NBER WORKING PAPER SERIES

INFORMATION LOST (APOLOGIES TO MILTON)

Catherine L. Mann

Working Paper 19526
http://www.nber.org/papers/w19526

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2013

Excellent research assistance from Alok Mistry, who experienced his own data breach (stolen laptop) during the course of this project. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Information Lost (Apologies to Milton)
Catherine L. Mann
NBER Working Paper No. 19526
October 2013
JEL No. F29,F5,L19,L86

ABSTRACT

Vast amounts of information result from business and consumer search, communication, and transactions. All this information can enhance market efficiency and consumer surplus as firms tailor products to buyers. But, there is increased risk of information loss. What issues should be on the Digital Agenda with regard to information loss, and what data are available to inform and generate incentives for consumer, business, and policy interactions in the information marketplace? 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 aggregation and protection? (2) Quantification and data: What is the evidence on the prevalence and nature of information loss, and what are the costs of information loss, and to whom? (3) Market and Policy Response: What do we know about the efficacy of market vs. other approaches to incentivize market participants to avoid loss or remediate after information loss? Throughout, of particular interest is the international dimension of the information marketplace. What issues arise when countries differ in their attitudes and policies toward the information marketplace?

Catherine L. Mann
415 South St
Brandeis University
Waltham MA
CLMann@Brandeis.edu

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 consumers' ease of transacting, range of complementary purchases, targeted news and advertising, and other directed goods, services, and information, all of which increase customer value—but which also raise the probability and consequences of 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 which also raise the probability of losing or abusing information. The Digitization Agenda can help frame and balance the benefits to firms and consumers of information gained with the risk and costs of information lost, particularly in the context of increasingly global flows of information and transactions.

A first priority is a conceptual framework. Three key elements in the structure of the information marketplace influence the valuation and balancing of benefits and costs. First, information exhibits economies of scale and scope, which challenges the ability of the market to efficiently price information. Second, participants in the information marketplace are not atomistic, rather are asymmetric in terms of market power, which affects the incidence and distribution of benefits and costs. Third, information loss is a probabilistic event, but with unknown distribution, which challenges the valuation of benefits and, particularly, costs. A final element is that the information marketplace is global, populated by heterogeneous firms and consumers, and by policymakers who differ in their policy responses to the imperfections in the marketplace.

A second requirement is empirical analysis of the frameworks. Mandated disclosure in the United States of data breaches was the watershed to enable this study, and its references. Without disclosure, it is impossible to investigate the risks and potential costs of information loss as against the benefits of information collection and aggregation. Disclosure helps reveal to consumers, firms, and policymakers the nature of data loss, and may change incentives and affect the incidence and balancing of costs and benefits. However, disclosure can be along a spectrum from every incident being announced to everyone to only critical incidents being communicated to a few. In fact, there is no global consistent approach to disclosure, nor even to the notion of disclosure at all, so the window into the empirical valuation of costs and benefits of the information marketplace is narrow.

Even so, evidence on how disclosure works is starting to emerge. If the market response to disclosure is sufficient to apportion and balance costs and benefits, then, in principle, no policy intervention into the marketplace is needed. So far, this does not appear to be the case. More information on the nature of data breaches, on incidence of benefits and costs, of market participant 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 conceptual frameworks with which we can analyze the 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, including differences in attitudes and priorities toward data security. 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). Wallsten (in this volume) continues the work to value the consumer benefits of using the Internet. Yet, with rapidly changing technology and social interaction, it is hard to pin down to exactly how large the increase in consumer surplus might be and so there is much more work to do.

Using the Internet generates the information that is the basic building block of the marketplace for information. A conceptual framework for the incidence and balance of costs and benefits of information in this marketplace includes the valuing of consumer gain from using the Internet, but it is a more complex framework with more players. For simplicity, suppose the information marketplace is populated by originators of information (say consumers, as they reveal their preferences through search and transactions); intermediaries of the information (say, firms that transmit data, and those that collect, aggregate, and retain information); and final users (say, firms that call on the aggregated data to improve products). How should we model the interactions between the information and the three players? Is the information atomistic or are there economies and scale and scope in aggregation of the information into a database? Are the players atomistic and equally numerous, or do they differ in concentration and market power in their economic relationships? What about the nature of uncertainty? Answering these questions help to determine to what extent the information marketplace is 'classic' in the Adam Smith sense and 'complete' in the Arrow-Debreu sense, or whether it is a market with imperfections.

Various authors have taken up the challenge of modeling the information marketplace, some explicitly in the context of data breaches. The several papers reviewed below are put into the context of a general framework that focuses on a market structure that includes economies of scale and scope in data aggregation, multiple non-atomistic players, an in an environment of uncertainty over the nature of data breaches and consequences. In this kind of market structure it is challenging to value the cost and incidence of a data breach. Further challenges are that the basic building blocks of information may be valued differently across geographies and cultures, and the probability of a data breach may have important attributes that factor into the calculation of costs and benefits in the information marketplace.

Complete markets: The benchmark market structure

The purpose of outlining the characteristics of the perfectly competitive marketplace—the Adam Smith marketplace—is to provide a benchmark against which the structure of the global information marketplace 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.

In Adam Smith's market, one-off transactions generate unique prices for each transaction. In this classic marketplace, buyers, intermediaries, and sellers are all atomistic. There are no databases with a history of a specific buyer's transactions or those of buyers of similar characteristics that create correlations between transactions across time or across individuals. No information is retained, so no information can be lost. Balancing the benefits of information exchange with the potential cost of information lost is not an issue.

An extension of Adam Smith allows for transactions across time, proximity, currency, and uncertainty. In the so-called Arrow-Debreu 'complete' market (Arrow and Debreu, 1954), 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 through time, space, and under uncertainty and yields a unique and market-determined price for that transaction in a frictionless world.

Whereas these transactions may be correlated and/or uncertain, the correlations of transactions (such as interest rates and exchange rates) and uncertainties (probability of default) are fully known (in the complete market), and therefore will be efficiently embodied in the relevant prices. In a complete-markets framework, both private and social optimum outcome can be achieved because there is a perfect (complete) and frictionless match between transactions and atomistic market participants over all possible states-of-nature and time. With full information about correlation and uncertainty, prices will fully reflect benefits of information exchange, which can then be balanced against the potential cost of information lost. There are no market imperfections.

The Information Marketplace: Violating the complete-markets framework

In a number of ways the information marketplace violates key assumptions of the complete markets framework, which makes pricing information difficult, and opens up for consideration the topics of market imperfections and problems of ranking the second-best. More specifically, without accurate prices, the benefit-cost calculation surrounding information exchange as against information lost through a data breach will be very challenging.

The first violation is the assumption that transactions are one-off or uncorrelated as in Adam Smith. In fact, information is characterized by economies of scale and scope.

That is, the value of information over a series of transactions for an individual is greater than the sum of the individual transactions because of the correlations across the individual's behavior: Economies of scale. The value of information aggregated over many individuals is greater than the sum of any individual's set because of the correlations across individuals: Economies of scope.

Even if each unique piece of information had a uniquely matched price, there would be an incomplete mapping between the value of that morsel of information by itself, its value in one database, and its value if two (or N) databases are merged together.[1] Databases, which are the product of the information marketplace, are characterized by economies of scale and scope so that the pricing of information is imperfect, unless there is full information about all the correlations. Because the information marketplace and its players are evolving rapidly with technology and Internet use, it is clear that the correlations needed for the complete-markets framework cannot be known.

The second challenge that the information marketplace brings to the complete-markets framework is the nature of uncertainty. Uncertainty enters the information marketplace through the possible mis-use of information. A complete markets set-up could, in theory, price insurance that pays-off in the case of a data breach, but such price determination in the information marketplace, in practice, is nearly impossible.

The information marketplace exhibits two types of uncertainty that are difficult to price. First and most challenging, is the potential correlation over time of information lost. A data breach today cannot be valued with certainty because the value of the information lost today is a function of all possible data breaches in the future. Future data breaches matter for today's valuation of information lost because of the unknown relationship between the information lost in today's breach with the information lost in a future breach. Economies of scale and scope in information in the future affect valuation in the present.

The second type of uncertainty is that information lost may not be information abused. The cost of information lost should differ depending on whether the lost information is used maliciously or not. The two uncertainties together make valuing information lost quite difficult. The insurance contracts, which are key instruments in the complete market framework, are not likely to exist.

The third violation of the assumptions that underpin the complete-markets benchmark model is that the players are not atomistic. Recall that the information marketplace has consumers (originators of information), intermediaries (transmitters and aggregators); and firms (that use information to improve products). Consumers are numerous. Firms are numerous. Transmitters and aggregators are concentrated and have several types of market power that will affect the price and value of information and therefore influence

[1] Another way to think about the economies of scope in information in databases is to consider the diversification gains associated with merging two not-identical financial portfolios.

cost-benefit calculations associated with information exchange, information security, and information loss. Moreover, the degree of market power and the rules under which the intermediaries operate vary substantially across countries and policy environments.

These violations of the complete markets framework offer jumping-off points for research. The following selected papers focus on modeling the information marketplace. Some papers specifically address how to model the cost-benefit calculation in the case of information lost. [2]

Applying the Pollution Model to Information Flows

Pollution seems like a good analogy for the information marketplace: Pollution has (negative) economies of scale, asymmetric market position of participants (upstream-downstream), and uncertainties as to costs and benefits of exposure and remediation. Hirsch (2006) uses the pollution model and focuses on the negative economies of scale. He presumes 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 is an outcome of production, so too is information aggregation an externality of 'production' (search and transaction on the Internet). In the pollution-model of the information marketplace, no data breach is necessary to generate harm. Aggregation alone departs from the complete-markets framework. With the economy of scale inherent in information aggregation, there will be a price wedge between the valuation of information by the consumer and by intermediaries and firms in the marketplace.

Hirsh continues with the pollution analogy and reviews the evolution of policy strategy from 'command and control' compliance (quantities) to 'second-generation' (prices) or 'outcome oriented' policy whereby the regulated entities find their own cost-effective strategy to achieve the legislated goal. Tang, Hu, and Smith (2007) take these strategies to the information marketplace. They model information collection looking through the lens of consumer preferences for trust. Standardized regulation does not map into the heterogeneity of consumer preferences for trust (with some consumers being 'too' regulated, others not enough) so overall economic well-being is reduced by such an approach. In contrast, they find that under circumstances of 'clarity and credibility' self-regulation can achieve a nuanced strategy that meets the heterogeneous preferences in the marketplace. On the other hand, Ioannidis, Pym, Williams (2013) argue the 'information stewardship' internalizes the social costs of data loss (e.g. pollution). With the prodding

[2] The literature addressed in this paper focuses on the benefits of information exchange and the costs of information lost. Other research focuses more specifically on the topic of privacy. For more on modeling privacy see U.S. Dept of Commerce, NTIA chapter compendium of articles. Roberds and Schreft (2009), Anderson (2006), and references therein.

of such a steward, firms internalize some of the costs of data loss and therefore undertake higher investments in information security than they would have. Social welfare is enhanced.

Whereas environmental economics offers a model for the information marketplace, the analogy is stretched because consumers and firms do gain from information aggregation, whereas it is hard to imagine anyone actually gaining from downstream pollution. Moreover, although the pollution model allows for market power and uncertainty, so far, researchers have not put all three elements of economies of scale/scope, market power, and uncertainty together in the context of the information marketplace.

Too Much Information: Trade-offs with Limits to Rationality

Full information and frictionless markets are key in the complete market framework. Acquisti (2010) starts by arguing that that the information marketplace is all about trade-offs. "In 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 intangible, and often ambiguous trade-offs. … But trade-offs are the natural realm of economics."

But then, he notes that limited consumer rationality and transactions costs make calculating these trade-offs difficult. Both of these issues affect the pricing of information, as well as the distribution of benefits and costs of information aggregation and potentially of its loss. If consumers don't know the value of their information, they cannot calculate the trade-off between allowing collection and aggregation against the possible cost of a data breach. These issues are the departure from the complete markets and Acquisti's (2010) jumping-off point for his modeling of the cost-benefit calculations. How significant are these departures in the information marketplace from the complete-markets framework?

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 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). Another way to value personal information is to calculate the cost to firms of the inability to use individual and aggregate personal information to target advertising (Goldfarb and Tucker, 2010). The limited empirical work on value of information to the consumer suggests that limited rationality is an important problem.

Policy makers and now businesses differ in their approach to the limited rationality of consumers. The EU Privacy Directive is at one extreme, disallowing the collection and retention of personal information on the grounds that consumers don't know what they are giving up. Other policy approaches require active consent (opt-in) or more transparency ('this website uses cookies… .click here for our cookie policy'). Some firms are finding a market opportunity in responding to the limited rationality problem.

Incorporated into the website are easy-to-use tools that allow customers to edit the information stream associated with their search and transactions activity and thereby improve the accuracy and targeting of their own information. [3]

However, the presence of economies of scale and scope in information aggregation as well as the nature of uncertainty regarding data breaches means that the analysis of the balancing of the benefits from information transmitted against the potential cost of information lost is more complex than the limited rationality of individuals.

Multiple players, market power, and the role for disclosure

Much of the literature that addresses benefit of information aggregation vs. cost when information is lost uses a two-player 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). In fact, there is a third player in the information marketplace—the intermediaries—through which information 'transits' and/or 'rests.' Examples range from VISA, Amazon, and Google to less familiar companies such as ChoicePoint or Acxiom.

Atomistic interaction among market players is an important underpinning of the complete-markets framework, but is clearly violated in the information marketplace. In particular, intermediaries are very highly concentrated: Google accounts for about 70 percent of all search[4], collecting and retaining all that information; VISA accounts for about three-quarters of all U.S. card transactions, creating a thick financial and purchase trail[5], and Amazon accounts for 15 percent of all U.S. on-line sales and is ranked fifteenth among all retail companies, collecting reams of data along the way.[6] On the other hand, there are billions of consumers and merchants that use Google and VISA and shop with Amazon. Virtually none of them interact with an intermediary such as ChoicePoint or Acxiom, although their information 'rests' there. The differential interactions and differential concentration are important for the valuation of information and magnitude and incidence of costs in the case of a data breach.

Considering interactions and concentration, Romanosky and Acquisti (2009) use a systems control strategy to map alternative legislative approaches to reducing harm from information loss. 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 intermediaries to, say, Payment Card Industry Standards. But, these authors argue that ex ante standards focus on inputs (encryption) rather than outcomes (harm); so are not efficient. Second, ex poste liability

[3] NYTimes: If My Data Is an Open Book, Why Can't I Read It? http://nyti.ms/12UHryv.
[4] Multiple sources as of April, May, June 2013.
[5] http://www.forbes.com/sites/greatspeculations/2013/05/03/visa-and-mastercard-battle-for-share-in-global-shift-to-plastic/
[6] http://www.prnewswire.com/news-releases/amazoncom-captures-28-of-top-online-retailer-sales-205427331.html

law (think law suits) could include fines for negligence in the protection of information. But, ex-post litigation may be ineffective 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 approach is disclosure of data breaches. Disclosure of data breaches is a key ingredient to calculating costs and benefits of providing and protecting information, and of apportioning responsibility and costs in the case of a data breach. But Romanosky and Acquisti note that consumer cognitive bias (misperception of risk) and costs of disclosing the data breach itself (disclosing what to whom) are important caveats for the effectiveness of disclosure.

Romanosky and Acqusiti use their framework to outline an empirical example of where cognitive bias and disclosure costs are less significant because of the concentrated market structure of intermediaries. Specifically, they analyze 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 who has provided that information cannot demand such protection.

Why does disclosure help align (some of the) private interests? First, a sufficient number of data breaches have occurred such that these costs have begun to be quantified (to be discussed in Section III and IV 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 intermediary, the data aggregators/holders. Third, the chain of causation between information loss and required remediation is revealed because of data-breach disclosure laws. The disclosure laws along with quantification of costs as well as the small number of players promote 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 cost of the data breach was internalized in this example.

However, the costs of information loss borne by individual card-holders was not transferred to those firms where the data breach occurred. The market power of individuals was insignificant, and in a transactions-sense, the individuals were distant from the data aggregators/holders. Individuals can change card issuers, but they have no power to affect the relationship between their card issuer and what firm aggregates the transactions of that card. Thus, the cost of the data breach incurred by individuals was not internalized by the intermediaries, and the individuals had no market power to affect such an internalization. Unlike the atomistic players in the complete market framework, the information marketplace has disparities in concentration and market power that affect the distribution of costs of a data breach, as well as the price and willingness to pay for techniques to avoid such a breach. (See more on disclosure in Section III and IV.)

The Probability Distribution of Data Breaches

The third key underpinning of the complete-markets framework is the pricing of uncertainty. For a number of reasons, it is challenging to estimate, and therefore price, the uncertainty of incurring and then the uncertain consequences of a data breach. Nevertheless, in the face of costly data breaches (see Section III) firms increasingly are turning to risk modeling for the decision to invest in information technology security.

However, the shape of the probability distribution of data breach events is crucial to calculate both the costs of a breach and benefits of undertaking security investments. Assuming that data breaches follow a 'normal' distribution will yield a different calculation than if data breaches are characterized by 'fat tails' or 'extreme outlier' distributions. (Thomas et al (2013) consider alternative probability distributions in a theoretical model of investment in information security.)

An analogy comes from the market for foreign exchange and the financial instruments that are priced and used in that market. Suppose a firm wants to put a floor on the value in the home currency of the revenue stream earned abroad in the foreign currency. In a complete-markets framework, the firm could buy an option that will pay-off when the home-to-foreign currency exchange rate reaches a particular value. In a complete markets framework, the probability distribution of exchange rate movements is fully known. The option would be priced exactly so as to make the firm indifferent to buying it or not (and on the sell side, the seller indifferent to selling the option or not.) The factor inducing one firm to buy the option and the other to sell the option is differences in risk appetite.

Suppose the probability distribution is not accurately parameterized. For example, suppose that exchange rate fluctuations are assumed to follow a 'normal' distribution, but the true distribution has 'fat tails'. The probability of the foreign currency depreciation that triggers the option will be underestimated relative to its value under the true probability distribution. The firm will not buy the option, and it will experience an uncompensated loss. On the other hand, if the firm assumes the extreme outlier distribution is correct, when the true distribution is normal, then the firm will buy too expensive an option, given the very small likelihood of the extreme event.

In the information market place there is a similar problem of deriving the correct probability distribution of a data breach. Information on the probability of incurring a data breach is limited; and incurring a data breach is not identical to the probability of data abuse. Without knowing the correct probability distribution, too much investment in information security or too little are equally possible. Moreover, whether the correct market player is the target of the security effort remains unclear. For example, Anderson et. al. (2012) point out that one automated spammer accounted for about one-third of global spam in 2010 and profited $2.7 million. But the 2010 worldwide spending on preventing spam exceeded $1 billion. So neither the level of spending nor the target appeared to have been correct.

But the challenges run deeper because of the economies of scale and scope in the information and the differential market power of the players. Does the cost-benefit calculation for information security differ as to many small breaches (say, the normal distribution) compared to a rare, but large, data breach (the black swan event). Is a large data breach more likely to lead to abuse of data, or less likely? The hypothesis of economies of scale and scope in information suggests that large data breaches, experienced over time, accumulate to enhance potential abuse of all revealed information, whether abused before or not. Finally, differential market power has already been seen to shift the burden of costs of a data breach; it could similarly shift the burden of responsibility to invest in information security. Free riding and moral hazard are other aspects of differential market power that cause the information marketplace to deviate from the complete-markets framework.

Information Marketplace: Challenges to pricing and balancing benefits and costs

The information marketplace violates the class complete-markets framework in three ways. First, information is characterized by economies of scale and scope, so it is difficult to price and value. Moreover, the benefits of aggregation increases, but so may the cost in the case is of information lost. Second, the various market players are not atomistic. The relationships between the originators of information; the intermediaries that transmit, aggregate, and hold information; and the users of the aggregated data to enhance products are characterized by differential marker power. The differential market power affects the distribution of both benefits of information and the potential costs when lost. Finally, there is substantial uncertainty about the probability distribution describing a data breach and potential abuse of information, so it is hard to value information lost. Collectively, these departures from the complete-markets framework point to potential inefficiencies in market pricing and in participant behavior. Whether such inefficiencies suggest policy-maker intervention requires more analysis.

III. Trends in Information Lost

The literature and framework presented in Section II pointed to a variety of data needs: Value of information incorporating economies of scale and scope. Nature of the market-power relationships between different market actors. Parameters of the probability distributions of information lost and/or mis-used. All this information is needed to evaluate whether the information marketplace is efficiently balancing the value of information aggregated against the costs of information lost.

Against this series of data needs, this section presents evidence on only 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, probability of abuse given a breach, and so on, including in the global context.

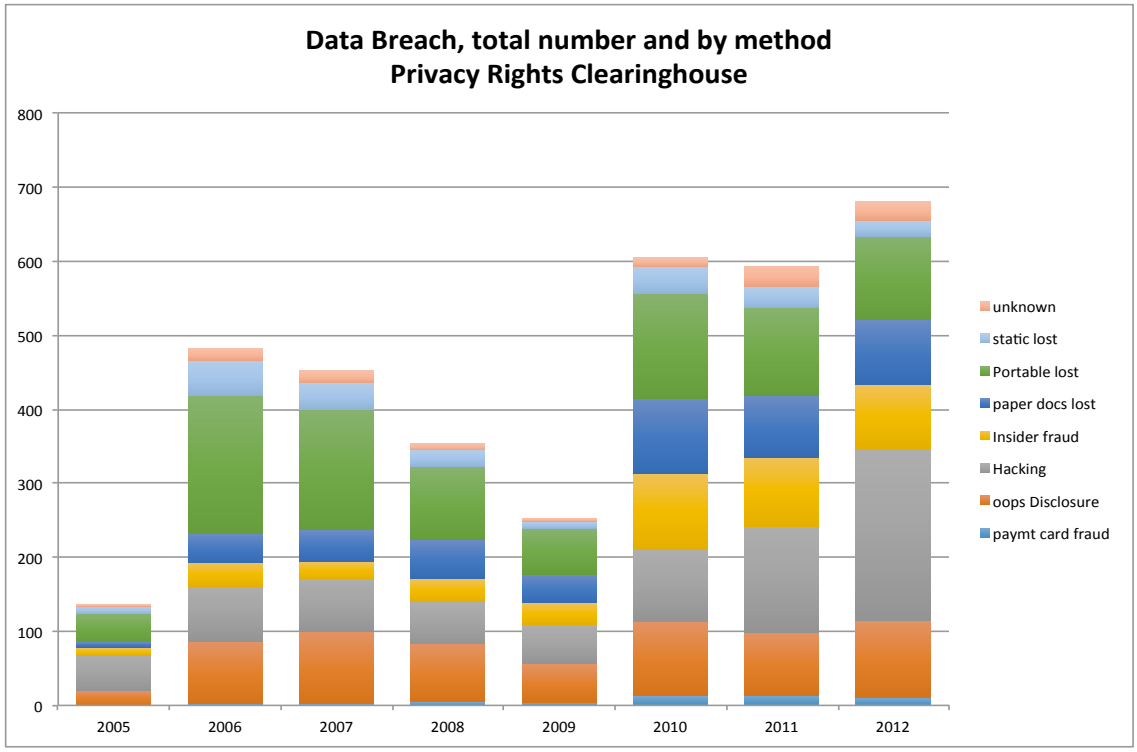
The raw data come from several sources including: Privacy Rights Clearinghouse and Open Security Foundation, which draw from public news sources. A number of consulting firms that employ industry surveys: Poneman Institute, Symantec, Verizon,

Javelin Security Research, KPMG Europe. The Federal Trade Commission and the Department of Justice, which draw on the consumer fraud on-line report database. Only some of the raw data are available for research use; most is proprietary.

How much information is lost? And by what means?

Privacy Rights Clearinghouse (PRC) data for 2005 to 2012 show that after a notable drop in data breaches in 2009, during the depth of the recession, data breaches are on the increase again. [7] (The number of records lost in each breach, which is a different measure of information lost, will be discussed below.) PRC disaggregate breaches into various types: [8] Losing paper documents or losing computers (static desktop or portable); inadvertent disclosure (such as cc vs. bcc in an e-mail); and various types of fraud (by an insider employee, by an outsider hacker, through payment card). The first three types of information lost are more 'by mistake,' although the disclosed information could still be mis-used. The three types of fraud are presumed with malicious intend.

Hacking dominates, and insider fraud is the increasingly important source of data breaches. But a surprising number of data breaches still take place the 'old-fashioned way' by losing paper documents or laptops and through unintended disclosure.

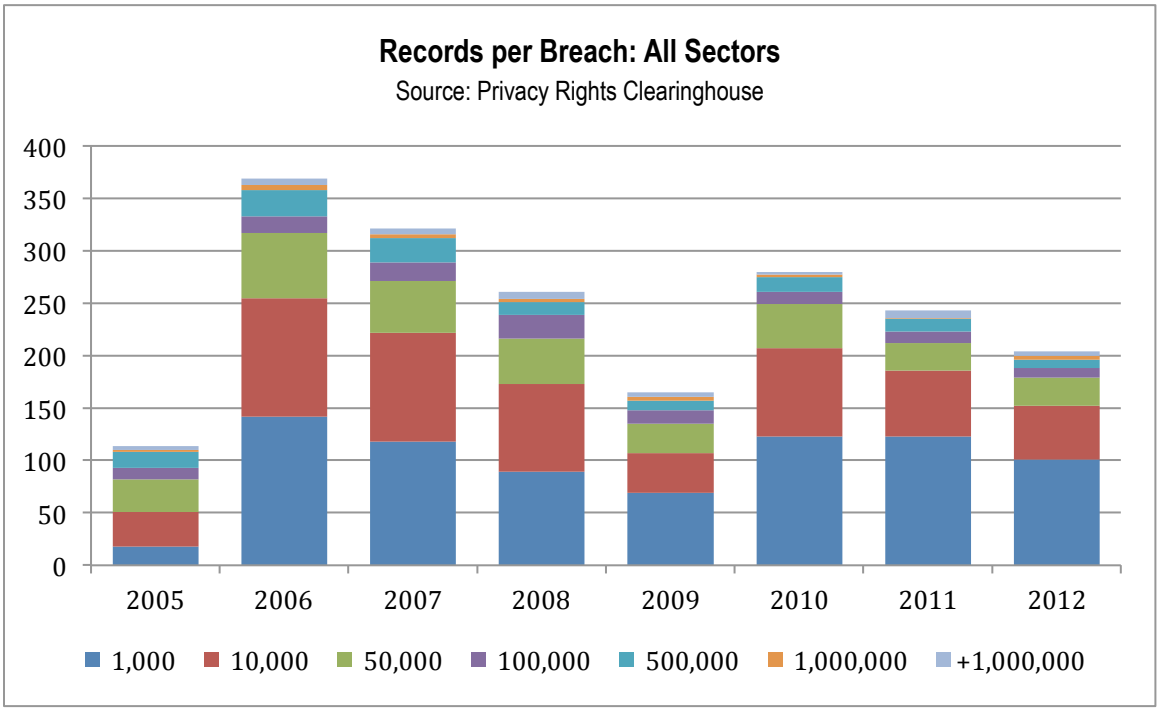


[7] The California disclosure law (discussed in Section IV) passed in 2003. The jump in breaches from 2005 to 2006 is more likely a consequence of more widespread reporting of data breach announcements and collection into the database than it is an actual dramatic jump.

[8] Open Society Foundation also uses this classification scheme.

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 about an individual. 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 breaches that reveal social-security numbers 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 that the most frequent breach is small, involving 1-10,000 records. There is 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. In particular, huge breaches (1,000,000 and up) though infrequent have not been controlled. This histogram of breaches offers an insight to the probability distribution of a breach event. A cross-tabulation of the type of breach with the size of the breach could help target investment in information security. However, not known is whether huge breaches are more likely to lead to information abuse, or whether data from small breaches are more likely to be mis-used.

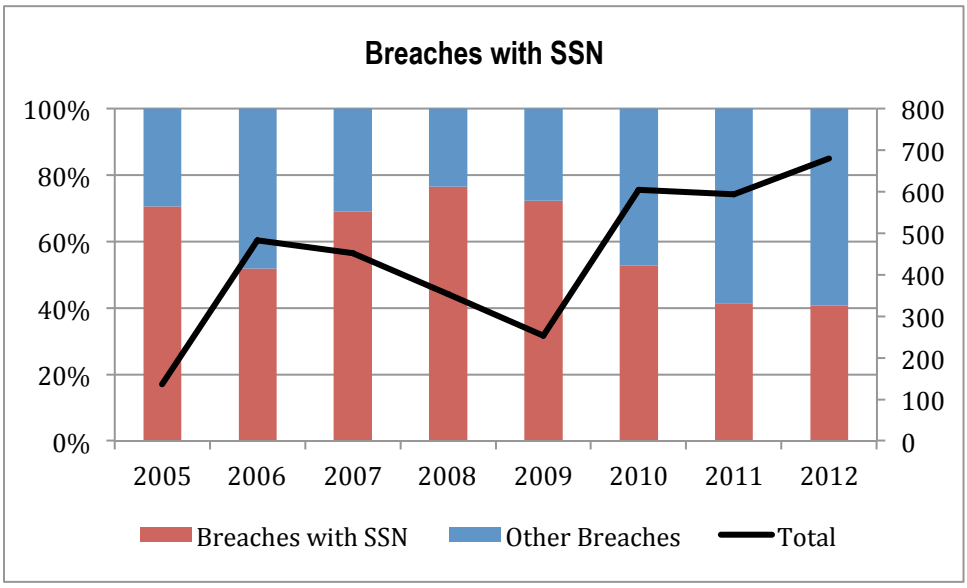


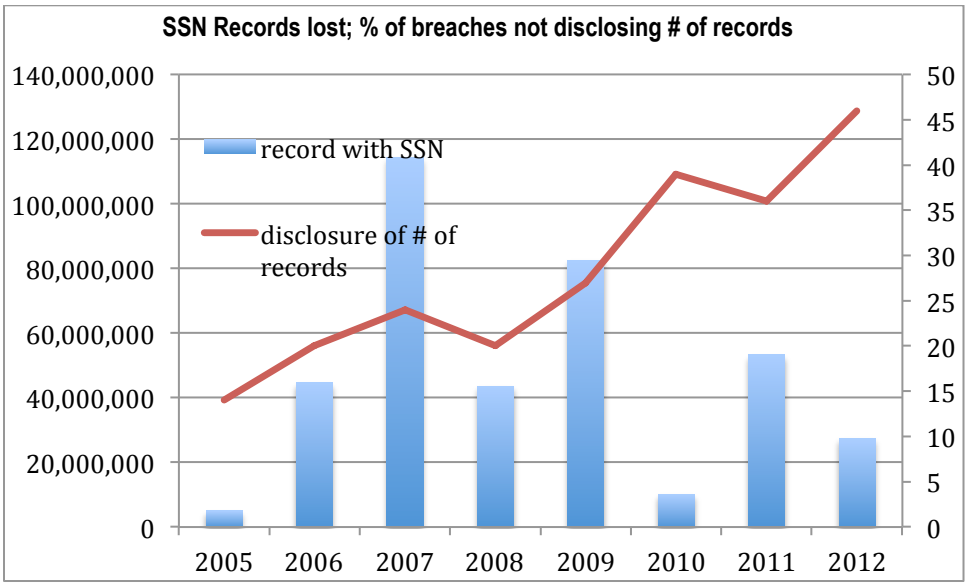
What kind of information is lost?

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 evidence in Section IV). Based on the PRC data, there is a mixed picture of whether more or less high-value information is

being lost. In part, this mixed picture appears to be because reporting of SSN losses in increasingly incomplete.

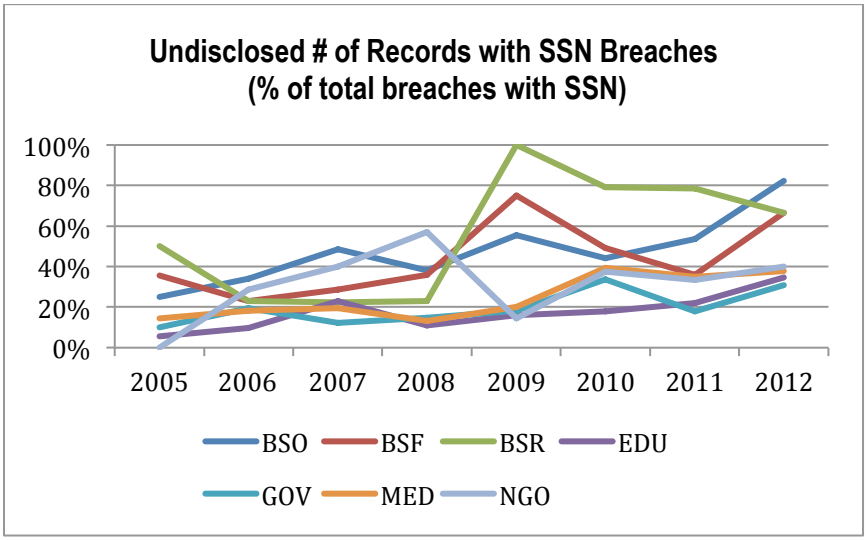
Over the time period, the number of breaches that reveal SNN has increased; but as a share of all data breaches, those that reveal SSN has declined. And, the number of reported records where the SSN was compromised declined from a peak in 2007, although not in trend fashion. So, this suggests that SSN breaches are becoming less prevalent, perhaps because of enhanced security.





On the other hand, recall that not all breach announcements reveal the number of records lost. Within breaches that compromise SSN, the share of those breaches that do not disclose the number of SSN-related records lost has increased over time. Considering a sectoral decomposition of data breach announcements, the 'Business-Other' (BSO)

category is the largest sector that does not disclose whether SSN records have been compromised. Sectors that perhaps are under greater scrutiny, such as Medical (MED), Financial (BSF), and Retail (BSR) appear to disclose more information.



In sum, interpreting the data on SSN breaches and required disclosure requires more analysis. Required disclosure may have led to security investment and thus fewer SSN-related breaches. Or required disclosure may just have prompted less transparency in public reporting.

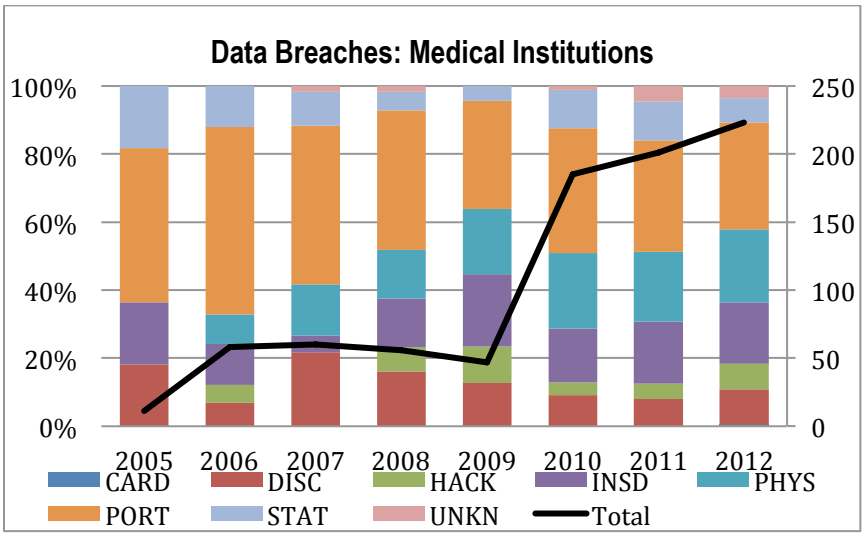
Is there differentiation by sector?

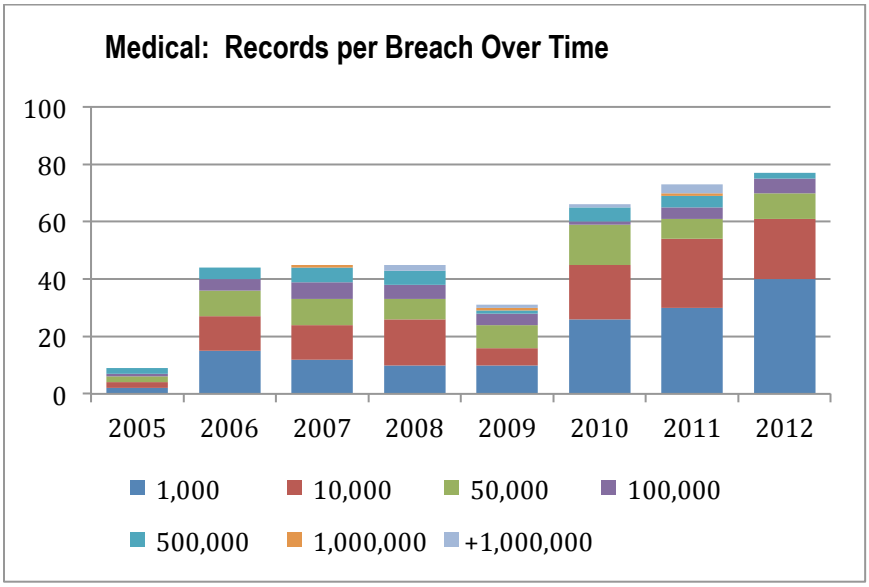
Looking behind the averages, are there differences by sector? Which sectors are the most prone to data breaches, by what means, and does the size of breach and information revealed differ by sector? The PRC data can be aggregated into business sectors (finance, retail, medical, other), government, education, and NGO.[9]

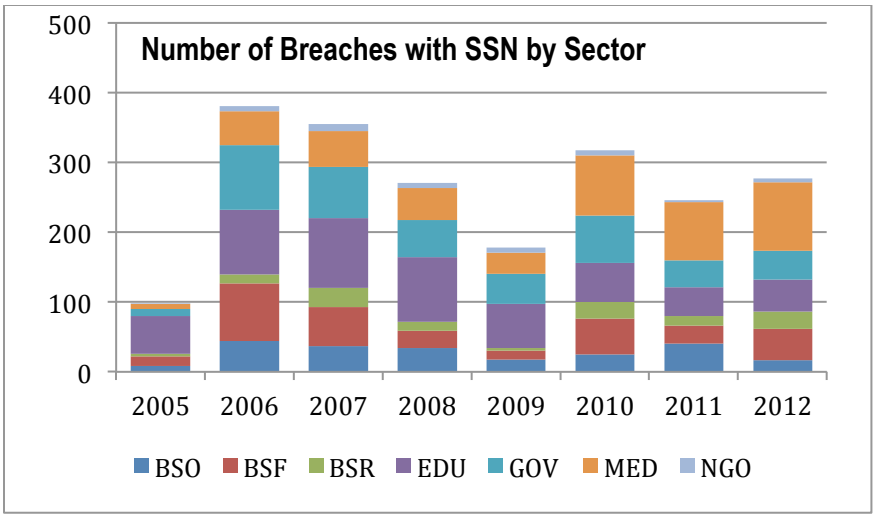
Data breaches in the medical sector are about double any other sector, with a huge increase in the last couple of years. This could be a fact, a function of disclosure, or a function of disclosure and reporting. In contrast to the aggregated data, the main source of data breach in the Medical sector is lost paper documents and lost laptops. But insider fraud has a rising role. (Recall that for the aggregated data, outsider 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.

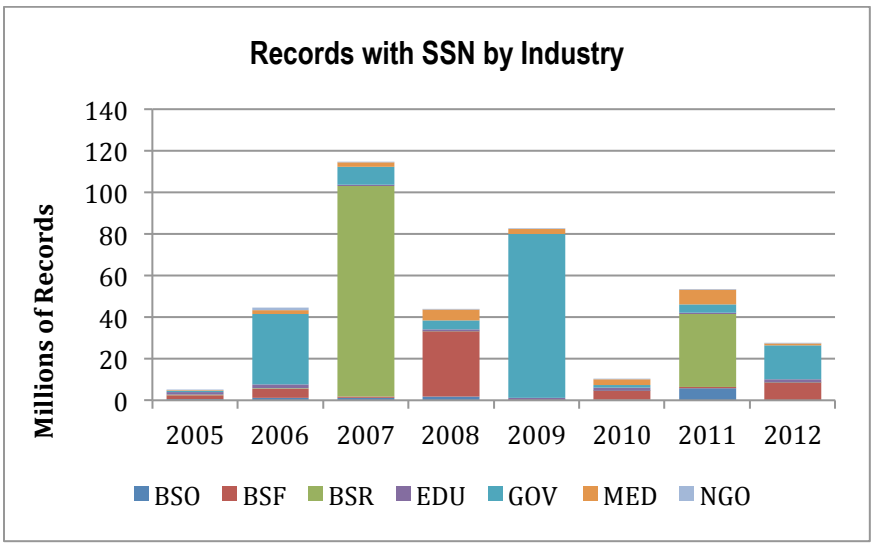
[9] More granular data, including firm identifiers, can be obtained directly from the PRC website. The Open Security Foundation did have a public on-line database (until 2007, see it used in reference Karagodsky and Mann, 2011), but it now is behind a permission wall. Efforts to obtain access were not successful. These two sources both draw from public announcements of data breaches. Cursory analysis comparing the two databases for overlapping years shows similarity, but they are not identical.

However, when the number of records lost with SSN is considered relative to other sectors, medical is not the largest problem sector.

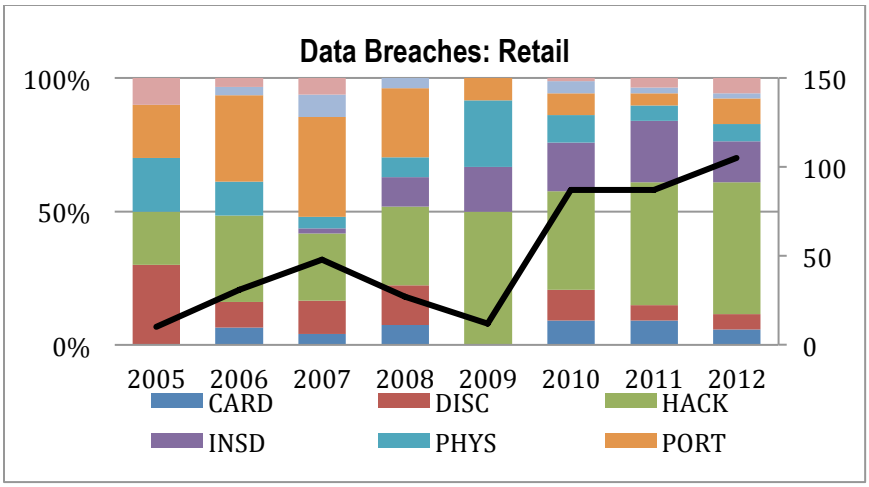


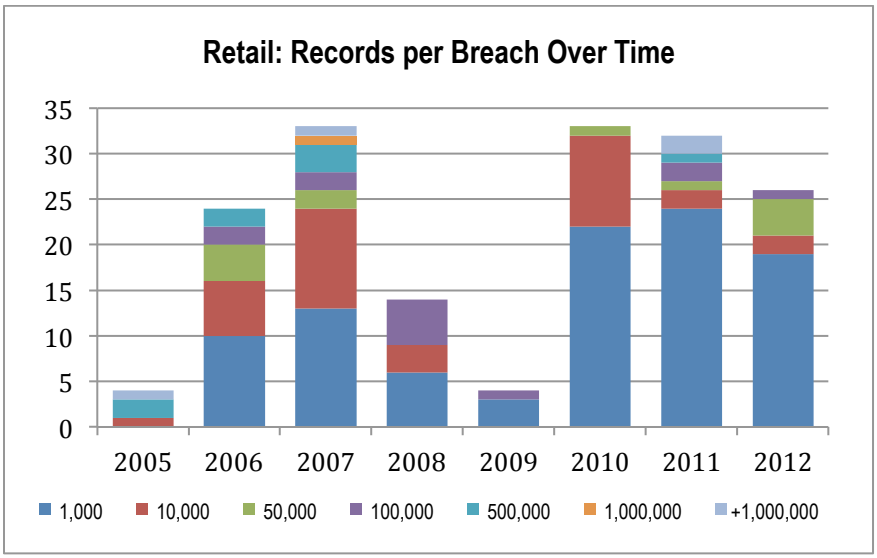




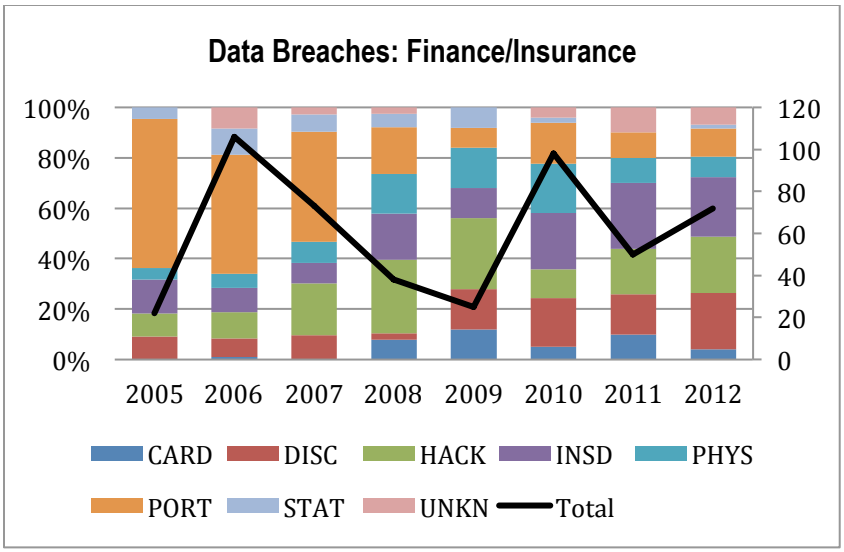


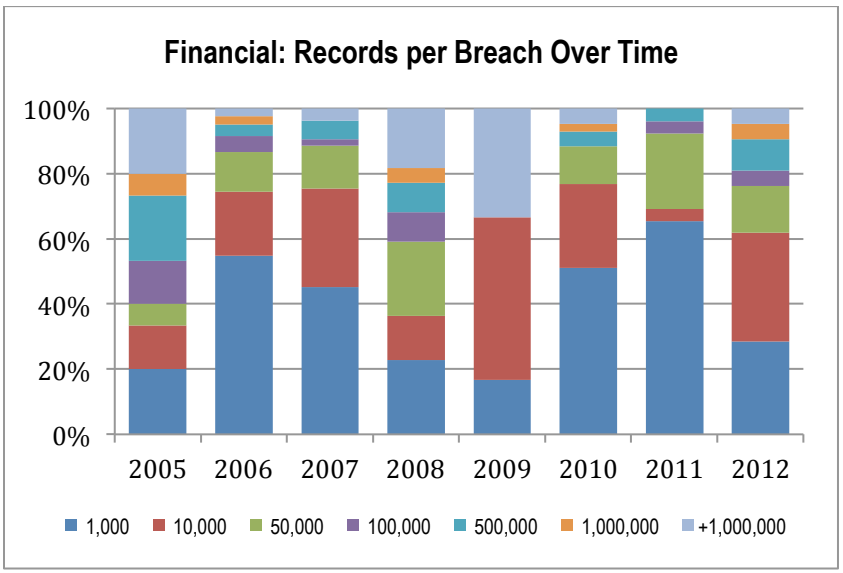
The chart above on records lost that compromise SSN reveals that retail is another sector that has a lot of data breaches. As shown below, the vast majority of data breaches in retail are by hackers. The number of records lost per breach is very small, generally, and the number of breaches that reveal SSN is quite small generally. But, when the 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 may warrant further analysis.



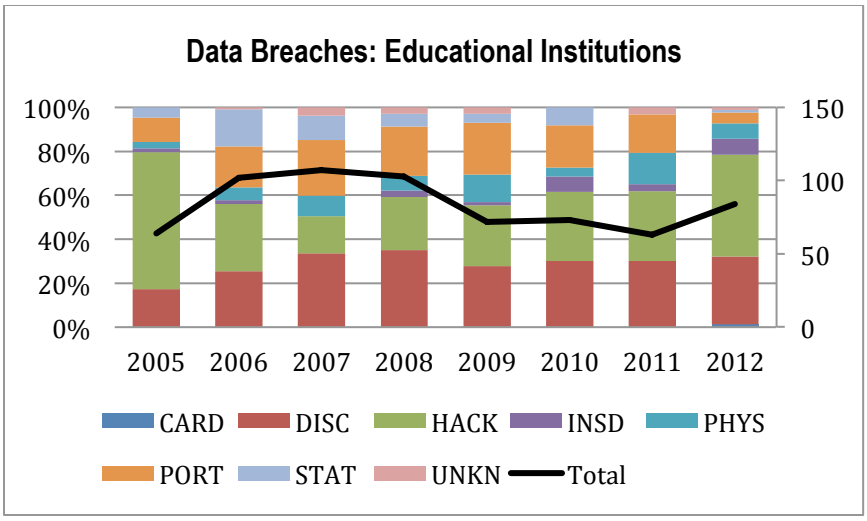


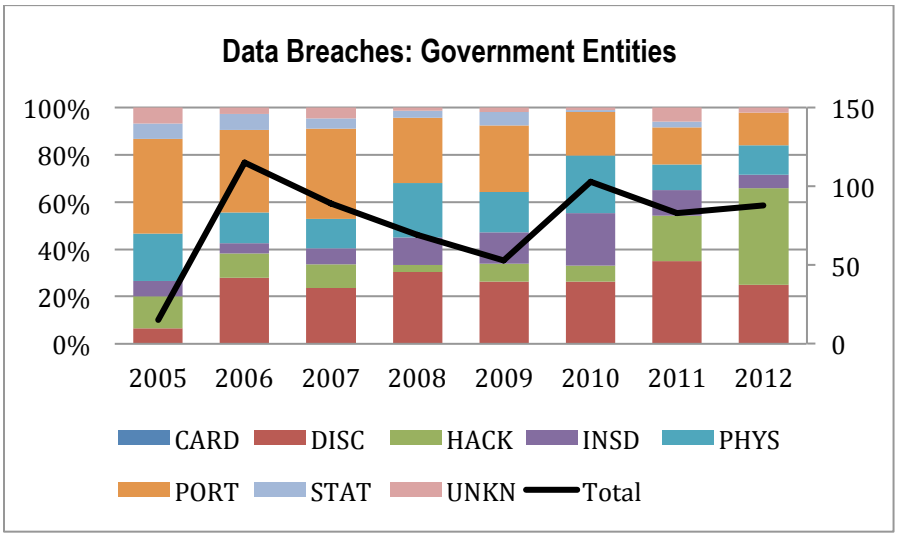
A third sector of particular interest is financial and insurance institutions. The number of data breaches appears to be under control. However the origin of the breach through insiders is significantly greater share than other sectors, and both hackers and unintended disclosures also are large. Very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

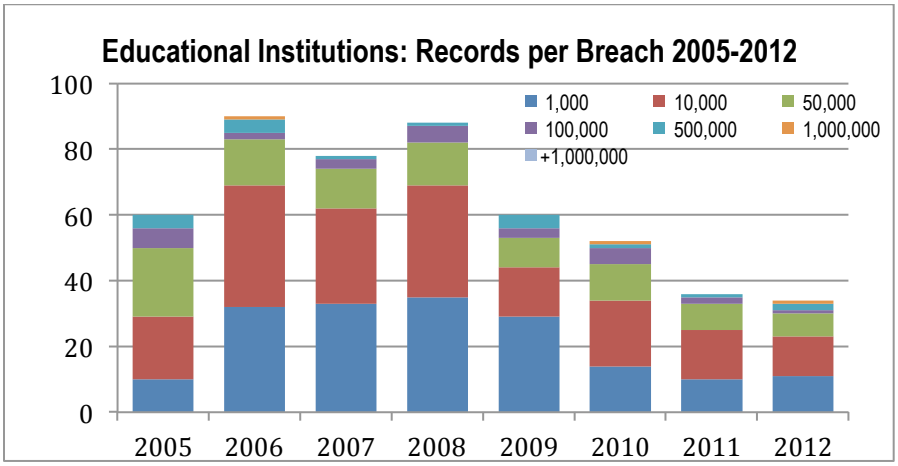


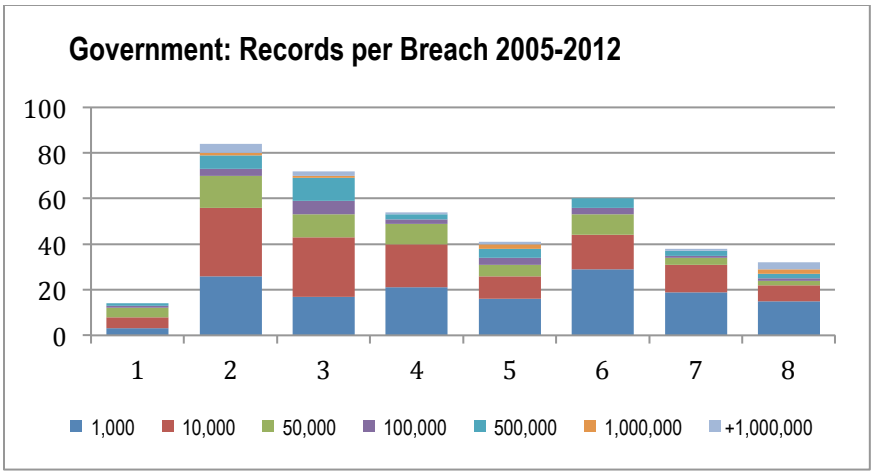


Government and educational institutions lose data both from hacking and from unintended disclosure. The bulk of the losses in the education sector are small, but the government has experienced some very large losses, and with a large number of records containing the SSN.


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


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


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



In sum, the sectoral decomposition of the data suggest that a one-size-fits-all approach to evaluating the costs of data breaches or the approach to data security is not appropriate. The sectors differ in terms of how data are lost and which size breach is most prevalent.

Cross-border data 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. First, only the U.S. has, since 2003, required public announcement. So a time series of public and reported disclosure is, at present, only available for U.S. firms. When U.S. firms incur a data breach, further information about the cross-border incident is obtained through firm survey by consultancies such as Verizon or Ponemon/Symatec. Data on consumer's information exposed is self-reported to the US. Federal Trade Commission or by other survey. In short, data on international breaches is spotty and incomplete.

Verizon reports that about 20 percent 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 point-of-sale or other skimming-type devices (this is consistent with the prevalence in the PRC data on small breaches in the retail sector). Note however, the very large share of incidents where the origin of the data breach 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		

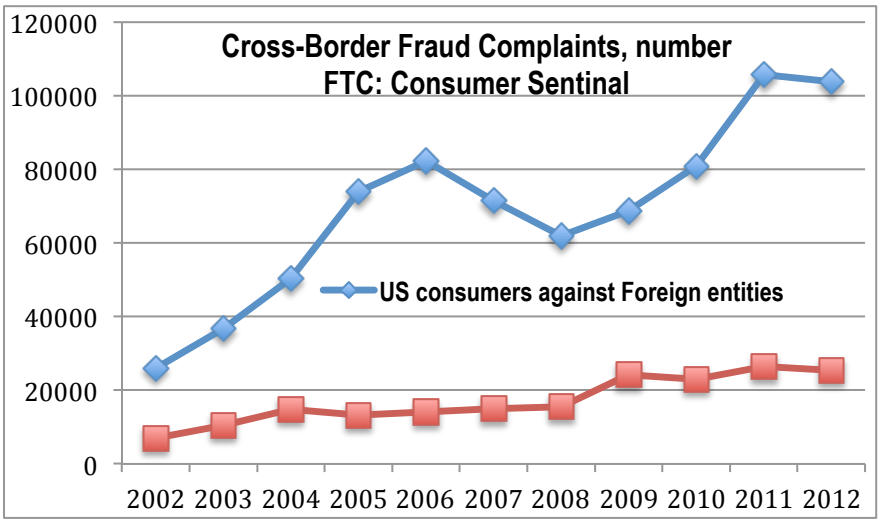
* external threats from China: 30% of threats; only 2% focused on financial (as opposed to industrial) information… the focus of this section of the paper is not on industrial espionage.
nr/ not reported

With regard to specific countries in CEE, Bulgaria and Romania especially (singled out in the 2012 Verizon report) joined the European Union in 2007. At that point, these countries 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 the EU has focused security strategies on data-in-transit, not data-at-rest. This is because the EU Directives stringently limit data retention. In contrast, the U.S. security emphasis is on data-at-rest. It is possible that the different security focus of U.S. vs. EU methods 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.) Sullivan (2010) and McCarthy (2010) discuss security and gaps in approach in the global payment card industry.

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 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 of security protection be on the numerous small breaches or the very few disastrous breaches? Finally, sectoral variation and variation in the size of the breach and in how information is lost may be relevant when considering the role for a domestic focus vs. a global emphasis on 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 balance the benefits of information aggregation against the costs of when information is lost. If market discipline is sufficiently robust, regulatory intervention or private legal action may not be necessary.

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) 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) 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.

Data on the costs of breaches is an integral part of the analysis—but how to measure costs, and costs born by whom? Should the focus be on costs to prevent or costs to remediate? 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 relative to the costs of enhanced security?

The Role for Disclosure

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, or to a government entity that can force remediation—is less clear. As noted in the framework, individuals are, by definition, atomistic and therefore lack market power to respond to disclosure and discipline whoever lost the data. Further, with cognitive bias and limited rationality, disclosure may not yield the right incentives to protect ones own information, nor improve the outcome if data are lost. Intermediaries are fewer in number, but this market position may reduce the incentive to prevent or remediate a data breach, disclosed or not. With layers of intermediaries, moral hazard is another issue as one intermediary may free-ride off the security approach of another. So, disclosure to a government entity that can press for remediation, and induce investment to avoid future losses might be socially optimal.

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 make a public announcement if the security of the information has been compromised. The legislation further stipulates that the organization responsible for the breach must notify each individual for whom it maintained information—this is a direct cost rather than simply an indirect cost of, say, loss of reputation. 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.[10]

The U.S. approach of such broad-based disclosure to individuals is unique around the world. [11] 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 incidents that must be disclosed has been narrowed on account of 'excessive' information flooding the agency. Australia is considering whether a specific disclosure law is necessary or whether existing law addresses disclosure. The European Union's approach here-to-fore under the 2003 Directive on Privacy and Electronic Communications has not required disclosure of information loss. But the promulgated and evolving General Data Protection Directive will require disclosure of material breaches to a supra-governmental unit. (European Commission, 2012.)

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 U.S. breach 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 U.S.-style disclosure probably is too costly relative to the gains.

There is evidence that consumer limited rationality is a problem. 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%). But, about 60% thought that the communication informing them of the breach was 'junk mail.' So, data security has salience, but people don't necessarily respond to disclosure as expected. Retzer (2008) considering broad-based disclosure rules desensitizes the recipient to the announcement, which works counter to the role that disclosure should play as a disciplining device.

Even so, firms need to be aware that loss-of-trust matters. Nearly 90% in the Ponemon study said they had or might discontinue their relationship with the firm over a data breach. Consumers have reason to punish firms that lose their data. Around one-quarter of consumers who received a data-breach notification experienced identity theft.

---

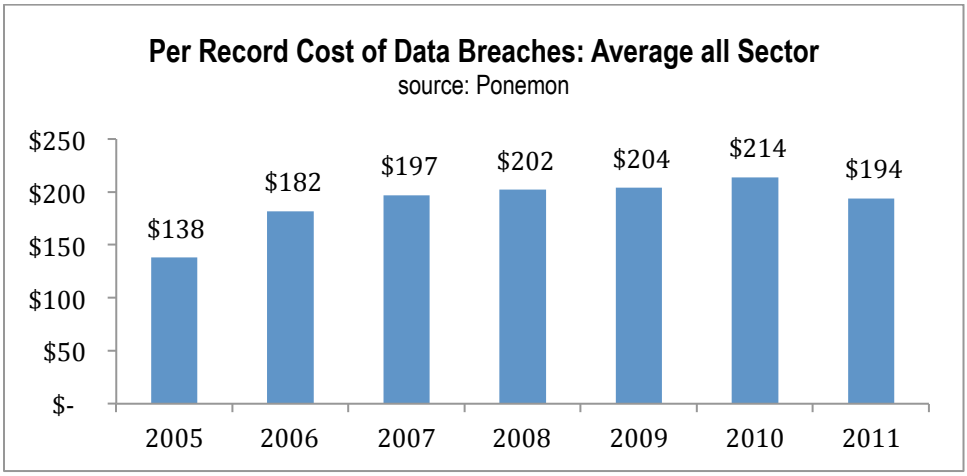
[10] There is federal legislation protecting children (COPPA), health (HIPPA), and financial data (Grahm-Leach-Bliley Act) but generalized 'personal information' is not protected. At the state level there is a patchwork of legislation that protects some information in some states – for example, databases 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).

[11] See Global Privacy Alliance (2009) for a comprehensive review.

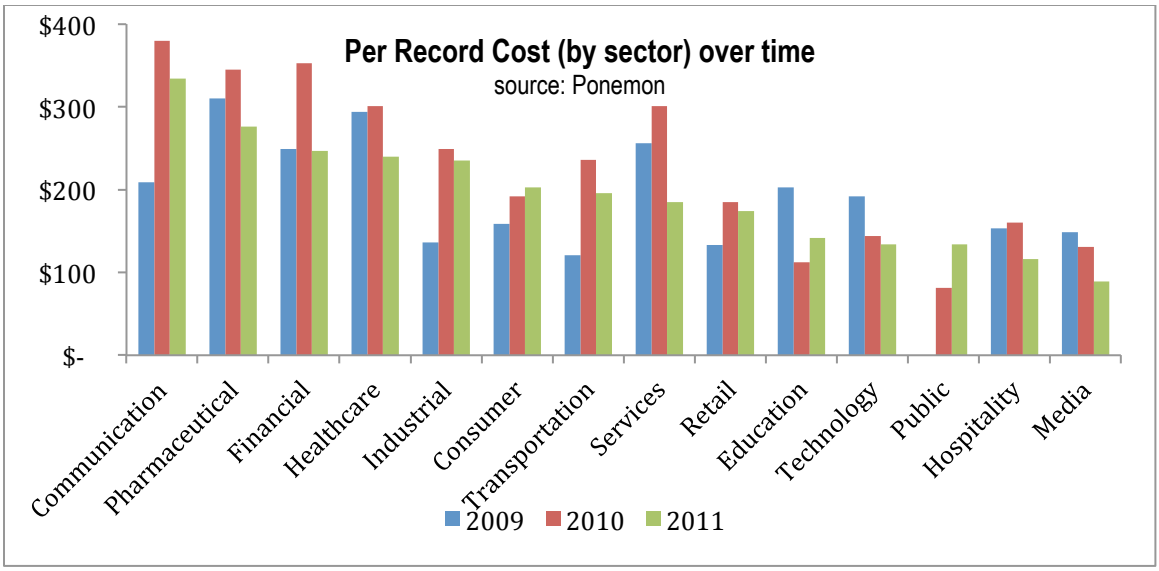
Consumers whose SSN were compromised were 5 times more likely to experience identity theft. The average out of pocket consumer costs of a data breach ranged from $400 to $700 (2005- 2012, Javelin Strategy Research, 2012). This is substantially higher than the per-record costs that businesses bear, as will be discussed below.

Trends in Business Costs: Lost Business and Remediation

If remediating after information loss is sufficiently directly costly to a business, then presumably that business will undertake action to improve information security. 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). [12] There is substantial variation across sectors in the cost per record loss. Financial and healthcare were two sectors that had a high prevalence of fraud and a high density of SSN losses. The higher costs of remediation could be due to these factors.



[12] 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/



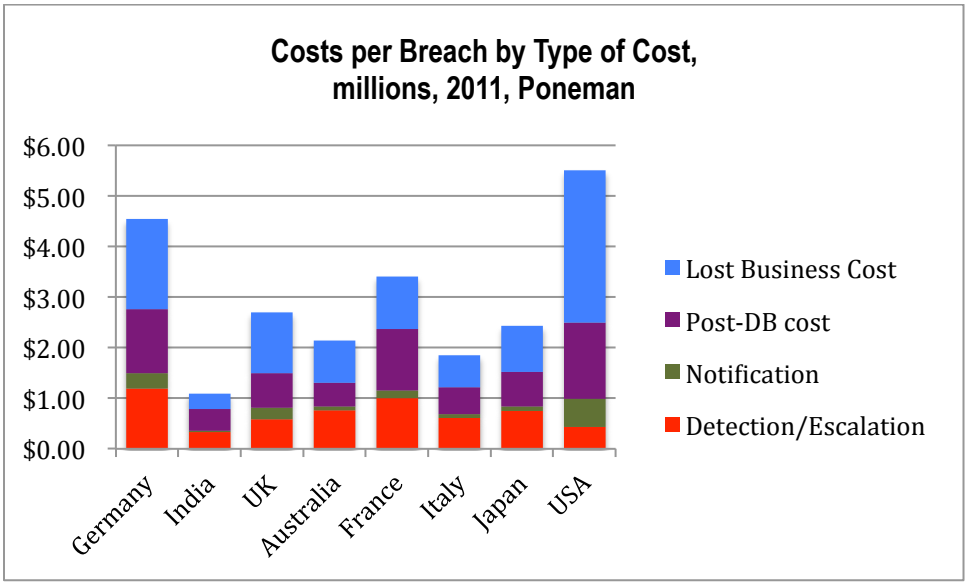
Comparing the business costs of losing data across countries reveals some interesting observations. There is substantial variation in costs per record lost, and also substantial variation in the components of the costs per record lost as disaggregated into detection, notification, post-breach costs, and lost business costs. However, 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 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 significant variation in the cost per record lost. These differences in cost, despite the same regulatory environment, could be due to different sectoral distribution of firms, or could reflect heterogeneity in consumer attitudes and response to data breaches.

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	

Next, 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., which may also be consistent with the U.S. disclosure law. 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 made detection by U.S. firms more a matter of course (and therefore less expensive) rather than a special one-off event.



Market Value of Stolen Information

Another approach to measuring the value of information lost through a data breach is to go directly to the market. How much is stolen data worth? Getting this information in a systematic way is very difficult. Snapshots of value in the market show very large ranges of value of stolen data and it is hard to map the valuation of stolen data to the sectors that have experienced data breaches. As a point of comparison, however, the value of 'bank account credentials' at anywhere from $30 to $850 could be compared to the business cost-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 business cost-per-record lost in retail is $174. If we consider just the business cost, perhaps the most efficient response is to just pay the thieves. But, recall that the consumer cost was multiples of the business cost.

Value of Data	**2007 (avg.1H, H2)**		**2008**		**2009**	
	Low	High	Low	High	Low	High
Credit Card Information	$0.45	$12.50	$0.60	$30.00	$0.85	$30.00
Bank Account Credentials	$20.00	$725.00	$10.00	$1,000.00	$15.00	$850.00
Full Identities	$6.50	$15.00	$0.70	$60.00	$0.70	$20.00
Email Accounts	$2. 50	$35.00	$0.10	$100.00	$1.00	$20.00

Discipline by Equity Market

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 (CAR)—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 as a strategy for aligning private incentives when it comes to protecting information against loss.

The predominant conclusion is that there is a negative, short term, statistically significant effect of a breach disclosure announcement on the equity market price of the announcing firm. The conclusion appears only when SSN are 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 (2011), suggests that relatively larger CAR losses are associated with data loss by

banks and by healthcare firms, when SSN are lost, and when data are lost stolen through an intrusion by a hacker. This is consistent with the higher per-record costs for these sectors as noted earlier.

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. [13]

Policy intervention: Standardization Amid Globalization

A back-of-the envelope calculation suggests that the macroeconomic business cost of information lost in the U.S. in 2011 was $40 billon, and triple that if consumer costs are added.[14] Is this large or small in a $14 trillion economy? Seems small. But, considering just the business costs, it is relatively larger when the reference is net business investment in equipment and software of $107 billion, (BEA NIPA 5.2.5). 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.

Policy intervention has many possible faces: standardized regulations and enforcement through fines. A key question is, who should be the policy target? The numerous consumers and final using firms, or the few intermediaries (transmitters and/or aggregators) of information? The market-power analysis already suggests and data confirm that cost to remediate a data breach per consumer exceeds most per business-related costs.

Oussayef (2008) and Orr (2012) suggest that the focus should be on the consumer – the originator but also ultimate user of information value, and the point of greatest cost of loss according to the data. A response to the consumer's limited rationality would be to

[13] Research on security costs include: Brecht and Nowey, 2102, Demetz and Balchnecher, 2012.

[14] $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). Assuming consumer costs of $400 (low end of the Javelin estimate) triples the total.

standardize communications with them, for example, standardize privacy policies. But, generally this is not the direction that regulation or the market is going.

Another approach would be to standardize regulations for the relatively fewer intermediaries. However, countries differ in the focus for security (data-in-transit vs. data-at-rest) as noted earlier. How do firms that operate in both the U.S. and the EU address this problem? 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, and submit an enforcement policy to the U.S. Department of Commerce. Fundamentally, the Safe Harbor remains a self-regulatory mechanism that rests uncomfortably against the mandate-oriented approach of the EU Directives. [15]

Within the EU, the issue of standardization is emerging with the January 2012 EU General Data Protection Regulation. This regulation addresses 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 (in Germany for example) and others to be tightened. It extends these rules to all foreign companies processing data for EU citizens and EU firms could not transmit data to countries with insufficient protection.[16] Disclosure of data breaches would be within 24 hours, which is quite a switch from no disclosure now. Fines for data breaches could be up to 2% of global revenue, which is potentially huge. Disclosure, assigning blame, and enforcement are all issues.

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 the location of their information. Third party breaches can yield large and broad based thefts. [17]

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 for Spokeo under

[15] For more on market vs. mandate approach to international data, see Mann (2001) and Mann and Orejas, 2003.
[16] At present 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.
[17] NYTimes: Cyberthieves Looted A.T.M.'s of $45 Million in Just Hours http://nyti.ms/ZKTw5H.)

the Fair Credit Report Law [18]. But these fines don't always work to change behavior. (Wyndham has been fined three times)[19]. Another strategy is the mandated audit: Many years and a big price-tag could change the balance between which is more costly, to protect data or to experience a breach and incur both immediate and long-lived audit costs. A caveat to the firm decision is that the advocacy of the FTC has a political lifespan.

Finally, the issue of cross-border regulation of information 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 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 and regulatory guidelines.

Legal recourse: Evolving Notion of 'Standing'

The role of the legal profession is evolving, and may play a more important role in firms undertaking appropriate security. (See Fryer, Moore, Chown (2013) for an extensive review of theory and practice.) Initially, and still true in general today, courts have 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 (although this is changing, as per discussion above). Simply losing data (as in losing property) is not sufficient grounds for a case, 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, as per data above). Without threat of legal action, there is reduced incentive for firms to improve data protection.[20]

However, research suggests that the legal approach may be beginning to have traction as a disciplining device in the cases of breaches of financial and medical information.

---

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

[19] "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/

[20] 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)

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 and if a class action lawsuit is a threat.

More generally, the risk of class action suits (or 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 percent of costs (2006) to 15 percent 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 against them. But these potential legal costs also cause firms to push-back against disclosure, particularly of the magnitude and sensitivity of information lost.

Increased Information Security: The Costs

Given the costs of a data breach, it is not surprising that firms are investigating whether the benefits of investing in information security are greater than the costs of incurring a data breach. A full review of this literature is beyond the scope of this paper. There are theoretical articles and practitioner analysis, and many firms whose business it is to sell information security solutions. [21] Selected research includes Aurora, Hall, Pinto, Ramsey, Telang, 2004; Gordon and Loeb, 2006; Bojanc and Jerman-Blazic, 2007; Carty, Pimont, Schmid, 2012. It is not just the cost of investment but also the organizational behavioral considerations of information security management (Kwon and Johnson, 2012).

Once information security is institutionalized, will investment costs decline? An important consideration is whether the investment in information security is against the 'normal' intrusion or the 'black swan' event. Do security investments differ for these two types of event? Is there a ratcheting-up of security, even if the threat remains the same?

From an organization behavior standpoint, the security professional will want to ensure that the rare event never happens on their watch. But if the way in which a rare event takes place changes with the technology frontier, it may imply an ever increasing security budget. Neuhaus and Plattner (2012) address this issue in the narrow context of software security patches. But the principle needs broader recognition.

To get at the general issue of data-breach hype, a Google Alert 'data breach' was employed for 300 days. There was at least one article every day. On average, there were 1-2 articles in the business press, with a single spike over the 300 days at 5 articles. The general press, averaged about 2 articles per day, with several days spiking to 7 articles. In media directed to security professionals, the average daily article count was 3 and on more than 25 days the number of articles was 6, 7, or 8 articles. There is a lot of

communication among security professionals. Is this about best practices (which could reduce incidence of data breaches) or about threats (which could increase investment in security)?

V: Considerations for the Digital Agenda

Conceptual Framework

There are three characteristics of the information marketplace that pose challenges to the pricing of information and the balancing of costs and benefits of securing data: (1) Information economies of scale and scope. (2) Differential market power among participants. (3) Unknown probability distribution of data breaches and of abuse of information.

The conceptual challenge is to put all three elements into a model. Probably the most difficult is to combine the economies of scale/scope and the probability function because the cost-benefit calculation regarding today's data breach is a function of all future possible data breaches. Numerical simulation models are one possible approach but these need parameterization using data.

Data needs and analysis

An increasingly amount of data is becoming available on the costs of data breaches to businesses and consumers. More analysis of these data as to the relationships between sectors, types of data lost, various costs associated with data loss, nationality of perpetrator and victim of data loss is beginning to become possible. Research will be fruitful, but depends on access to a wide range of what often is proprietary data. Ensuring wider availability would enhance research.

However, key variables are not known. What is the probability of a data breach? The 'dog that didn't bark' (that is, transactions or databases that were not compromised) would seem to be greater than those that are, but we don't know. Getting a handle on the probability of a data breach by sector, size of firm, national jurisdiction, is key for starting to evaluate and balance the costs of protecting data against the costs of intrusion. In addition, understanding the relationship between data lost and data abused is important. Most data appear to be lost 'the old fashioned way,' not stolen. Is the likelihood of abuse low or even zero for that kind of data breach? In this regard, the large but infrequent malicious data breaches would seem to be more costly than the small, but frequent breaches, but more analysis of data breaches by size, sector, and type of breach is needed.

Cost of investing in data security and costs of limiting data aggregation are not widely available. One area for research is the intersection of the investment strategy with the probability of a data breach. Investing against the infrequent but disastrous event may be

inefficient, particularly in a rapidly changing technological environment. Similarly, evaluating the cost of limiting data aggregation as lost benefits also depends on the probabilities.

International jurisdiction

The information marketplace and many firms are global, but policy jurisdictions and consumers are still local. Countries differ in their emphasis for information security, data-in-transit vs. data-at-rest and these differences can create exploitable gaps for hackers. Considering ways to reduce international arbitrage offers an area for research in the political, economics, and technological spheres.

Countries also differ in their approach to information aggregation, protection, and breach disclosure. These differences are not likely to be overcome, since deep-seated cultural values play some role. More work is needed on the value of information, the value of protection, and the costs of a breach both within and across international borders.

References

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.

Anderson, Ross and Chris Barton, Rainer Bohme, Richard Clayton, Michel J.G. van Eeten, Michael Levi, Tyler Moore, Stefan Savage (2012) "Measuring the Cost of Cybercrime," WEIS 2012.

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

Ashish Arora, Christopher M. Forman, Anand Nandkumar and Rahul Telang (date), "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

August, Terrence and Tunay I. Tunca (2011) "Who Should be Responsible for Software Security? A Comparative Analysis of Liability Policies in Network Environments, WEIS March 2011.

Demetz, Lukas and Daniel Bachlechner (2012) "To Invest or Not To Invest: Assessing the economic viability of a policy and security configuration management tool" WEIS 2012.

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

Brecht, Matthias and Thomas Nowey (2012) "A Closer Look at Information Security Costs" WEIS 2012.

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*

European Commission (2012) "Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

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

Fryer, Huw, Roksana Moore, and Tim Chown (2013) "On the Viability of Using Liability to Incentivise Internet Security", WEIS 2013.

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.

Global Privacy Alliance (2009) BREACH NOTIFICATION LEGISLATION KEY ELEMENTS TO CONSIDER, August.

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.

Ioannidis, Christos, David Pym, and Julian Williams (2012) "Sustainability in Information Stewardship: Time Preferences, Externalities, and Social Co-Ordination" WEIS2012.

Javelin Strategy and Research (2013). "Data Breaches Lead to Identity Fraud".

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?

Kwon, Juhee and M. Eric Johnson (2012) "Security Resources, Capabilties and Cultural Values: Links to Security Performance and Compliance," WEIS2012.

MacCarthy, Mark (2010 "Information Security Policy in the U.S. Retail Payments Industry" WEIS 2010

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.

Neuhaus, Stephan and Bernard Plattner (2012) "Software Security Economics: Theory and Practice," WEIS 2012.

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.

Sullivan, Richard J. (2010) The Changing Nature of the U.S. Card Payment Fraud: Issues for Industry and Public Policy,” WEIS2010

Thomas, Russell Cameron, Marcin Antkiewicz, Patrick Florer, Suzanne Widup, Matthew Woodyard (2013) “How Bad Is It? A Branching Activity Model to Estimate the Impact of Information Security Breaches,” WEIS 2013

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

Black swan events were introduced by Nassim Nicholas Taleb in his 2001 book *Fooled By Randomness*, which concerned financial events. His 2007 book *The Black Swan*

Tang, Zhulei; Yu (Jeffrey) Hu, and Michel D. Smith (2007) "Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor," *Journal of Management Information Systems* vol 24:4 p 153-173.

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**