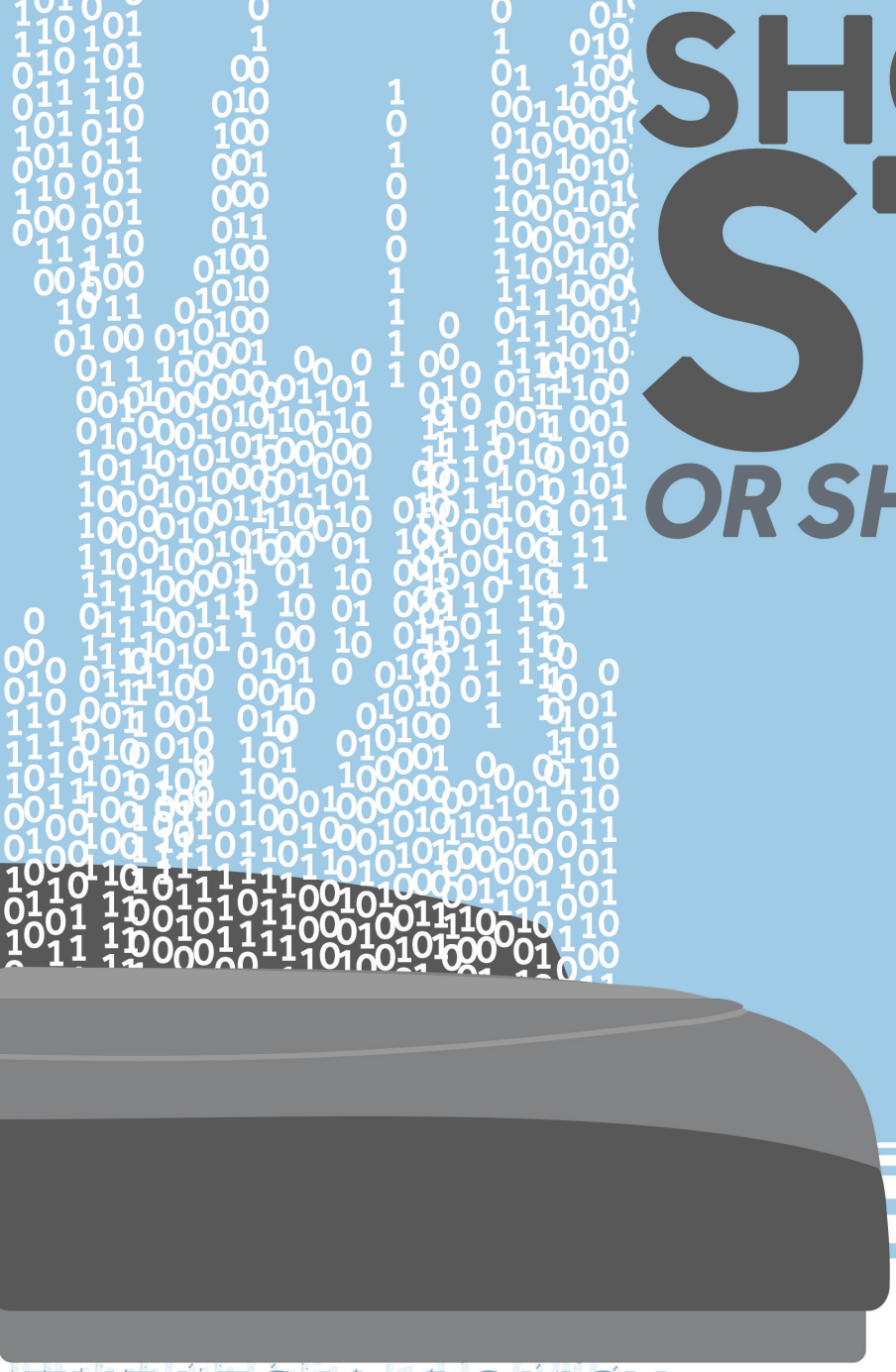
SHOULD IT STAY, OR SHOULD IT GO?



THE LEGAL, POLICY AND TECHNICAL LANDSCAPE AROUND DATA DELETION

cdt

February 2017

By Michelle De Mooy, Joseph Jerome and Vijay Kasschau
Center for Democracy & Technology

Published February 2017

Introduction

Delete is a word built into the vocabulary of users from the beginning of personal computing. When commanded to "del," an operating system appeared to erase a file completely. However, right from the start, a user's commonsense understanding of the command to "delete" differed from companies' practices; rather than erasing a file, "delete" meant "put in the recycle bin." "Deleted" files were not really gone but rather out of sight, available to be recovered if necessary. The rise of cloud computing, where files live remotely from their owners' devices and are frequently accessible from multiple devices, has further muddied the concept of deletion by saving all of a user's files in an ambiguous location until called upon. While this change may seem trivial, it represents a larger truth about our digital files: they are almost never really "deleted."

As our lives migrate more and more to digital platforms, companies collect and retain more data about us. Correspondingly, the range of practices for destroying that data has also grown. The technical challenges of deleting data or completely disposing of information, coupled with a permissive legal environment and economic incentives, have created an environment in which many companies retain information by default. But this practice underestimates the serious risks posed to companies and users when there is no plan to destroy data.

In this paper, we argue that companies should reconsider their concept of deletion and implement sound technical and policy processes to formalize their practices. We believe there comes a point when the value of data has been extracted and the costs (both operational costs and potential for liability) of retaining data outweigh the potential benefits of keeping it. While the price of physical storage may be plummeting, data management costs continue to grow. Data breaches are ubiquitous and massive,[1] and show no sign of abating. Retaining large amounts of data greatly increases the potential harms that could result from a data breach; the more robust a database is, the more appealing it is to malicious actors. Headline-grabbing breaches of major retailers, financial institutions, healthcare providers, and even government agencies have damaged companies' reputations, exposed individuals to identify theft and embarrassment, and undermined trust in both institutional and organizational security efforts.

The threat of this information being used for government surveillance purposes also hangs heavy over many entities that wish to that wish to be helpful to the government while also protecting user data. Concerns about government access to private information were reignited by the 2016 election, during which Donald Trump campaigned on creating a Muslim registry derived from existing digital databases. Since that election, municipalities like New York City have been confounded in their attempts to delete an immigration database, over which the city now fears losing control. [2] Advocating for data deletion, privacy groups have taken to reminding companies that they cannot be made to surrender information they do not have.[3] A group called NeverAgain.tech, made up of engineers, designers, and tech executives at high-profile companies like Google, Twitter, and IBM, pledged to delete datasets and backups to thwart efforts to build the registry.[4]

Finally, lengthy data retention defies public expectations. A Pew survey found that people believe that companies should place limits on how long records about digital activities are stored.[5] This includes not just data collection by online advertisers, but also social media companies, search engines, credit card companies, and utilities; twenty-seven percent surveyed believe it is reasonable that even companies and retailers they do direct business with should not retain any of their personal information. Pew also found in a recent survey that 64% of Americans have personally experienced a major data breach.[6]

For all industries, removing data, rather than indefinitely retaining it, is an undervalued and underutilized component of business data management that can reduce risk and liability and improve efficiency. Even guidance documents that state regulators reference as a baseline for reasonable data practices, such as the CIS Critical Security Controls, often gloss over data deletion and destruction as information management tools.[7]

The data life cycle has lengthened, and at the same time, guidance on data deletion best practices has been neglected. This paper will scope some of the business risks that can emerge from mass data retention; detail the practical, technical, and legal issues around data retention and disposal; and offer a set of policy and technical recommendations for promoting the destruction of information within a responsible data management program.

I. The Incentives for Retaining Data

A business's decision to retain or delete any given data is informed by a number of considerations that are specific to a company's business model and how it plans to achieve those objectives. This decision calculus is never the same for any two companies, as internal data uses, policies, procedures, and related business goals differ between companies.

A permissive legal environment

Legal and regulatory requirements shape corporate retention and deletion decisions. Rather than explicitly requiring organizations to dispose of old information, many laws and regulations instead often require that certain data be retained for a specific period of time. Frequently, records retention in and of itself has become an important, substantive component of regulations, and retaining records is often essential to establishing compliance with legal and regulatory requirements.[8] One example of this is the Sarbanes-Oxley Act, which requires accountants to retain audit information of publicly-traded companies for five years from the end of a fiscal period in which an audit took place.[9] However, this law is specific to audit records, and it is inapplicable to sales records, collected customer information, and internal emails. This makes proper data categorization for most companies essential.

Alternatively, the enactment of data disposal rules such as the Fair and Accurate Credit Transactions Act's Disposal Rule provides some impetus for companies to inventory and categorize information in the event they consider destroying any of their data holdings.[10] These laws generally direct entities to implement a set of reasonable practices, generally based on the sensitivity of information to be disposed of. However, it is important to note that they provide no clear mandate to delete data. Absent such a mandate, and with ambiguous requirements to retain different types of records, it's not hard to understand why companies end up defaulting to saving rather than destroying data.

The difficulty companies face in determining whether to retain or destroy data is exacerbated in the U.S. by the wide array of sectoral laws and regulations that can apply to different categories of information. Rather than one baseline law that applies to technologies and data-gathering practices generally, U.S. privacy and data security laws instead address the collection and use of information in certain specific contexts. Because sectoral laws impose different protections on different types of information, they can also impose different retention periods or recommend different disposal options. Again, in the face of legal ambiguity, the natural incentive is to retain rather than to destroy.

New sources of health information demonstrate some of the legal complexities in this sectoral environment; identifying appropriate retention and deletion requirements is particularly challenging

when information is generated from different devices, locations, and contexts. The proliferation of wearable devices, employee wellness programs, and health trackers and apps is expanding the number of companies that come into contact with health information, but only some of these entities are likely to be covered by the Health Insurance Portability and Accountability Act (HIPAA). HIPAA's Privacy and Security Rules provide broad protections to "protected health information," or PHI,[11] but neither rule requires any particular disposal method of PHI for covered entities.[12] Instead, companies are asked to "review their own circumstances to determine what steps are reasonable to safeguard PHI through disposal, and develop and implement policies and procedures to carry out those steps."[13] The law suggests that companies assess potential risks to individual privacy and the types and amount of PHI to be destroyed.

Though at least thirty-one states and Puerto Rico have enacted laws that require entities to destroy, dispose, or otherwise make personal information unreadable or undecipherable,[14] the parameters for when data should be destroyed are left to a company's discretion. Oregon's data disposal law, for example, obliges companies to dispose of information after it is "no longer needed for business purposes" or when "the business decides it will no longer maintain the records."[15] Forty-seven states have also enacted legislation requiring private, governmental or educational entities to notify individuals of security breaches of information involving personally identifiable information.[16] Though the contours and precise requirements of these laws vary from state-to-state, the generally open-ended language in most states places the ultimate decision to destroy information in the hands of a company's data management program.

More broadly, while privacy and consumer protection regulators continue to stress the importance of collection limitation in principle,[17] even the White House argued that governmental policies focused on limiting data collection are neither "broadly applicable" nor a "scalable strategy."[18] Self-regulatory efforts in the online advertising space have largely focused on restricting how data is used rather than on limiting initial collection of information or having requirements for retention or disposal of data.[19]

Against this legal backdrop, companies are incentivized to embrace data collection and retention as a default practice -- and this is amplified by the potential economic value of big data.

The value of data

There are sound economic reasons that companies may seek to collect and store as much data as possible. A 2013 study by Bain found that companies using big data analytics to change the way they do business or to improve their products and services were twice as likely to be in the financial top tier of their industry.[20] The ability to conduct broader and deeper analysis of data holdings can help businesses develop a multi-faceted view of their customers, enabling them to be more responsive to their needs and expectations while providing a platform for better personalization.

While the value of data may depreciate over time, not all information decreases in value or in the same way.[21] Even old data can help companies do real-time reviews of business transactions to prevent fraud or quickly identify errors in software and device malfunctions. The potential for value extracted from data in the future is also a driver of large-scale data retention, as this practice often pays off for businesses. Some argue that "serendipitous innovation" depends upon the exploration of data,[22] and that focused collection and minimization efforts limit the development of new services and the discoveries that may follow from valuable research.[23] Absent legal restraint, the potential economic utility of good and "bad" information alike has pushed companies to increase their data holdings.[24]

The technical challenges of deletion

While it may seem straightforward to delete or destroy information held by a company, it is anything but -- implementation challenges dominate the data destruction landscape. First and not least, complete data destruction stands in opposition to the cardinal rule of computer storage design, which is to protect user data at all costs.[25] Information is regularly recovered from devices which have been burned, crushed, or submerged in water. The Department of Education's Privacy Technical Assistance Center has warned that organizations face an environment where "modern data storage technologies are extremely resilient, and data recovery techniques have become highly advanced."[26] In order to improve drive performance and build in redundancy, hard drives isolate information, making it difficult to access and complicating efforts to securely delete it.[27]

Methods used for data destruction involve different trade-offs: the more certain the data destruction, the longer the process takes to complete.[28] Some software implementations promise complete data disposal using a method that stops short of physically destroying the entire storage medium, but at the cost of a great amount of time. Manually deleting large quantities of information is technically ineffective and time-consuming for employees. Overwriting is effective but also incredibly time-consuming, with wait times varying depending upon the size of the file or drive being rewritten as well the complexity of the erasure pattern.[29] Secure overwriting can take weeks or longer to execute on storage devices that are orders of magnitude smaller than those used at the enterprise level. Some destruction methods may also be designed to retain certain information metadata or provide limited levels of deletion granularity.[30]

The variety of formats and platforms for digital storage is another complicating factor in data destruction. Unlike the destruction of paper records, which can be as simple and routine as running documents through a shredder, complete digital erasure is more complicated. Digital systems have been designed to protect and retain user data.[31] Even when data is no longer readily accessible to an operating system or to the application that created it, it typically exists as a copy or backup on remote servers.[32] Data deletion efforts by companies can also be stymied by difficulties in appropriately categorizing and indexing data, making it hard to discern valuable data. Datasets are also increasingly intermingled and broken apart, amplifying the technical challenges for companies seeking to expunge certain types of data completely.

Finally, employees who would be trusted to carry out these technical tasks often lack basic training on how to do them. Some employees may be responsible for making decisions related to data destruction but basic training and know-how in data disposal techniques is not standard – a Ponemon Institute survey of individuals responsible for company document destruction found that while 55 percent of respondents trained employees in secure data disposal, only 38 percent were confident in their ability to securely dispose of information.[33] Another study found that nearly half of used hard drive disks available from online retailers contained residual data, and thousands of leftover emails, call logs, and other media were retrievable from 35 percent of used mobile devices, despite the fact that deletion attempts had been made on the majority of these products.[34] Some organizations turn to outside vendors for data destruction[34] to avoid these potential problems, which can expose the company's proprietary data to increased privacy and security risks.[36]

While each technical hurdle may be small in isolation, the combination of them presents a meaningful challenge for companies seeking to reduce their data stores. A vague and confusing legal requirement, strong economic incentives, and a series of technical hurdles have caused companies to seemingly forego the conclusion that data retention is necessary and beneficial.

II. The Argument for Deletion

But the rationale behind large-scale data retention fails to account for the risks that data can pose to companies and their business models. The decision to keep data should not be the default option for companies because the associated risks call for higher scrutiny. Indiscriminate data retention wastes resources on data that may not increase revenue, creates a temptation for hackers, and is a source of legal liability.

The legal risks

As discussed above, many laws on the books either loosely incentivize data retention or allow companies to handle decisions regarding destruction at their own discretion. However, taking those regulations at face value without considering broader legal liability or federal guidelines is insufficient at best, and at worst leaves companies open to serious consequences. Although the analytic value of data may diminish over time, the legal risk associated with that information increases. A survey by the Compliance, Governance and Oversight Council found that corporate information generally fits into one of four categories: one percent of data needs to be retained for litigation purposes; five percent needs to be retained for regulatory compliance reasons; 25 percent has business value; and 69 percent has little or no business value.[37] That means that almost 70 percent of a company's data assets serve mostly to create liability.

First, to comply with legal and regulatory requirements, companies must devote resources to performing time-intensive searches of their data, which is more difficult when there is more data to analyze. Companies in 2011 spent an average of 11.9 percent of their total IT budget on compliance-related activities,[38] a number that bloats when additional time or vendor resources are necessary to wade through a deluge of data.[39] These costs are exacerbated in the event of civil litigation, as well.

Electronic discovery costs have grown exponentially, and it can cost upwards of $18,000 to collect, process, and review a single gigabyte of data.[40] A study by the RAND Institute of Fortune 500 companies found that median discovery expenditures were $1.8 million, concluding that "many millions of dollars in litigation expenditures will be wasted each year until legal tradition catches up with modern technology."[41] A sophisticated data deletion strategy can help to avoid significant downstream e-discovery costs.[42]

Second, while it is true that many U.S. laws and regulations lack clear and specific guidelines around deletion and disposal, some espouse the benefits of deleting data and most regulatory frameworks recommend that entities establish reasonable and appropriate measures for getting rid of information. This includes the HIPAA Privacy and Security Rules discussed above. HHS has also released guidance[43] for covered entities on rendering protected health information "unusable, unreadable, or indecipherable to unauthorized individuals" if one of two requirements is met: the first requirement is storing the data using an NIST-approved encryption process.[44] The second requirement is that the equipment on which the PHI is stored has been cleared, purged, or destroyed in a manner consistent with NIST's media sanitization guidelines.[45]

The Gramm Leach Bliley Act's Disposal Rule applies to individuals (such as a landlord), and small and large businesses (such as lenders and employers), that use consumer reports. The Rule requires entities to establish destruction policies and mechanisms that consider "the sensitivity of the information, the costs and benefits of different disposal methods, and changes in technology."[46]

The Rule also asks entities to conduct due diligence when using an outside vendor for data disposal which might include requiring a vendor be certified by a trade association and/or reviewing an independent audit of the company's disposal policies and procedures.

These frameworks recognize the challenges inherent in data deletion, but they also offers companies considerable flexibility and leeway to pursue responsible data destruction programs or engage in third-party disposal options after appropriate due diligence.

Failing to enact a formal policy for data deletion may also leave a business vulnerable to general enforcement actions by the U.S. Federal Trade Commission (FTC). Data destruction policies frequently come under the Commission's watchful eye.[47] The agency has the ability to enforce against both unfair and deceptive trade practices as part of its consumer protection mission under Section 5 of the FTC Act, and it has directed considerable attention to monitoring how information is used across new technologies and business models.[48] Targeted enforcement actions have been the FTC's principal mechanism for addressing industry behavior, including various efforts at transferring, selling, or disposing of information.

The FTC may bring enforcement actions against companies that fail to comply with their own stated policies. It has also taken the position that failure to implement reasonable security measures is, by itself, an unfair business practice.[49] The growing universe of data security enforcement actions provides a set of baseline data management practices, and companies trafficking in consumer data are wise to familiarize themselves with the FTC's expectations.[50]

Finally, entities that collect information about European Union citizens must also be cognizant of the forthcoming General Data Protection Regulation (GDPR), which obligates data controllers (those processing and storing EU citizens' data) to "erase personal data without undue delay" in a number of circumstances.[51] While much of Article 17 focuses on the "Right to Be Forgotten," it also suggests that companies must take reasonable steps which take into account costs and available technologies to erase personal information. The GDPR's erasure right is not unlimited, but it is far more clearly defined than most American statutory data disposal requirements.

The cost of losing trust

While businesses will have seen big successes from widespread data retention, they may have also experienced a constant stream of smaller or less tangible costs. When considered collectively, this chips away at the economic argument for data retention.

The most significant and concrete legal risk of retention is that it leaves more data to be breached. Insecure database storage and targeted phishing attacks have resulted in many data breaches, which pushes security concerns from not only the devices, but to secure company retention and deletion practices as well. For example, entities in Washington state experienced data breaches between July 2015 and July 2016 of more than 450,000 total consumer records. The Washington Attorney General's report states that cyberattacks constituted the largest share of the breaches, with a substantial number of breaches coming from unauthorized third-party or employee access.

Data breaches have real costs for companies. A 2016 IBM/Ponemon study found that the average cost of data breach for companies this year was $4 million, with the loss or theft of sensitive or confidential data reaching $158 per record.[52] When breaches occur and are made public, companies are exposed to the possibility of multiple lawsuits from a host of entities including insurers, shareholders, government regulators, and consumers.[53] They may also mark the end of company leadership,

as was the case for Target's CEO Gregg Steinhafel after the company's massive 2014 data breach.[54]

In addition, the true costs of a data breach are more than the dollar value of legal fees, payments to injured customers, and fines from regulators. Awareness of data incidents, let alone loss of data, damages consumer trust, and consumers have reported that they would be less likely to use that company's products in the future should the company suffer such an attack.[55] Maintaining consumer confidence may be especially important for companies entering the Internet of Things market, where they must ensure protection of traditional personal information such as names and passwords and new varieties of sensitive sensor data generated by in-home or in-car connected devices.

Quality over quantity

The technical challenges to deletion are not without technical payoffs. First, the constant threat of attack is diminished when data is minimized, secured, and well-managed. Second, the future of analytics is in higher quality, not higher quantities, of data. The more data there is to wade through, the harder it becomes for companies to separate and extract quality data from "noise" -- one study found that most of the data held by companies is "redundant, obsolete, or trivial."[56] In some ways, the benefits of mass data collection are the heart of its most difficult challenges. Noted statistician Nate Silver has said that the misconception that massive datasets are intrinsically valuable is actually dangerous: it ignores the very real problem of interjecting human bias over neutral data, something commonly done to support arguments in areas like public health and politics.[57]

Finding the diamonds among the slurry in datasets without bias requires far more than analytics tools -- it requires objective systems of categorization and organization that assign relevance and value, and then systematically push out unnecessary data once it loses relevance or value. A PricewaterhouseCoopers and Iron Mountain study found that 43 percent of companies are getting little actual value from their big-data holdings, with 23 percent getting none at all.[58] This points to the problem of over-collection and retention, which makes it harder and more costly to derive value.[59]

Storage costs also become non-negligible as companies grow. Despite the decrease in the physical costs of data storage, excessive data storage actually costs companies huge sums of money in terms of personnel time and technical infrastructure.[60] Even with dwindling storage costs, one study found that companies are still spending a great deal of money -- an estimated $5 million per petabyte -- to retain old information.[61] Unless a company paying for such expensive storage churns out a hugely profitable idea borne from it, they will end up with a net loss of revenue.

Finally, productivity is hampered by efforts to implement software and train staff to sift through and analyze larger quantities of potentially useless or distracting data. Even when the data management is outsourced to third parties that have extensive training in secure data storage and disposal, these vendors must still be regularly and adequately monitored and audited, at a cost. This is compounded in complexity when there is a diverse set of data sources and types, as is the case in the IoT supply chain, with device manufacturers, software makers, and platforms all playing a role in managing and storing data.

C:\>del ... but how?

There are different ways both to think about and to categorize how information is destroyed. The National Institute of Standards and Technology (NIST), for example, writes in terms of "sanitizing" information, such that access to data is rendered infeasible based on certain degrees of effort.[62]

NIST defines data destruction as a process that renders paper records unreadable and, more importantly, digital data irretrievable. It results in electronic information that is forgotten, erased, deleted, completely or reliably removed, purged, sanitized, revoked, or destroyed.[63]

In the most basic terms, there are three general ways to remove information from an electronic device: deleting, overwriting, and physically destroying the drives or infrastructure holding information.[64] Within these broad categories, however, are a number techniques and variations of those techniques that work to ensure information is irretrievable and unrecoverable.

US-CERT (Computer Emergency Readiness Team)

Soft Deletion	*Basic "deletion" operations do not remove information. When a file is deleted from a personal computer, what really gets removed is the pointer to the sector of the disk where the file exists. This sort of file de-indexing simply frees up the space available to a file system to store new information, but the deleted file's content is still there until it is explicitly overwritten. Average users, both consumers and employees, frequently believe that a file's content is erased when the file name has been deleted.[65]*
Overwriting	*Overwriting is a method for destroying digital data. In the most basic terms, overwriting works to place new information in place of the data sought to be destroyed, erasing from existence the previous data values. Different methods work in a similar manner but use different implementations. For example, the Write Zero method overwrites data with a series of zeros, while Random Data uses random characters.[66] More complicated processes like the Schneier method use a combination of multiple passes of random characters as well as zeros and ones.[67]*
	The benefit of overwriting is that it allows entities both to repurpose and reuse drives and protect consumers via the same data destruction strategy. Moreover, overwriting can be done via software and deployed selectively, making it a relatively easy and cost-effective option.[68] *However, overwriting is not a data deletion panacea. For one, overwriting cannot be used where devices are damaged or non-rewritable, and it may not completely address all areas of a device where information is stored.[69] It can be particularly challenging to effectively overwrite flash-based storage media.[70]* *Overwriting can also be time-intensive. Overwriting a large-capacity drive can be a lengthy process, and this is compounded by the fact that experts disagree on how many times data must be overwritten in order to be successfully destroyed. Some believe that government agencies can recover data that has been overwritten multiple times, but research also suggests that one overwrite is sufficient to sanitize most drives.[71]*

	NIST's 2006 Media Sanitization Guidelines[72] explain that "clearing" is a level of media sanitization that "does not allow information to be retrieved by data, disk, or file recovery utilities." Overwriting is described as a an acceptable method for clearing media.[73] *NIST also describes methods appropriate for "purging" data from a system or storage device with the intent that data not be reconstructed by laboratory techniques. In many respects, the terms "clearing" and "purging" have converged.[74]*
Physical Destruction	*Physical destruction of storage media is the most extreme method to ensure information is irrecoverable. Specialized services can disintegrate, burn, melt, or pulverize devices and drives. Magnetic drives can also be degaussed, which involves applying a strong magnetic field to a drive to eliminate the data stored on the disk. Unfortunately, while physical destruction can work to make logical information impossible to retrieve, the physical matter on which the data existed is also utterly destroyed, rendering storage drives unusable in the future.*

Information disposal can also be facilitated through the use of encryption. Once data is encrypted and the encryption keys are erased, the information is rendered irretrievable.[75] After destroying an encryption key, this kind of "deletion-by-encryption" works to effectively make information irretrievable to undesired parties, a method that is increasingly viewed as viable for protecting information in cloud-based environments.[76]

Storage matters

The methods in which data is retained is a factor in determining whether or how it might be managed, secured, and eventually destroyed. For most business applications, data is stored either locally or remotely for quick access, for instance on storage area networks (SANs) or network-attached storage (NAS).[77]

New approaches to storage and processing technology may also offer some solutions to mitigating the risks associated with holding massive amounts of data. Rather than retaining data and sending it for processing to large data centers, "fog" or "edge" computing pushes data processing away from centralized data centers to devices like smartphones and laptops. The ability to send sensor data to a smartphone using the phone's computing processing capabilities allows much more rapid feedback than transmitting data through a phone and then to a central processing location. In IoT devices like fitness trackers, the movement from the core to the edge of the network could help ameliorate the processing bottleneck that occurs when data is transmitted from devices using cellular service to central databases. Another key benefit of edge computing is that it minimizes the amount of data a company needs to wrangle, retain, and record, reducing the storage costs, privacy and security risks, and potential legal liability. Edge computing also allows companies to save money by not wasting resources on transferring every bit of data -- including data that cannot be monetized -- to central servers for processing. By reducing the strain on valuable network and storage resources, companies can focus on sending useful data to servers, rather than being forced to transmit less useful, noisy data.

Cloud storage is a common choice for many businesses because it can provide comparable functionality to traditional databases, but tends to be cheaper and more reliable.[78] To manage or re-

duce data, some cloud service vendors provide clients with data sharding options. Sharding is the process of obscuring specific data by partitioning multiple tables with fewer rows and hosting each portion of the database on separate servers to spread load across many machines. Vertical sharding, the splitting of databases by column rather than row, can be used to separate types of information (for example, names from credit card numbers) to ensure that data security is maintained and sensitive information is not exposed.[79] The advantage of sharding is that it partitions data sets into smaller chunks across multiple storage devices, allowing datasets to be securely stored with more flexibility[80] and more search efficiency.

Archiving is a longer-term form of data storage that can reduce data loads by improving storage efficiency. Archives are specialized repositories used to preserve, control, maintain authenticity and integrity, accommodate physical and logical migration, and guarantee access to information and data objects over their required retention period. Data backups, which are sometimes used instead of archiving, provide some redundancy and reliability for short-term data uses but are often overwritten as critical reliability needs fade. Archiving, on the other hand, is focused on information retrieval at the file level once pieces of information in the file stop changing and need to be accessed much less frequently.

Regardless of the remote storage methods they use, businesses should ask storage vendors about their deletion and destruction policies and practices in order to properly ensure remote data is minimized and destroyed with as much fidelity as data stored locally.

IV. Recommendations

There is currently a lack of guidance for companies on effective data retention and destruction best practices that consider data value, company risk, and legal liability. Below, we make recommendations on the strategic, policy, and technical approaches companies might take to reduce their risk and liability, assess the value of their data holdings, and determine when and how to minimize, dispose of, and hold onto data.

In general, the principles of Privacy by Design (PbD)[81] offer some direction for companies. Data deletion is a key component in implementing a PbD strategy throughout a company's policies and practices. PbD says that companies should promote privacy throughout their organization and at every stage in the development of new products and services, and should maintain comprehensive data management procedures throughout the lifecycle of their products and services.[82] Embracing a PbD approach might also be useful for businesses facing increased regulation and privacy compliance challenges. The FTC's enforcement strategy is informed by the notion of PbD and compliance with the EU's General Data Protection Regulation, coming into effect in 2018, which will almost certainly require some form of PbD.

The following recommendations are intended to provide companies with actionable guidance for data management with a focus on disposal and responsible storage of data.

Audit data holdings to determine how much of stored data is adding value, such as generating revenue, to the company.

- Implement a system for categorizing the value of data, based on specific uses for information.
  - *Articulate the actual value for each data types by applying a specific use case for it. Consider whether data has real or potential value.*

- *Automated systems for categorizing data may be helpful for administrators because they perform quick evaluations of key markers of data value, such as when it was created, what type of data it is, and the number of times it has been accessed, and use this to organize and index the data.*
- *Hire or include a Data Archivist at the start of auditing data holdings. Because of their training in analytics and preservation techniques, data archivists offer valuable perspective to companies as they seek to build sustainable data management policies and protocols.*

- Review how often data has been accessed or used by the company.
  - *Criteria for disposal might include data that has not been accessed in a considered period of time,[83] data that is redundant, short-term reference files, orphaned files, outdated drafts, technical duplicates[84] and data owned by employees that are no longer with the company.*

- Determine sensitivity of data types.
  - *Most data can be categorized in terms of its value to the company and relative to its sensitivity (i.e., confidential, highly sensitive, sensitive, not sensitive, and public).*
  - *Review the 18 categories of sensitive identifiers in the HIPAA "Safe Harbor" method for de-identification.[85]*
  - *Review state data breach notification laws for particular categories of information that raise concerns. While there is no uniform definition of what constitutes sensitive data, many very different data types may be considered sensitive depending upon jurisdiction.[86]*
  - *Review methods of categorization developed by researchers handling privacy-sensitive data, such as DataTags, an open source tool born out of Harvard University's Berkman Klein Center for Internet and Society.[87]*
  - *Data managers should create a formal policy that details what the company considers sensitive and non-sensitive data.*

Create and implement formal retention and destruction policies over the life cycle of data.[88]

- Based on the audit, data managers should then create a formal policy that details company mandates for removal or disposal of different data types.
  - Create a system that deploys the categories created by an audit of the data holdings to schedule deletion dates for the categories of data based on datasets not being accessed for certain periods of time.
  - Deletion requests should also be logged so regular audits of deletion practices can be performed and provide a basis for companies to modify their deletion schedules as needed.

- Consider which software or technical approach is needed (or which vendor) to protect the privacy and security of the types of data held by the company, and determine the levels of destruction needed for each data type.
  - *Apply technical approaches, such as edge computing or differential privacy, to datasets. Complete anonymization is difficult to achieve and de-identification should not be relied upon as the sole method of data obfuscation.*
  - *Create a data life cycle that includes requirements for the regular disposal of unnecessary data.*

- Apply methods such as data archiving or encryption for data that is considered to have long-term value.
  - *Highly sensitive data may warrant stronger data destruction methods that will render the data completely irretrievable or inaccessible while less sensitive data may call for overwriting[89] or other less rigorous forms of deletion.*

- Enable data minimization.
  - *Companies must take steps to limit the amount of consumer data collected, shared, and stored. Data minimization pairs well with a deletion policy to ensure that data is kept for the minimum amount of time necessary to extract its business value before deleting it. Consider what the business need is for the data being collected and use it only for its stated purpose, then securely delete data once it outlives its usefulness.*

Use deletion-by-encryption for all data holdings at rest and in transit, regardless of sensitivity.

- Review and implement the encryption requirements of the Massachusetts information security regulation, which is generally offered as a data management baseline for protecting personal information.[90]

- NIST offers a guide on encrypted storage technologies that includes several useful standards for different contexts, including full disk encryption, virtual disk encryption, as well as individual file or folder encryption and includes guidance on steps to design and deploy cryptographic solutions.[91]

- When possible, offer users the option to "delete" their data by encrypting it and securely destroying the encryption key.

Reduce access to data internally and externally.

- Train and educate company staff.
  - *Many data breaches have occurred because of simple mistakes made by employees in data storage, security, and destruction. Companies should educate their staff about disposal, destruction, and retention management policies. Training should be tailored to different job functions, with a combination of information that is digestible to average employees and more sophisticated instruction to employees with technical expertise/responsibilities.*
  - *Add or include a data archivist to different teams across the company, particularly in areas tasked with managing large amounts of data.*

- Solicit diverse internal stakeholder input on data retention and disposal compliance.
  - *Effective legal compliance generally requires input and buy-in from across an organization, but data security and privacy compliance is especially fraught with technical complexities. The manner in which data is categorized, inventoried, and ultimately destroyed may vary from department to department. Representatives from across different business units, particularly IT, HR, and legal, should participate in order to provide insight into how company systems operate, what information is being created, and how retention and disposal guidelines can be deployed to comply with applicable law.*

- Require third-party vendors to destroy any potentially identifiable information (and all copies) once its business purpose is complete.
  - *Companies using cloud-based storage should make sure their vendors have the net-*

- *work and data security expertise to securely manage data flows. If possible, use a cloud provider that has data sharding options.*
- *Include a provision in third party contracts requiring confirmation that data has been appropriately deleted as a part of basic auditing and monitoring obligations.*

- When using a federated database management system, it is possible to segregate sensitive data into different databases and create hashed identifiers for data as it is added to each database to make it more difficult to link data across datasets without a proper access request.[92]
  - To delete data, implement a two-stage process that begins by revoking access permissions for data that has reached its expiration date, then review and purge data that meets established deletion criteria.
  - If possible, create a default setting that automatically disposes of newly-created datasets once they are not in use, unless the querying employee has a specific justification for maintaining a copy of the data.

IV. Conclusion

From massive breaches to fear of litigation and rising costs of e-discovery, the urge to collect and retain large amounts of data in the hopes of squeezing value out of it has been tempered by evidence that holding data can actively work against a company's interest.

A key solution to reducing data stores is the application of data management policies and emerging technology techniques that reduce data holdings. Data deletion works to protect individuals' privacy and security, and it helps organizations save resources that would otherwise be spent hosting information without significant value. Though businesses suffer compliance costs and damage to their reputation when data is breached or otherwise exposed, it is also incumbent on companies to also consider the effects on their customers; individuals may experience consequences such as identity theft, reputation damage, embarrassment, and public ridicule.[93] Lengthy data retention periods support the idea that information can never be deleted, which can have a chilling effect on free expression and individual autonomy.[94]

Ultimately, data management policies and procedures that include retention and deletion strategies provide a platform for improving the overall sustainability of a company, allowing it to make faster and more efficient decisions about products and services. Data deletion should be part of every company's data management tool kit.

V. Addendum

Below we offer examples of how legal mandates differ in their approaches to data deletion, including a review of FTC cases related to deletion and retention policies and practices. We also provide a closer look at how security risks in IoT devices might be mitigated through data deletion.

Sectoral Highlights

When a given dataset falls under different legal mandates, creating procedures and policies for proper disposal and deletion can be challenging or confusing. For example, the Children's Online Privacy Protection Act (COPPA) and the Family Educational Rights and Privacy Act (FERPA) touch on different types of data produced by school-age children, though FERPA applies to students' "educational records" while COPPA covers a vast array of personal information, including technical identifiers, collected from children under the age of thirteen.[95] *Different rights also attach to each law, and they address getting rid of data differently.*

Deletion in COPPA

COPPA requires companies to establish reasonable data retention and deletion procedures. But the FTC, which is primarily responsible for enforcing COPPA, provides limited detail into what these procedures might entail. When the FTC updated its COPPA Rule in 2013, it explained that its choice of language that operators maintain children's information "for only as long as is reasonably necessary" and deploy "reasonable measures" to dispose of children's information was designed to avoid "rigidity" and permit companies to determine their own data retention and destruction practices, as well as their own data deletion capabilities.[96]

The Commission expressed the hope that the COPPA Rule would encourage companies to give further thought to data destruction by requiring entities to anticipate the reasonable lifetime of data and apply the same security protections to data destruction as to its initial collection and storage.[97]

Deletion in FERPA

The U.S. Department of Education has said that FERPA does not require entities to destroy education records as part of regular business.[98] *Many local jurisdictions actually require lengthy retention periods for certain information in order to measure student attendance, graduation rates, and perform statewide analytics and academic achievement measurements. The Department recognizes that local schools and districts often elect to establish their own record retention policies, time frames, and destruction policies in order to minimize IT costs and reduce the chance of disclosure of sensitive information.*[99]

FTC Action Highlights

The Federal Trade Commission (FTC) has long promoted data minimization as a core privacy-protecting practice. The agency encourages companies to examine their data practices and business needs in order to develop policies that impose reasonable limits on the collection and retention of data and ensure data is disposed of once it is no longer needed.[100] Failure to responsibly dispose of data once it is no use to an organization leaves companies vulnerable to enforcement actions by the FTC, and the Commission's existing history of privacy and security enforcement work provides some basic lessons in data destruction.

From Low-Tech Dumpster Diving to Securing Digital Drives

The FTC has brought a number of major enforcement actions against companies for tossing sensitive information into dumpsters,[101] and its first action pursuant to the Disposal Rule in 2008 came against a mortgage company that left hundreds of documents in an unsecured dumpster in open trash bags.[102] The FTC has also brought enforcement actions against companies for selling surplus hard drives and computer equipment without sufficiently deleting data first, exposing the sensitive information of some 34,000 customers.[103]

The FTC has brought actions against companies for failing to secure information on drives when in transit. Cases have been brought against companies whose employees' negligence resulted in the theft of devices, such as laptops, backup tapes, and external hard drives containing sensitive information, from the employees' cars.[104]

The FTC's Data Disposal Rule for Consumer Reports

The FTC has also played an active role in regulating the use and disposal of consumer reports under the Fair and Accurate Credit Transactions Act (FACTA) and earlier Fair Credit Reporting Act (FCRA). Under FACTA, the FTC promulgated a Disposal Rule that addresses some of the privacy and security risks that can arise from haphazard disposal of consumer reports, and it requires companies to take "appropriate measures to dispose of sensitive information" found in the consumer reports they use.[105] It applies to anyone who uses consumer reports, including employers, landlords, and lenders and insurers. However, the Disposal Rule's standards permit companies and individuals to determine their disposal measures based on a combination of "the sensitivity of the information, the costs and benefits of different disposal methods, and changes in technology."[106] This could include burning, pulverizing, or shredding physical documents, which is straightforward enough, but recommendations that electronic media be destroyed or erased may be less helpful.

Encouraging Data Minimization to Avoid Unreasonable Risks

BJ's Wholesale Club collected customer credit and debit card information to process retail transactions, and then stored this information for days after any sale was complete. After hackers stole account data and used it to make counterfeit payment cards, the FTC brought an action arguing that the company's retention policy not only violated banking rules, but also created an unreasonable risk to consumers by holding onto information absent a legitimate business need.[107]

False Promises Around Data Deletion

In 2016, the FTC settled allegations against the adult dating website, Ashley Madison, that it not only failed to protect account information but that the site failed to delete account information in response to user requests.[108] While Ashley Madison offered a basic deactivation option, the company promised individuals that its system was "100% secure" because users could delete their "digital trail."[109] It marketed a $19 "Full Delete" feature that promised a way for customers to remove their profiles from Ashley Madison entirely, removing profiles and messages as well as site usage history and other personally identifiable information.[110] The FTC charged that this information was frequently not removed from the site's internal systems and that consumers were informed after paying for the "Full Delete" that certain information could still be retained for six to twelve months.[111]As part of a settlement with the FTC, numerous state attorneys general, and several foreign privacy regulators, Ashley Madison was forced to pay a $1.6 million dollar fee and prohibited from misrepresenting the terms and conditions for deleting user information.

Mergers and Acquisitions

RadioShack attempted to sell consumers' names, addresses, email addresses, and purchase histories to pay its debts, and the FTC intervened to note that this proposed sale ran counter to several promises made to consumers by RadioShack in its privacy policy.[112]

The agency recommended a series of conditions be placed on the sale of such informational assets. Specifically:

- *Customer information had to be bundled with other assets and not sold individually as a standalone asset;*
- *Customer information must be sold only to another company in substantially the same line of business -- and the buyer needed to agree to be bound by Radioshack's existing privacy policies.*
- *Before customer data could be used by the buyer in a materially different way than that described by Radioshack's privacy policy, the buyer would need to provide notice and obtain affirmative consent from Radioshack's customers.*[113]

Deletion to reduce security risks in the IoT

In the IoT, a data deletion policy and responsible data stewardship program can mitigate some security risks to companies and consumers. Security risks facing IoT manufacturers and vendors stem, in part, from the amount and type of data collected, and how this data is transferred and stored.

The security of the device itself is paramount. Two key security concerns are the theft of the device resulting in unauthorized access of data and a breach of local wireless communication protocols like Bluetooth to access the device owner's information. In conjunction with data minimization and ongoing removal of non-valuable data, companies should give customers the ability to automatically delete their information from a device, especially if the device is stolen, and make them aware of ways to protect the security of their device. To address the insecurity of wireless communications, IoT companies should use the guidelines developed by NIST on implementing secure Bluetooth technology.[114]

Attacks on the data held by an IoT company itself is another vector of insecurity that may be mitigated by deletion policies that include encryption. IoT companies may have data that is highly sensitive, such as information about a user's habits in their home, and may have multiple connections to other devices or apps. When customer information lives in centralized databases, it can be accessible by employees, which subjects the company to attacks should authorized employee login credentials be improperly disclosed through phishing or other attacks.

In 2014, eBay experienced a data breach wherein hackers gained access to and copied parts of a database that housed 145 million customer records, including customer passwords, email addresses, birth dates, mailing addresses, among other information.[115] *Luckily, the passwords were stored in an encrypted format, making the information extremely difficult for the hackers to convert into usable information. Deletion-by-encryption is a valuable tool for IoT companies to protect against data breaches and the resulting costs and harm to consumers and brand; encrypting data stored in central servers and using secure protocols to transfer data to be stored for later use are two critical ways to render data useless for hackers. Encryption should be a part of a responsible data stewardship program, which begins with establishing a data life cycle that details retention and destruction policies.*

Endnotes

[1] Yahoo revealed in December of 2016 that over a billion records had been stolen from its user databases. See Lily Hay Newman, *Yahoo Hack Leaves One Billion Accounts Compromised*, WIRED (Dec. 14, 2016), https://www.wired.com/2016/12/yahoo-hack-billion-users/.

[2] Mimi Onuoha, *What It Takes To Truly Delete Data*, FiveThirtyEight (Jan. 30, 2017), https://fivethirtyeight.com/features/what-it-takes-to-truly-delete-data/.

[3] *Id.*; *Your Threat Model Just Changed*, Electronic Frontier Foundation (2016), https://www.eff.org/files/2016/12/20/eff_wired_ad_2016_fixed.jpg.

[4] http://neveragain.tech/

[5] Lee Rainie & Maeve Duggan, Privacy and Information Sharing (Pew Research Ctr. 2016), http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/.

[6] Kenneth Olmstead & Aaron Smith, Americans and Cybersecurity (Pew Research Ctr. 2017), http://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/

[7] *The CIS Critical Security Controls for Effective Cyber Defense*, Center For Internet Security (August 31, 2016), https://www.cisecurity.org/critical-controls/documents/CSC-MASTER-VER61-FINAL.pdf.

[8] R. Thomas Howell, Jr. & Rae N. Cogar, Records Retention - an Essential Part of Corporate Compliance, in Am. Bar & Records Retention and Destruction Current Best Practices 1-10 (Am. Bar 2003), http://apps.americanbar.org/buslaw/newsletter/0019/materials/recordretention.pdf.

[9] 18 U.S.C. § 1520(a)(1), *available at* https://www.gpo.gov/fdsys/pkg/PLAW-107publ204/html/PLAW-107publ204.htm.

[10] Disposal of Consumer Report Information and Records, 16 C.F.R. pt. 682 (2004), *available at* https://www.gpo.gov/fdsys/pkg/FR-2004-11-24/pdf/04-25937.pdf.

[11] 45 C.F.R. § 160.102.

[12] *Id.* at 160.103.

[13] U.S. Department of Health & Human Services, *What do the HIPAA Privacy and Security Rules require of covered entities when they dispose of protected health information?*, HIPAA FAQ (Feb. 18, 2009), http://www.hhs.gov/hipaa/for-professionals/faq/575/what-does-hipaa-require-of-covered-entities-when-they-dispose-information/.

[14] *See generally, Data Disposal Laws,* Nat'L Conference of State Legislatures (Dec. 1, 2016), http://www.ncsl.org/research/telecommunications-and-information-technology/data-disposal-laws.aspx (showing the applicability of each state's data disposal laws and providing links to the relevant statutes).

[15] *See, e.g.,* Ore. Rev. Stat. § 646A.622; Nev. Rev. Stat. Ann. § 603A.200(1).

[16] *See generally, Security Breach Notification Laws*, Nat'L Conference of State Legislatures (Jan. 4, 2016), http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx (providing links to the security breach notification laws for the states that have passed such statutes).

[17] Opening Remarks of FTC Chairwoman Edith Ramirez Privacy and the IoT: Navigating Policy Issues International Consumer Electronics Show (Jan. 6, 2015), *available at* https://www.ftc.gov/system/files/documents/public_statements/617191/150106cesspeech.pdf; *see also* FTC Staff Report, *Internet of Things: Privacy & Security in a Connected World iv* (Jan. 2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

[18] Executive Office of the President, President's Council of Advisors on Science and Technology, *Big Data and Privacy: A Technological Perspective x* (May 2014), https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

[19] For example, both the NAI and DAA self-regulatory codes emphasize limits on the use of data for online behavioral advertising, but they do not restrict companies' ability to collect data in the first instance. Once information is collected, companies are instructed only to retain it so long

as they have a legitimate business need, which while sensible, provides little actual guidance in practice.

[20] Rasmus Wegener & Velu Sinha, *The Value of Big Data: How Analytics Differentiates Winners*, Bain & Company (2013), http://www.bain.com/Images/BAIN%20_BRIEF_The_value_of_Big_Data.pdf.

[21] Viktor Mayer-Schönberger & Kenneth Cukier, Big Data: A Revolution that Will Transform how We Live, Work, and Think 110-111 (2013).

[22] Daniel Castro & Travis Korte, *Data Innovation 101*, Center for Data Innovation (Nov. 2013), http://www2.datainnovation.org/2013-data-innovation-101.pdf.

[23] Chris Wolf & Jules Polonetsky, *An Updated Privacy Paradigm for the Internet of Things*, Future of Privacy Forum (Nov. 2013), https://fpf.org/wp-content/uploads/Wolf-and-Polonetsky-An-Updated-Privacy-Paradigm-for-the-%E2%80%9CInternet-of-Things%E2%80%9D-11-19-2013.pdf.

[24] *See* Mayer-Schönberger & Cukier, *supra* note 19.

[25] Gordon Hughes & Tom Coughlin, Tutorial on Disk Drive Data Sanitization, http://cmrr.ucsd.edu/people/Hughes/documents/DataSanitizationTutorial.pdf.

[26] Privacy Technical Assistance Ctr., Best Practices for Data Destruction 5 (2014), http://ptac.ed.gov/sites/default/files/Best%20Practices%20for%20Data%20Destruction%20(2014-05-06)%20%5BFinal%5D.pdf.

[27] Diesburg, S. M. and Wang, A. A., *A survey of confidential data storage and deletion methods*. ACM Comput. Surv. 43, 1, Article 2 (November 2010), *available at* https://www.cs.fsu.edu/~awang/papers/csur2010.pdf.

[28] *Id.* at 30.

[29] *Id.* at 32.

[30] *Id.*

[31] Hughes & Coughlin, *supra* note 23.

[32] *Data destruction*, TechTarget, http://searchstorage.techtarget.com/definition/data-destruction (last visited on Dec. 20, 2016).

[33] Ponemon Inst., Security of Paper Records & Document Shredding (2014), http://www.cintas.com/customer_applications/DM-Ponemon-Study-And-Webinar/pdf/Ponemon%20Study%20-%20The%20Security%20of%20Paper%20Records%20and%20Document%20Shredding.pdf.

[34] See Michael Kan, *Used hard drives on eBay, Craigslist are often still ripe with leftover data*, PCWorld (Jun. 28, 2016), http://www.pcworld.com/article/3089343/security/resold-hard-drives-on-ebay-craigslist-are-often-still-ripe-with-leftover-data.html.

[35] Ponemon supra note 31.

[36] Though encryption can serve as an extra safeguard when transferring data to third parties and vendors for disposal, the use of deletion-by-encryption or other deletion-by-encryption methods introduces cryptography overhead into the process of information disposal and so it less commonly used. Such information could also potentially be recovered in the future as decryption efforts advance.

[37] Michael Osterman, *Can you defend your decision to delete data?*, Titus (July 14, 2014), http://www.titus.com/titus-blog/2014/07/can-you-defend-your-decision-to-delete-data/.

[38] Ponemon Inst., The True Cost of Compliance, http://www.tripwire.com/tripwire/assets/File/ponemon/True_Cost_of_Compliance_Report.pdf.

[39] Hmong Vang, *5 Hidden Costs of Data Security and Compliance*, Corporate Compliance Insights (April 20, 2016), http://corporatecomplianceinsights.com/5-hidden-costs-data-security-compliance/.

[40] Matthew Werdegar & Benedict Hur, *E-Discovery Trends You Can't Afford to Ignore*, Corporate Counsel (Sept. 26, 2014), http://www.corpcounsel.com/id=1202671501269/3-Ediscovery-

Trends-You-Cant-Afford-to- Ignore?slreturn=20150030155830; *see also* Press Release, New IDC Forecast Shows Worldwide eDiscovery Market Surpasses $10 Billion in 2015 (Jan. 4, 2016), https://www.idc.com/getdoc.jsp?containerId=prUS40881916.

[41] Nicholas M. Pace and Laura Zakaras, Where the Money Goes: Understanding Litigant Expenditures for Producing Electronic Discovery 17, xx (2012), available at http://www.rand.org/content/dam/rand/pubs/monographs/2012/RAND_MG1208.pdf.

[42] Dean Gonsowski, *E-discovery costs: Pay now or pay later*, Inside Counsel (May 23, 2012), http://www.insidecounsel.com/2012/05/23/e-discovery-costs-pay-now-or-pay-later.

[43] *Guidance to Render Unsecured Protected Health Information Unusable, Unreadable, or Indecipherable to Unauthorized Individuals*, Dept. of Health and Human Services, https://www.hhs.gov/hipaa/for-professionals/breach-notification/guidance/index.html (last visited on Feb. 9, 2017).

[44] Karen Scarfone, Murugiah Souppaya & Matt Sexton, Guide to Storage Encryption Technologies for End User Devices (2007), http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-111.pdf.

[45] Richard Kissel et al., Guidelines for Media Sanitization (Nat'l Inst. of Standards & Tech. 2014), http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r1.pdf.

[46] *Disposing of Consumer Report Information? Rule Tells How*, Fed. Tradde Comm'n. (June 2005), https://www.ftc.gov/tips-advice/business-center/guidance/disposing-consumer-report-information-rule-tells-how.

[47] *See* In the Matter of BJ's Wholesale Club, Docket No. C-4148, *Decision and Order* (F.T.C. Sep. 23, 2005); *see also* 16 C.F.R. pt. 682 (2004).

[48] Fed. Trade Comm'n, *Privacy & Data Security Update (2015)* (Jan. 2016), https://www.ftc.gov/reports/privacy-data-security-update-2015. In its update, the FTC declares that it "has unparalleled experience in consumer privacy enforcement" across practices "offline, online, and in the mobile environment." The FTC policies privacy and data security through targeted enforcement, as well as via studies and reports, public workshops, and targeted educational materials for consumers and businesses.

[49] *Fed. Trade Comm'n v. Wyndham Worldwide Corp.*, No. 14-3514, at 6 (3rd. Cir. Aug. 24, 2015), *available at* https://www.ftc.gov/system/files/documents/cases/150824wyndhamopinion.pdf. This decision by a federal appeals court upheld a lower court ruling that detailed how the FTC could use Section 5's prohibition on unfair practices to challenge the data security lapses.

[50] Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 Colum. L. Rev. 583 (2014) (explaining that FTC enforcement actions, settlement orders, and staff opinions provide a "rich jurisprudence that is effectively the law of the land for businesses that deal in personal information.").

[51] 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), Article 17.

[52] Ponemon Inst., *2016 Cost of Data Breach Study: Global Analysis*, IBM (June 2016), http://www-03.ibm.com/security/data-breach/.

[53] Melissa Maleske, T*he 6 Lawsuits All GCs Face After A Data Breach*, Law360 (December 9, 2015), http://www.law360.com/articles/735838/the-6-lawsuits-all-gcs-face-after-a-data-breach.

[54] Eric Basu, *Target CEO Fired - Can You Be Fired If Your Company Is Hacked?*, Forbes (June 15, 2014), http://www.forbes.com/sites/ericbasu/2014/06/15/target-ceo-fired-can-you-be-fired-if-your-company-is-hacked/#7e47910f7bc1.

[55] *Beyond the Bottom Line: The Real Cost of Data Breaches*, FireEye (2016), *available at* https://www2.fireeye.com/WEB-Real-Cost-of-Data-Breaches.html.

[56] *Id.*

[57] Matt Asay, *Nate Silver Gets Real About Big Data,* ReadWrite (March 29, 2013),

http://readwrite.com/2013/03/29/nate-silver-gets-real-about-big-data/.

[58] *Seizing the information advantage*, Iron Mountain (September 2015), http://www.ironmountain.com/Knowledge-Center/Reference-Library/View-by-Document-Type/White-Papers-Briefs/S/Seizing-The-Information-Advantage.aspx.

[59] https://redowl.com/2016/11/behind-scenes-threat-intel-data-doesnt-always-equal-insights/, http://kinshipdigital.com/2016/09/12/more-data-doesnt-mean-better-data/

[60] Cindy LaChapelle, *The Cost of Data Storage and Management: Where is it Headed in 2016?*, Data Center Journal (2016), http://www.datacenterjournal.com/cost-data-storage-management-headed-2016/.

[61] *The Databerg Report: See What Others Don't*, Veritas (2016), https://www.veritas.com/content/dam/Veritas/docs/reports/scd_veritas_strike_summary_a4-ls-usa_final.pdf.

[62] Kissel et al., *supra* note 45.

[63] Joel Reardon, David Basin & Srdjan Capkun, SoK: Secure Data Deletion (IEEE Symposium on Sec. & Privacy 2013), http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6547117.

[64] Linda Pesante et al., Disposing of Devices Safely (United States Computer Emergency Readiness Team 2012), https://www.us-cert.gov/sites/default/files/publications/DisposeDevicesSafely.pdf.

[65] Diesburg, S. M. and Wang, A. A., *A survey of confidential data storage and deletion methods.* ACM Comput. Surv. 43, 1, Article 2 (November 2010), *available at:* https://www.cs.fsu.edu/~awang/papers/csur2010.pdf

[66] Tim Fisher, *What is the Schneier Method?*, Lifewire (Aug. 22, 2016), https://www.lifewire.com/what-is-the-schneier-method-2626000.

[67] *Id.*

[68] Bob Violino, *The in-depth guide to data destruction*, CSO Online (Feb. 6, 2012), http://www.csoonline.com/article/2130822/it-audit/the-in-depth-guide-to-data-destruction.html.

[69] See Diesburg, *supra* note 27, at 8.

[70] Wei, M. Y. C., Grupp, L. M., Spada, F. E., & Swanson, S. (2011, February). Reliably Erasing Data from Flash-Based Solid State Drives. In FAST (Vol. 11, pp. 8-8). https://www.usenix.org/legacy/events/fast11/tech/full_papers/Wei.pdf.

[71] Brian Smithson, *The Urban Legend of Multipass Hard Disk Overwrite*, Infosec Island (Aug. 28, 2011), http://www.infosecisland.com/blogview/16130-The-Urban-Legend-of-Multipass-Hard-Disk-Overwrite.html.

[72] Richard Kissel et al., Guidelines for Media Santization (Nat'l Inst. of Standards & Tech. 2006), http://ws680.nist.gov/publication/get_pdf.cfm?pub_id=50819.

[73] *Id. at 7.*

[74] *Id. at 6.*

[75] Courtney Bowman et al., The Architecture of Privacy: On Engineering Technologies that Can Deliver Trustworthy Safeguards 121-22 (O'Reilly Media, Inc. 2015).

[76] *Id. at 122.*

[77] SANs provide access to consolidated, block level data storage primarily composed of disk arrays. The arrays are accessible to servers and appear to end users as locally attached devices. NAS functions similarly, but provides file-level data storage connected to a computer network, providing data access to a heterogeneous group of clients.

[78] Center for Democracy & Technology, *FAQ: HIPAA and Cloud Computing* (August 7, 2013), *available at:* https://cdt.org/files/pdfs/FAQ-HIPAAandCloud.pdf.

[79] *Partitioning*, Microsoft, https://technet.microsoft.com/en-us/library/ms178148(v=sql.105).aspx (last visited Dec. 20, 2016).

[80] *Data Partitioning Guidance*, Microsoft, https://msdn.microsoft.com/en-us/library/dn589795.aspx (last visited Dec. 20, 2016).

[81] Ann Cavoukian, *Privacy by Design: The 7 Foundational Principles*, Internet Architecture Bd. (2011), https://www.iab.org/wp-content/IAB-uploads/2011/03/fred_carter.pdf.

[82] *Protecting Consumer Privacy in an Era of Rapid Change* 22-32, Federal Trade Commission (Mar. 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

[83] An appropriate time frame may vary, but existing laws and regulations may provide a guide. For example, research activities could take guidance from the Common Rule, which governs federal policy for human subject testing. The Rule specifies that records relating to research must be retained for at least three years after the completion of a research activity. 45 C.F.R. §46.115(b).

[84] Aliye Ergulen, et al., Disposing of Digital Debris (2014), http://www.edrm.net/resources/edrm-white-paper-series/disposing-of-digital-debris.

[85] *Guidance Regarding Methods for De-Identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule*, U.S. Dept. of Health and Human Services (November 26, 2012), https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/.

[86] *See, e.g.,* Daniel Solove, *What Is Sensitive Data? Different Definitions in Privacy Law*, TeachPrivacy (July 31, 2014), https://www.teachprivacy.com/sensitive-data-different-definitions-privacy-law/. A Congressional Research Service report on information security and breach notification laws explains that "sensitive personally identifiable information" is a subset of PII which "typically includes any information about an individual (including education, financial transactions, medical history, and criminal or employment history) along with information that can be used to distinguish or trace the individual's identity (including name, address, or telephone number; date and place of birth; mother's maiden name; Social Security number or other government-issued unique identification number; biometric data; or unique account identifiers)." Gina Stevens, Congressional Research Serv., Data Security Breach Notification Laws (Apr. 10, 2012), https://fas.org/sgp/crs/misc/R42475.pdf.

[87] *See Datatags*, Harvard University Privacy Tools Project (2014), http://privacytools.seas.harvard.edu/datatags.

[88] *See generally, Governance and Compliance*, Box, https://www.box.com/security/governance-and-compliance (promoting the various tools Box has made avialable to ensure data lifecycle regulation compliance).

[89] Overwriting data can be performed to any of several data destruction standards. For example, the commercially available software, Blancco Drive Eraser, provides individual consumers the option to overwrite their storage device with up to thirty-five passes of random data. See What Download is Best for You?, Blancco, https://dban.org/.

[90] *See* 201 Mass. Code Regs. 17.00 (2009).

[91] *See* Scarfone *supra* note 44.

[92] For a more detailed discussion of privacy measures in federated databases, please see: Bhavani Thuraisingham, *Privacy Constraint Processing in a Privacy-enhanced Database Management System*, 55 Data & Knowledge Engineering 159 (2005), https://pdfs.semanticscholar.org/9a1e/a9587b1d3d16fd2a22914e7347664f7841d9.pdf.

[93] Kim Zetter, *Hackers Finally Post Stolen Ashley Madison Data*, WIRED (Aug. 15, 2015), https://www.wired.com/2015/08/happened-hackers-posted-stolen-ashley-madison-data/.

[94] Justin Brookman & G.S. Hans, Why Collection Matters: Surveillance as a De Facto Privacy Harm (2013), http://www.futureofprivacy.org/wp-content/uploads/BrookmanWhy-Collection-Matters.pdf.

[95] Dep't of Education, Office of Educational Tech., Privacy, https://tech.ed.gov/privacy/ (last visited Jan. 25, 2017).

[96] 78 Fed. Reg. 3995 (Jan 17, 2013), *available at* https://www.ftc.gov/system/files/documents/fed

eral_register_notices/2013/01/2012-31341.pdf. *See also* 16 C.F.R. § 312.10.

[97] *Id.*

[98] *See* PTAC Best Practices, *supra* note 26.

[99] *Id.*

[100] Fed. Trade Comm'n, FTC Staff Report: Internet of Things: Privacy and Security in a Connected World iv (Jan. 2015), https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf.

[101] *Start with Security* 14, Fed. Trade Comm'n (June 2015), https://www.ftc.gov/system/files/documents/plain-language/pdf0205-startwithsecurity.pdf.

[102] United States v. American United Mortgage Company, No. 07C 7064 (N.D. Il December 18, 2007), *available at* https://www.ftc.gov/enforcement/cases-proceedings/062-3103/american-united-mortgage-company-united-states-america-ftc.

[103] Goal Financial, Docket No. C-4216 (F.T.C. April 9, 2008), *available at* https://www.ftc.gov/sites/default/files/documents/cases/2008/04/080415decision_0.pdf.

[104] Accretive Health, Docket No. C-4432 (F.T.C. February 24, 2014), *available at* https://www.ftc.gov/system/files/documents/cases/140224accretivehealthdo.pdf; CBR Systems, Docket No. C-4400 (F.T.C. May 3, 2013), *available at* https://www.ftc.gov/sites/default/files/documents/cases/2013/05/130503cbrdo.pdf.

[105] *Disposing of Consumer Report Information? Rule Tells How*, Fed. Trade Comm'n. (June 2005), https://www.ftc.gov/tips-advice/business-center/guidance/disposing-consumer-report-information-rule-tells-how.

[106] *Id.*

[107] *Start with Security 14*, Fed. Trade Comm'n, (June 2015), https://www.ftc.gov/system/files/documents/plain-language/pdf0205-startwithsecurity.pdf.

[108] *Operators of AshleyMadison.com Settle FTC, State Charges Resulting From 2015 Data Breach that Exposed 36 Million Users' Profile Information*, Fed. Trade Comm'n. (Dec. 14, 2016), https://www.ftc.gov/news-events/press-releases/2016/12/operators-ashleymadisoncom-settle-ftc-state-charges-resulting (providing a summary of the FTC action against AshleyMadison.com and the agreed upon terms of the agreed upon settlement).

[109] *Stipulated Order for Permanent Injunction and Other Equitable Relief* at 6, Federal Trade Commission v. Ruby Corp., Case 1:16-cv-02438 (D.D.C. Dec. 14, 2016), *available at* https://www.ftc.gov/system/files/documents/cases/161214ashleymadisoncmplt1.pdf.

[110] *Id.* at 7.

[111] *Id.*

[112] *See FTC Requests Bankruptcy Court Take Steps to Protect RadioShack Consumers' Personal Information*, Fed. Trade Comm'n. (May 18, 2015), https://www.ftc.gov/news-events/press-releases/2015/05/ftc-requests-bankruptcy-court-take-steps-protect-radioshack.

[113] *Id.*

[114] John Padget, et al, Guide to Bluetooth Security (Nat'l Inst. of Standards & Tech. 2016), http://csrc.nist.gov/publications/drafts/800-121/sp800_121_r2_draft.pdf.

[115] Jim Finkle, *Hackers Raid EBay in Historic Breach, Access 145 Million Records*, Chi. Trib., May 22, 2014, *available at* http://articles.chicagotribune.com/2014-05-22/business/sns-rt-us-ebay-password-20140521_1_passwords-hackers-rapid7.