the costs of notification and of ameliorating a data breach (for example, issuing new credit cards), as required by the California law, could be the main channel for market discipline. Similarly, fines imposed within the self-regulatory hierarchy (for example between merchants, card issuers and payment processors via the PCIS) offers a disciplining device, as do fines levied by a regulatory agency such as the Federal Trade Commission. Finally, legal suits brought by those suffering the information loss could be sufficiently threatening, or actually costly enough, to encourage firms to enhance their data security or design their information systems differently, although the international nature of theft adds another dimension to the legal challenge.

Is disclosure enough to discipline, and disclosure to whom?

The disciplinary mechanisms noted above all require that a data breach be acknowledged. But, to whom the data breach should be disclosed—those whose data are compromised or an intermediary whose responsibility it is to safeguard the data—is less clear. As noted in the framework, individuals are, by definition, atomistic and therefore may lack market power to respond to disclosure. Intermediaries are fewer in number, but may have little incentive to admit or remediate a data breach. With multiple level of intermediaries (encryption data-in-transit for example before it become aggregated into data-at-rest) there is the added potential of moral hazard of one type of intermediary free-riding off the security approach of another part of the information value chain.

A U.S. state law, first introduced in 2003 in California as Senate Bill 1386, mandates that organizations that maintain personal information about individuals must disclose if the security of the information has been compromised. Moreover, the legislation stipulates that if there has been a breach of a database containing personal information, then the responsible organization must notify each individual for whom it maintained information. The law forced every firm doing business in California to comply. By 2007, 46 of the U.S. states had adopted similar versions of a breach-disclosure law, although to date there is no federal legislation governing most personal data.[3]

To what extent are disclosure mandates a key foundation required for consumers, businesses, and/or policymakers to adequately assess the costs and benefits of engaging in the information-rich, but potentially information-risky, marketplace? Disclosing a data breach to market participants with power to punish may not mean public disclosure to the consumer. In fact, if the consumer has little market power, has limited rationality, or is 'distant' from the offending firm, then disclosing a data breach to the consumer may not be the most efficient approach to aligning costs and benefits.[4]

Indeed, the U.S. approach of such broad-based disclosure to individuals is unique around the world. Among other countries, only the United Kingdom has legislated disclosure, but the disclosure is to a governmental agency. Similarly, Japan requires disclosure to a governmental agency, but the scope of which type of incidents must be disclosed has been narrowed on account of 'excessive' information flooding the agency. (reference) Australia is considering whether a specific disclosure law is necessary or whether existing law addresses disclosure. (source) The European Union's approach here-to-fore under the 2003 Directive on Privacy and Electronic Communications has not required

[3] There is federal legislation protecting children (COPPA), health (HIPPA), and financial data (GLBA) but generalized 'personal information' is not protected. At the state level there is a patchwork of legislation that protects some information in some states – for example, databases with drivers license information are in the public domain in some states, but not others. For a complete review of US data security legislation, see Stevens (2012).

[4] Reference the Goldberg and Tucker, and AEI survey articles on business costs of disclosure without necessarily gaining consumer benefit.

disclosure of information loss. But the promulgated and evolving General Data Protection Directive will require disclosure of material breaches to a supra-governmental unit. (ref).

Does disclosure even work to reduce the incidence of data breaches, and at what cost? There is relatively little research on whether disclosure itself works to reduce data breaches, much more research (discussed below) on whether disclosure punishes firms, which presumably is the first step needed for firms to receive the signal to safeguard data. Romanosky, Teland, Acquisti (2011) find that data disclosure rules reduce ID theft by about 6 percent. On the other hand, Romanosky, Acquisti, and Sharp (2010) consider the optimality of U.S.-style disclosure. Considering parameters of cost to disclose, response of consumers to disclosure, consumer harm, and reduced rates of data breach, they find that disclosure probably is too costly relative to the gains.

In another domain of disclosure – software vulnerability and patches to fix them – the research is mixed on the role for market discipline because market structure affects the role and channels of disclosure. Arora, Forman, Nandkumar, and Telang (date) find that disclosure of software patches was quicker with more vendors subject to vulnerability, and that reputation (reputation is hurt if the patch is not released to customers fast enough) is an important disciplining channel. On the other hand, the model shown by Kannan and Telang (2005) concludes that market discipline over software vulnerabilities would be dominated by the proverbial 'social planner'; e.g. an omniscient and transaction-free regulatory agency not the market would best close the private-social gap created by the information externality.
There is evidence that consumer limited rationality problem noted above affects the role that disclosure can play. Survey evidence from Poneman Institute (2012) indicates that 85% of consumers are very concerned about data breaches. Comparing 2012 with 2005, twice as many consumers recall receiving a notification of a data breach (25% vs. 12%); the number of those who did not receive such a notification fell from 69% to 24%; but the share that could not recall increased from 19% to 51%. Tellingly, about 60% thought that the communication informing them of the breach was 'junk mail.' Thus, data security is a concern, but people don't recall whether they have received disclosure or not. From the standpoint of loss-of-trust disciplining the market, nearly 90% said they had or might discontinue their relationship with the firm over a data breach. So disclosure may matter, but communicating the right information seems to be a challenge. Further, Retzer (2008) notes that, in consideration of disclosure rules for the European Union, announcing every breach, regardless of size or potential harm, either to individuals (as in the U.S.) or to government agencies (as in Japan) desensitizes the recipient to the announcement, which works counter to the role that disclosure should play as a disciplining device.

What are the trends in type of disclosure and to whom? A Google Alert using the term 'data breach' has been running for about 250 days. The table below breaks down the alert into articles that are consumer-facing, articles that are firm-facing, and articles that are security-professionals facing. More articles are firm-facing and security-professionals facing, suggesting an internalization and standardization of information

protection as a response to data breach. This is consistent with survey evidence on the growing role and prevalence of Chief Information Officer.

Placeholder for table…

Market discipline through lost business and remediation costs

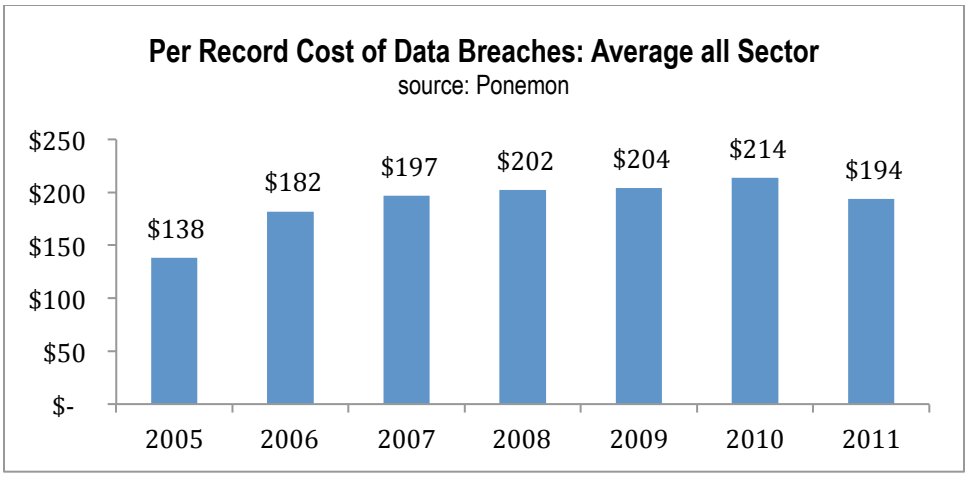
If remediating after information loss is sufficiently directly costly to a business, then presumably that business will undertake action to improve information security. The previous discussion of structure in the information marketplace is, however, quite relevant when considering this market signal. Moreover, quantification of *sufficiently and directly costly… to whom* in the information network remains an issue

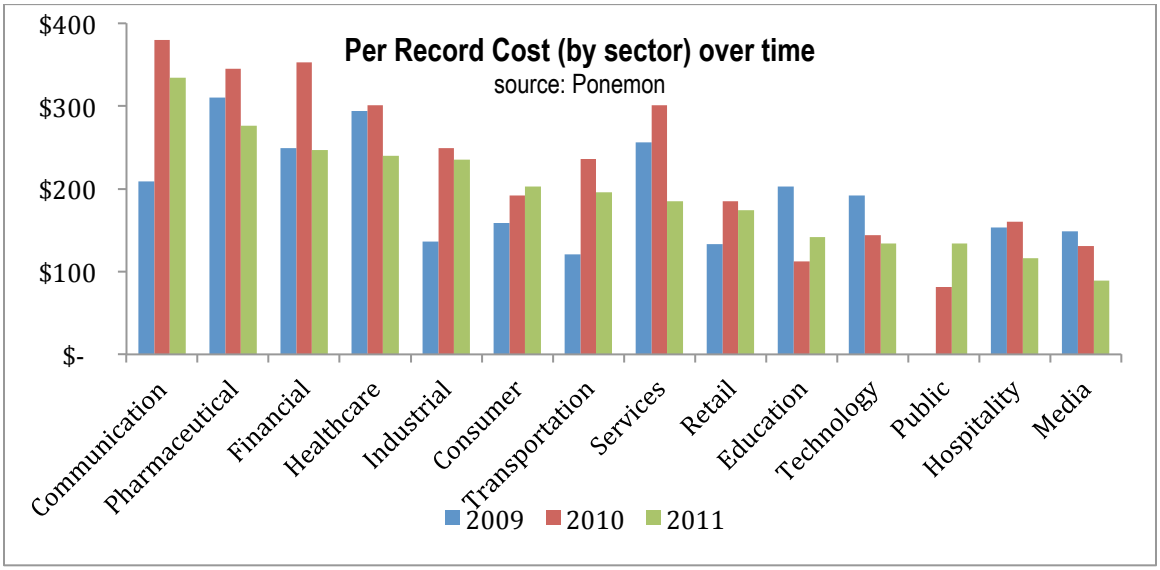
Calculations suggest that the loss of business due to customer turnover and reputation loss can be the relatively more important cost of a data breach, particularly in the U.S. On the other hand, remediation costs (customer notification, assistance, audits) are increasingly important. (Ponemon, 2011). Adding the two together, for the U.S. although the per record cost of a data breach remains relatively static 2005-2011, the number of breaches is rising, so the overall cost to society is increasing too. [5]

A back-of-the envelope calculation suggests that the macroeconomic cost of information lost in the U.S. in 2011 was: $40 billon.[6] Is this large or small in a $14 trillion economy? Seems small. But, is it large or small relative to business investment in equipment and software of $1 trillion gross, but $107 billion net (BEA NIPA 5.2.5)—more likely so. The distribution of these costs across sector and size of firm is key for whether the market discipline will work. But for policymakers, the macroeconomic size may be the most relevant for considering intervention.

[5] The cost per record lost as presented by Poneman is calculated only for breaches of 100,000 records or less. Against these costs, it is possible for a firm to assess the benefits of engaging in better information security. For example, Symantec now offers an on-line calculator for potential risk of information loss: http://eval.symantec.com/flashdemos/campaigns/small_business/roi/

[6] $200 per record lost and 200 million records lost. 50 million reported records lost with SSN, grossed up by 2 (about half of breaches reveal SSN) and grossed up again by 2 (about half of SSN breaches disclose the number of records).





However, whereas small breaches are the largest number of breaches, in any given year one or another sector experiences a disastrous information loss, which dwarfs all the other data breaches in terms of records and SSN compromised. So far, there is little systematic quantification of business and remediation costs, and how those costs are borne by the various market participants (consumers, aggregators, users). Moreover, information lost in the context of infrequent but disastrous breaches, and the probability of those breaches taking place in any particular sector, is much less well developed.

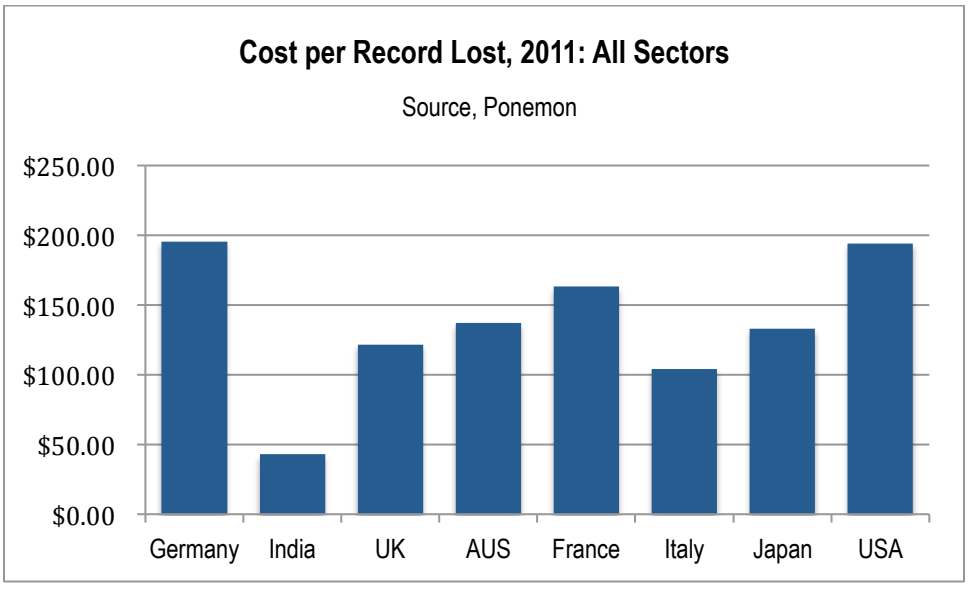
Further, is the likelihood of an infrequent but disastrous breach related to the nature of market structure and/or type of regulatory environment? A multi-player market place with large entities with market power (aggregators, transmission network) surrounded by information originators (consumers) and users (firms) may engender a self-regulatory focus on the small or distant market participants rather than the large players. And, a fragmented regulatory structure that focuses on sector-specific regulation (such as in the U.S. viz. HIPAA, GLB) may tend to make the regulators focus on one sector after a big

breach, rather than considering that the probability of a huge breach is not related to a sector but related to information type or type of breach.

Across countries and sectors there is substantial variation in costs per record lost, and also substantial variation in the components of the costs per record lost as disaggregated into detection, notification, post-breach costs, and lost business costs. However, notably, these costs do not seem to depend on level of income of the economy and vary substantially across countries within the same jurisdiction (e.g. the EU).

First, consider the comparison across countries of overall costs per record lost. Countries that have a lower per capita income (India) have a lower average cost of records lost. This reflects the lower domestic costs in general (the Balassa-Samuelson effect). On the other hand, countries with similar regulatory environments (Germany, France, Italy in the EU) have what looks to be significant variation in the cost per record lost. Is this a reflection of different costs in different sectors, and the share of the sectors in the economy or other factors?

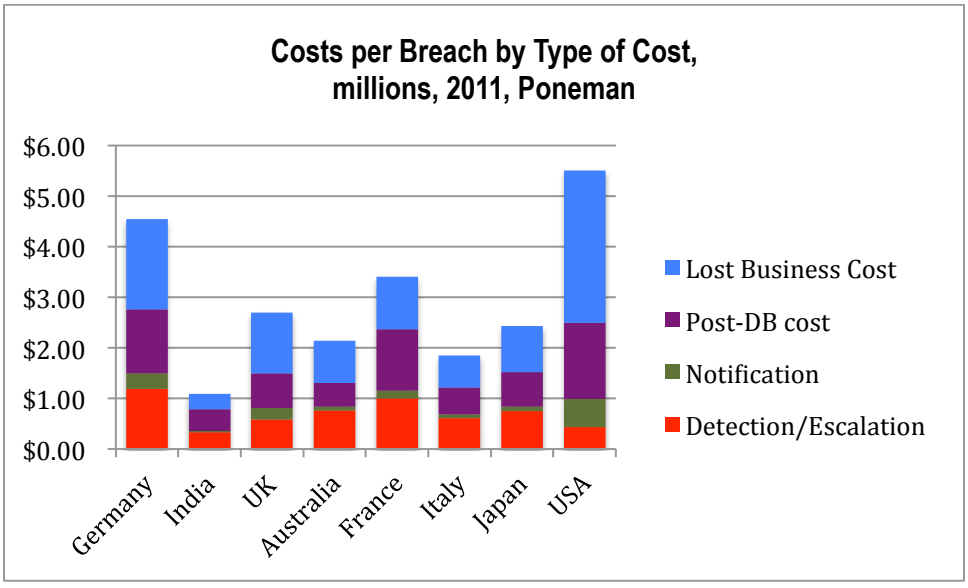
Comparing sectoral variation across countries, it is not the case that countries with high costs (Germany) or low costs (India) have the highest or lowest cost in all sectors. For example, costs in the communication sector in Germany are quite low but costs in the communication sector in India are rather high. Considering particular sectors across all countries, costs in the financial sector are highest among sectors (although not in India). But otherwise there is not a clear pattern where costs in certain sectors are always highest or lowest.



Per record by sector (2011)	Germany	India	UK	AUS	France	Italy	Japan	USA

Services	$344.92	$64.77	$135.59	$125.99	$176.87	$104.28	$195.39	$185.00
Industrial	$318.18	$38.43	$103.24	$134.92	$111.63	$101.60	n/a	$235.00
Hospitality	$290.11	$148.20	$147.92	$99.21	$145.05	$90.91	n/a	$116.00
Financial	$275.40	$40.57	$158.71	$199.40	$218.85	$147.06	$365.88	$247.00
Consumer	$203.21	$52.64	$147.92	$164.68	$189.84	$70.86	$105.08	$203.00
Retail	$149.73	$31.53	$92.45	$84.33	$100.80	$52.14	$96.45	$174.00
Technology	$147.06	$64.00	$92.45	$169.64	$195.45	$188.50	$328.51	$134.00
Public Sector	$129.68	$23.70	$95.53	$101.19	$75.94	$62.83	$84.18	$134.00
Communications	$89.57	$140.40	n/a	n/a	$128.61	$89.57	$124.96	$334.00
Pharmaceutical	n/a	n/a	$184.90	n/a	n/a	$131.02	$150.26	$276.00
Overall (not average of sectors)	$195.19	$42.85	$121.73	$136.90	$163.10	$104.28	$132.77	$194.00
Exchange Rate (2011)	0.748	49.124	0.649	1.008	0.748	0.748	82.931	

Consider a decomposition of what types of costs are incurred by a data breach (as opposed to costs per record lost). In all countries the notification cost is the smallest component (although largest in dollar terms in the US which makes sense given the California law), and the lost business cost (from customer churn etc) is the largest, and relatively larger for the U.S. On the other hand, detection of a breach is relatively larger for other countries compared to the U.S., suggesting that the California disclosure law may, over time, have make detection by U.S. firms more a matter of course (and therefore less expensive) rather than a special one-off event.



Another approach to considering how to discipline the market is to consider the value of the information lost. The variance of value in the market is quite large, and it is by information type, rather than by sector, so assessment is more difficult. As a point of comparison, however, the value of 'bank account credentials' at anywhere from $30 to $850 is compared to the per record lost of $250 in the financial sector (U.S. data).

Against the value of credit card information (worth 50 cents to $30 on the open market), the cost of each record lost in retail is $174. More information about the value put on different types of information would be useful to assess this approach to understanding the market for information loss and security response.

Value of Data	2007 (1H)		2007 (2H)		2008		2009	
	Low	High	Low	High	Low	High	Low	High
Credit Card Information	$0.50	$5.00	$0.40	$20.00	$0.60	$30.00	$0.85	$30.00
Bank Account Credentials	$30.00	$400.00	$10.00	$1,000.00	$10.00	$1,000.00	$15.00	$850.00
Full Identities	$10.00	$150.00	$1.00	$15.00	$0.70	$60.00	$0.70	$20.00
Cash-out Services			10%-50% of total amount		8%-50% or flat rate of $200-$2000/item		$0-$600 + 50-60%	
Email Accounts	$1.00	$350.00	$4.00	$30.00	$0.10	$100.00	$1.00	$20.00
Email Addresses (per mb)	$2.00	$4.00	$0.83	$10.00	$0.33	$100.00	$1.70	$15.00
Mailers	$8.00	$10.00	$1.00	$10.00	$2.00	$40.00	$4.00	$10.00

Stock market discipline appears to have limited effect

The direct cost of a data breach is not the only way in which market discipline can work. A number of studies investigate whether the stock market 'punishes' firms that lose customer data. (See Table 1 Appendix) These papers use the same methodology--cumulative abnormal returns—but differ somewhat in the time horizon over which they calculate the 'normal' return as well as the window over which they calculate the CAR. They differ in the measure of the market against which to assess the abnormal return. There also can be a difference in terms of whether to measures losses as a percent of stock market value or in dollars. On balance the stock market discipline appears limited in most cases as a strategy for aligning private and social incentives when it comes to protecting information against loss.

With regard to loss in shareholder value in percentages terms, the predominant conclusion, however, is that there is a negative, short term, statistically significant effect of a breach disclosure announcement on the financial performance of the announcing firm. The conclusion appears only when classified customer information was lost. Campbell sums up the findings: "we do not find a significant market reaction when we examine security breaches that are not related to confidentiality. In contrast, we find a highly significant negative reaction for those breaches that relate to violations of confidentiality." Considering sector-specific comparators, rather than the broad market indicators, as in Karagodsky and Mann, suggests that relatively larger CAR loss associated with data loss by banks and by healthcare firms and when SSN are lost. This is consistent with the aggregated research that found losing personal financial and health data were punished to a greater extent.

Firm characteristics may play a role, although the conclusions are mixed. Gatzlaff and McCullough who find strong and persistent effects up to 35 days after the event do not

find that the type of data lost matters, although firm characteristics, such as higher market-to-book ratio exacerbate CAR whereas larger firms mitigate the negative impact. Karagodsky and Mann find a smaller negative CAR when data are lost by mistake by an 'insider' as compared to data stolen through an intrusion by a hacker. Cavusoglu et. al. find that stock valuation of larger firms appear to be less affected. In contrast, Acquisti et. al. who also consider firm characteristics suggest that large firms might be more significantly affected by negative reports about their privacy practices as a result of irreversible damage to their reputation.

Are these results economically large, that is, compared with what it might cost to put into place security systems and procedures to avoid information loss? Karagodsky and Mann evaluate the dollar losses for four representative firms, one from each sector (Bank, Retail, Computers, Health) by using the findings on CAR and calculating the cumulative decrease in the firms' value 30 days following the breach announcement event. The cumulative dollar loss ranged from $170,000 (J.P. Morgan Chase and Gap), to $1 billion (IBM) and $7.5 billion (Pfizer). This calculation depends not only on the loss per share, but also the number of shares outstanding. Firms with more shares outstanding experience a larger dollar loss, and the loss can be quite large. Whether such dollar losses are large enough to incentivize firms to increase information security depends on the cost of those systems and procedures, a topic beyond the bounds of this paper, but which is critical to the cost-benefit analysis of the Digitization Agenda.

Multifaceted policy intervention: Standardization, regulations and fines

Policy intervention has many possible faces: standardization, regulation, and enforcement through fines. A key question is, who should be the policy target? The consumer or the various firms that acquire, transmit, and aggregate information throughout the information value chain? It is worthwhile to remember the nature of the gap between social and private benefits and costs, since any policy intervention is warranted by the gap.

Oussayef (2008) and Orr (2012) suggest that the focus should be on the consumer – the originator but also ultimate user of information value. A response to the consumer's limited rationality would be to standardize communications with them, for example, standardize privacy policies. But, generally this is not the direction that regulation or the market is going. Standardizing an approach toward the consumer fails to acknowledge the other aspects of the imperfect market: relative market power of the actors and the aggregation benefits of the information in database form. If standardization is warranted as a strategy, a focus on the firm (aggregator and user of aggregated information) or intermediary (holder or transmitter of information) would be a more economically efficient approach since the value of aggregated information is greater than the sum of each individual piece of information.

However, standardization of regulations with regard to these two (or more) types of firms in the information marketplace also is problematical. It it is not obvious who should decide on what standards to require. Coordination failures along the firms in the

information transmission and aggregation chain leave information exposed to loss. Even without a market failure, some countries may want a specific level of safeguard for transactions or for some sub-set of information in their jurisdiction, which makes standardization of information transmission across international borders particularly problematical. Both between the US and EU and within the EU these issues are particularly germane.

Data-in-transit vs. data-at-rest as the locus for security attention. US focuses on data-at-rest. EU focuses on data-in-transit. This is because of the underlying view on the role for information in driving private sector activities built on information. EU has limited the length of time private information can be retained, initially under the Data Protection Directive (1995) and reaffirmed under the Data Retention Directive (2006). U.S. has never limited the private sector's length of time for data retention; for public data there is a five-year retention.

The US-EU Safe Harbor Agreement discussed in more detail in Mann, Eckert, Knight (2000) remains the operational agreement governing cross-border information flows between the U.S. and Europe. US firms operating within Safe Harbor engage in self-certification, fill out forms with the Department of Commerce, and must have some type of privacy enforcement program in place. However, the Safe Harbor Agreement does not address the transmission of information across third party jurisdictions. For example, does information associated with packages flown by FedEx through Singapore need to abide by EU Directives, is it covered under the US Safe Harbor?

The January 2012 EU General Data Protection Regulation as promulgated tries to address information security both within the EU and between EU firms and firms in other countries. It would harmonize regulations for all members. Presumably this may cause some standards to be loosened—Germany? and others to be tightened. It extends these rules to all foreign companies processing data for EU citizens. Whether the US-EU Safe Harbor, or other individually-approved safe harbor arrangements would continue as is is a question. Moreover, whether all entities within a multinational enterprise would have to abide by the EU standards is a question. Disclosure of data breaches (to a supra-national entity or to consumers?) would be within 24 hours, which is quite a switch from no disclosure now. Fines for (any size?) data breaches could be up to 2% of global revenue. EU firms could not transmit data to countries with insufficient protection. Determination and enforcement are both issues; right now data transit is allowed to Argentina, Canada, Iceland, Norway, US under the Safe Harbor, and to various important financial centers--Switzerland, Lichtenstein, Isle of Man and Guernsey.

Within the U.S. the Federal Trade Commission has been playing a more active role. The grounds for FTC action is Contract Law: Firms that lose data are breaking the terms of service based on privacy statement. Fines can be large: $800,000 fine Spokeo under Fair Credit Report Law [7]. But they don't always work to change behavior. (Wyndham has

[7] Edward Wyatt, F.T.C. Levies First Fine Over Internet Data NYTimes.com, June 12, 2012,

been fined three times)[8]. Another strategy is the mandated audit: Many years and a big price-tag should change the balance between which is more costly: to protect vs. probability multiplied by cost of loss. Importantly, however, the advocacy of the FTC has a political lifespan.

Legal recourse: Evolving notion of 'standing' and scope

The role of the legal profession is evolving, and may play a more important role in firms undertaking appropriate security. However, given the international origins of data breaches, the legal approach may work by demanding greater security attention by moving the costs of a data breach from the originator (in a foreign country) to the U.S. intermediary.

Initially, and still true in general today, courts have not found that data breach cases have no standing because the link between a data breach and any future potential use of that data for harm cannot be proved ex ante. Simply losing data (as in losing property) is not sufficient grounds, since in the U.S. there is no right to privacy. As well, the costs of information loss have heretofore been unquantified (this is changing, see above). Without any legal standing, there is no incentive for firms to improve data protection. Yields problem of moral hazard—firms don't care. A potential new direction is to focus on 'industry standards': If firm experiencing a data breach did not employ 'industry standard' the courts are more likely to find against the firm, especially if data are used inappropriately. (YouRock)

Evidence also suggests that the legal approach may be beginning to have traction as a disciplining device in the cases of compromise of financial and medical information. Romanosky, Hoffman, and Acquisti (2012) find that the probability that a firm will be sued is 3.5 times higher when financial data are involved. Settlement is 30% more frequent when there is allegation of financial loss, even higher for compromised medical information. The possibility of a class action lawsuit also raises the likelihood of a settlement.

Finally, the risk of Class Action suits ('Ambulance chasing') appears to be increasingly important to legal consultants. (Gibson Dunn) Poneman indicates that legal defense costs have risen steadily, from accounting for 6% of costs (2006) to 15% of costs in 2011. Increased legal costs and threats of legal costs increase incentives for firms to take evasive action/or protect data to avoid becoming embroiled, even if the case won't go

[8] "The Federal Trade Commission filed suit against global hospitality company Wyndham Worldwide Corporation and three of its subsidiaries for alleged data security failures that led to three data breaches at Wyndham hotels in less than two years. The FTC alleges that these failures led to fraudulent charges on consumers' accounts, millions of dollars in fraud loss… http://www.ftc.gov/opa/2012/06/wyndham.shtm In response, Parsippany, N.J.-based Wyndham moved to dismiss the complaint… saying.. that the FTC "singled out" Wyndham in "unprecedented litigation." … that the commission has neither the expertise nor the statutory authority to establish data security standards for the private sector," http://www.scmagazine.com/wyndham-hotels-challenges-ftc-security-suit-over-breaches/article/258559/

against them. But these potential legal costs also cause firms to push-back against disclosure, particularly of the magnitude and sensitivity of information lost.

V: Conclusions and Considerations for the Digital Agenda

1. Market structure is characterized by externalities where the value of aggregated data is greater than the sum of any individual's observations and by differential market power (atomistic individuals vs. concentrated aggregators and transmission). This suggests that a pure-market solution cannot be achieved and that there is a role for explicit government intervention.
2. Infrequent but disastrous events happen in all sectors. More data on the probability, characteristics, and consequences of the large infrequent breaches is needed to consider an appropriate market and policy response.
3. Information value, cost of loss, and size of breach differ by sector. The market approach tends to focus on small numerous breaches, whereas the large infrequent breaches 'happen.' Simple technological strategies may address small breaches by reducing the costs of prevention and encouraging best-practice security implementation.
4. Sectors differ in the intensity of key information collected, aggregated, used and potentially lost (e.g. prevalence of SSN varies by sector). An information policy that focuses on the information type may be warranted, rather than a sector-based strategy.
5. Countries differ in their emphasis for information security, data in transit vs. at rest. There should be a presumption of using technological best practice, which would also stem arbitrage across jurisdictions. How to implement technological best practice when firms vary in size or countries vary in resources to spend on security is less clear.
6. Countries differ in their approach to disclosure. Disclosure is key to guide firm and policy response. However, who should be notified is less clear.
7. How countries implement policy in the information marketplace has some foundation in historical and cultural underpinning, e.g. EU mandate vs US market emphasis. Unlike technological best practice, it is much more challenging for these two approaches to co-exist.
8. Jurisdictional arbitrage and enforcement problems are significant issues that leave information exposed. Whereas name-and-shame strategies may prevent firms from routing through or storing data in low-protection jurisdictions, without a closed network (such as inter-bank SWIFT) there will always be gaps. This may point, again, to technological approaches to protecting certain key information to reduce the costs of information lost.
9. Should there be an 'insurance fund' to pay-off individuals facing financial damage? Tiny taxes on transactions would create a big fund quite quickly, and therefore would create the problem of moral hazard. That is, with the insurance fund in place, there is less incentive to follow data safeguard mechanisms. (Consider the incentives created by flood insurance, for example

References

Articles

Acquisti, Alessandro (2010) “The Economics of Personal Data and the Economics of Privacy” draft dated November 21, 2010.

Acquisti, Alessandro, Allan Friedman, and Rahul Telang (2006) Understanding the Impact of Privacy Breaches *35th Research Conference on Communication, Information and Internet Policy (TPRC)*.

Anderson, Horace E. (2006) “The Privacy Gambit: Toward a Game-Theoretic Approach to International Data Protection,” Pace University Law Faculty Publications, 1-1-2006.

Arora, Ashish, Anand Nandkumar, Rahul Telang (XXXX) Does information security attack frequency increase with vulnerability disclosure? An empirical analysis”

Ashish Arora, Christopher M. Forman, Anand Nandkumar1 and Rahul Telang (xxxx), “Competitive and strategic effects in the timing of patch release”

Arrow, Kenneth J. and Gerard Debreu (1954) "Existence of an equilibrium for a competitive economy". *Econometrica* **22**: 265–290. doi:10.2307/1907353

Bamberger, Kenneth A. and Deirdre K. Mulligan (2011) “Privacy on the Books and On the Ground,” Stanford Law Review, Vol. 63:247, 247-315.

Campbell, K., L.A. Gordon, M. P. Loeb and L. Zhou, (2003) The Economic Cost of Publicly Announced Information Security Breaches: Empirical Evidence from the Stock Market, *Journal of Computer Security, Vol. 11, No. 3.*

Carty, Matt, Vincent Pimont, David W. Schmid (2012) Measuring the Value of Information Security Investments, IT@Intel White Paper, January.

Cavusoglu, Huseyin, Birendra Mishra, and Srinivasan Raghunathan. (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 / Fall, Vol. 9, No. 1, pp. 69–104*

Federal Deposit Insurance Corporation (2004) Offshore Outsourcing of Data Services by Insured Institutions and Associated Consumer Privacy Risks, June.

Gatzlaff, Kevin M. and Kathleen A. McCullough. (2010) The Effect of Data Breaches on Shareholder Wealth, *Risk Management and Insurance Review, Vol. 13, No. 1, 61-83*

Gibson Dunn (2012) “2011 Year-End Data Privacy and Security Update” February 7.

Greenstein, Shane and Ryan McDevitt (2011) "The Global Broadband Bonus: Broadband Internet's Impact on Seven Countries," in The Linked World: How ICT Is Transforming Societies, Cultures and Economies, The Conference Board.

Greenstein, Shane and Ryan McDevitt: (2009) "The Broadband Bonus: Accounting for Broadband Internet's Impact on U.S. GDP," NBER Working Paper #14758.

Hann, Il-Horn, Kai-Lung, Hui, Tom S. Lee, I.P.L Png, (2002) "Online Information Privacy: Measuring the Cost-Benefit Trade-Off," Twenty-Third International Conference on Information Systems.

Hirsh, Dennis D. (2006) "Protecting the Inner Environment: What Privacy Legislation Can Learn from Environmental Law" *Georgia Law Review*, vol 41 no. 1, p1-62.

Kannan, Karthik, Jackie Rees, and Sanjay Sridhar. (2007) Market Reactions to Information Security Breach Announcements: An Empirical Analysis, *International Journal of Electronic Commerce / Fall, Vol. 12, No. 1, pp. 69–91*

Karagodsky, Igor and Catherine L. Mann (2011) "Do Equity Market Punish Firms that Lose Customer Data?

Mann, Catherine L. (2001) "International Internet Governance: Oh, What A Tangled Web We Could Weave!," *Georgetown Journal of International Affairs*, Summer/Fall.

Mann, Catherine L. and Diana Orejas (2003), "Can the NAFTA Partners Forge a Global Approach to Internet Governance?" in North-American Linkages, Richard G. Harris, ed. Ottawa: Industry Canada, 2003.

Mann, Catherine L., Sue E. Eckert, Sarah Cleeland Knight (2000) Global Electronic Commerce: A Policy Primer. Institute for International Economics: Washington DC.

Morton, Fiona Scott (2006) "Consumer Benefit from Use of the Internet", in Adam B. Jaffe, Josh Lerner, and Scott Stern eds. Innovation Policy and the Economy, vol 6 . The MIT Press.

Oussayef, Karim Z. (2008) "Selective privacy: Facilitating Market-Based Solutions to Data Breaches by Standardizing Internet Privacy Policies," Boston University Journal of Science and Technology Law, Vol. 14:1, 104-131.

Orr, Madolyn (2012) "Foxes Guarding the Henhouse: An Assessment of Current Self-Regulatory Approaches to Protecting Consumer Privacy Interests in Online Behavioral Advertising," *www.ftc.gov/os/comments/privacyreportframework/00231-57343.pdf*

Ponemon Institute (2012) "Consumer Study on Data Breach Notification," sponsored by Experian Data Breach Resolution, June.

Retzer, Karin (2008)"Data Breach Notification: The Changing Landscape in the EU,"

Roberds, William and Stacey Schreft (2009) "Data Breaches and Identity Theft," Journal of Monetary Economics, 56, pp 918-929.

Romanosky, Sasha and Alessandro Acquisti (2009) "Privacy Costs and Personal Data Protection; Economic and Legal Perspectives" Berkeley Technology Law Journal vol 24 no 3, pp 1061-1101.

Romanosky, Sasha, Alessandro Acquisti, and Richard Sharp (2010) "Data Breaches and Identity Theft: When is Mandatory Disclosure Optimal?", TPRC 2010. Available at SSRN: http://ssrn.com/abstract=1989594

Romanosky, Sasha, Rahul Telang, Alessandro Acquisti (2011) "Do Data Breach Disclosure Laws Reduce Identity Theft?" *Journal of Policy Analysis and Management*, Vol. 30, No. 2, pp. 256-286.

Romanosky, Sasha, David A. Hoffman, and Alessandro Acquisti (2012) "Empirical Analysis of Data Breach Litigation," Temple University Legal Studies Research Paper No. 2012-30. Available at SSRN: http://ssrn.com/abstract=1986461 or http://dx.doi.org/10.2139/ssrn.1986461

Stevens, Gina (2012) "Data Security Breach Notification Laws," Congressional Research Service, R42475.

US Department of Commerce, National Telecommunications and Information Agency (date?) "Chapter 1 – Theory of Markets and Privacy," http://www.ntia.doc.gov/page/chapter-1-theory-markets-and-privacy

Tang, Zhulei; Hu, Yu Jeffrey; and Smith, Michael D. (2007) "Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor," Heinz Research. Paper 49. http://repository.cmu.edu/heinzworks/49

Yan Chen, Grace YoungJoo Jeon, Yong-Mi Kim (2012) "A Day without a Search Engine: An Experimental Study of Online and Offline Searches," (location)

Appendix Table 1: Summary of literature review of equity market effect of data breach

Author	Days to calculate market model	Market index	Interval for CAR calculation	# events in the dataset	Time period covered	Mean CAR % loss by window (reported if significant)
Campbell, et. al.	121	NYSE AMEX NASDAQ	-1 to +1	43	1997-2000	−0.02
Acquisti, et. al	92	NYSE NASDAQ	0 to +1 0 to +2 0 to +5 0 to +10	79	2000-2006	-0.58 -0.46 0.21 1.3
Cavusoglu, et. al	160	NASDAQ	2 days Day 0 Day +1	78	1996-2001	Not signif –0.0086 –0.0123 (check magnitudes
Kannan, et. al	50	SIC codes control group S&P 500 index	-1 to +2 -1 to +7 -1 to +29	72	1997-2003	-0.65 -1.4 2.22
Gatzlaff and McCullough	245	Value-weighted S&P500 index	Day 0 0 to 1 0 to x in one day increments to 0 to +35	77	2004-2006	-0.57 -0.84 avg: -0.74
Health study						
Karagodsky and Mann		NYSE, NASDQ, Ken French Sectors 1. banks 2. health 3. technology 4. retail 5. insurance 6. brokers	Day +1 Day -1 to + 7 Day -1 to +7			-0.7% range: 1% - 1.3% 1.2% 2.5% no loss 1% no loss no loss

Data sources (incomplete) :

DLDOS, open security foundation public database. Further information about the database is available at http://attrition.org/dataloss/dldos.html. [this database is no longer, as of first quarter 2012, available for immediate download].

Privacy Rights Clearinghouse, www.privacyrights.org/**data-breach**

Ponemon Institute LLC (2011) "2011 Cost of Data Breach Study United States', Sponsored by Symantec, March.

….

http://www.verizonenterprise.com/resources/reports/rp_data-breach-investigations-report-2012_en_xg.pdf?__ct_return=1
http://www.symantec.com/threatreport/
http://www.idtheftcenter.org/artman2/publish/headlines/Breaches_2011.shtml
http://www.symantec.com/content/en/us/about/media/pdfs/b-ponemon-2011-cost-of-data-breach-global.en-us.pdf