

# Protecting Personal Information: Obstacles and Directions

Rachel Greenstadt and Michael D. Smith Harvard University

{ greenie,smith }@eecs.harvard.edu

May 4, 2005

## ABSTRACT

We present a framework for comparing and evaluating approaches for achieving electronic informational privacy. Our framework focuses on the issues of decisionmaking, negotiation, and enforcement, which are fundamental in determining what personal information is considered private and what uses of that information are considered improper. Our analysis of four leading approaches finds that none adequately address all three issues. This analysis explains why marketbased approaches are currently infeasible, and why regulatory approaches are a reasonable interim solution. We also suggest research directions that could lead to more flexible approaches.

## INTRODUCTION

Violations of information privacy, also called data protection, occur when personal information is improperly collected, used, or disclosed [29, 36]. Setting aside for a moment the question of when an action is improper, these violations are a cost that individuals suffer in the information exchanges that drive the global economy. In particular, it costs businesses essentially nothing to retain and correlate personal information collected electronically during a commercial transaction. However, if this information is later used, for example, to produce unwanted solicitations, to enable discriminatory practices, or to steal an identity, an individual's quality of life directly suffers. For privacy to improve, this negative externality must be addressed. Our goal is to identify an approach that accomplishes this while simultaneously balancing the privacy choices of individuals with the information needs of businesses and governments.

When electronic information privacy issues started to receive attention, many economists feared that personal information would be either restrictively protected by reactionary legislation or rampantly abused by profitdriven businesses [21, 40]. The former would stifle an economy with a growing dependence on the flow of electronic information, and the latter would incur costs on the individuals whose personal information was being bought and sold without their participation or consent. Though neither of these Draconian scenarios came to pass, no single model has emerged that adequately addresses information privacy.

In this paper, we construct a framework for discussing and comparing four influential models in the literature that claim to address electronic information privacy and ameliorate the negative externality that exists today. Our framework focuses on three fundamental issues that each model must successfully address in order to achieve the claims. At the heart of these issues are the questions of what personal information is considered private, and what uses of that information are considered improper.

The framework helps to explain how each existing model falls short as a solution to information privacy, and it helps to identify technology and policy questions that cut across all four models. With this understanding, we believe it is possible to chart a research path in economics, computer science, and public policy that will result in material progress toward a working model for information privacy that ameliorates the negative externality, provides for individual privacy choices, and supports the information needs of businesses and governments.

At a high level, the literature in information privacy describes four different models: selfregulation, government regulation, thirdparty regulation, and information markets. These models differ in where the responsibility for the protection of personal information is placed.

The most common models in use today are selfregulation and governmental regulation. Selfregulation places the responsibility in the hands of those (e.g., businesses) that gather, use, and sell personal information. Except for a few categories of personal information, this is the predominant model in the United States. Government regulation, based on either comprehensive or sectoral laws, relies on the judicial and legislative branches of a government for the protection of personal information. The United States has historically preferred to regulate particular sectors of the data

space, such as health care or financial data, while the European Union has enacted general data protection rules.

Thirdparty regulation and information markets have been proposed as models with the potential to overcome the shortcomings of selfand government regulation. Thirdparty regulation can be thought of as a generalization of government regulation, where some independent but trusted third party assumes responsibility for the protection of personal information used by businesses and governments. Though this model has traditionally been driven by technology considerations [20, 25], it may grow in popularity as governments consider greater uses of personal information in law enforcement and the fight against terrorism—as governments begin to look less and less like a disinterested third party. Finally, information markets promote the concept of personal information as a property right, providing individuals with direct control over their personal information. Advocates for information markets claim that such an approach is more powerful, flexible, and economically efficient than the existing regulatory approaches [21].

Our goal is to be able to compare and evaluate these different models within a unifying framework. Tang, Hu, and Smith [35] present the only other similar effort of which we are aware. Their work compares selfregulation, government regulation, and thirdparty regulation via sealofapproval programs. It suggests that sealofapproval programs can provide socially optimal privacy protection if the fees and penalties associated with these programs are chosen effectively. Our work shows that this is a necessary but not sufficient condition. In particular, Tang, Hu, and Smith assume that consumer privacy is perfectly protected when a retailer chooses to protect privacy. They do not address the enforcement difficulties faced by different approaches, a key issue in our framework. Our work directly addresses the question of what it takes for retailers to protect privacy and for consumers to believe in that privacy protection.

Our position is that information privacy is too strongly an individual issue and pure regulation is too rigid and inefficient an approach for regulation to be the “right” answer in the long term. The flexibility and efficiency of market-based approaches make them appear better capable of handling the wide range of individual privacy preferences as well as the future information needs of businesses and governments. However, as our analysis makes clear, marketbased approaches to information privacy are significantly more complex than regulatory approaches and thus more difficult to get started. Still, we argue that it is possible to leverage advances in the regulatory models to move in the direction of the more flexible and efficient marketbased approaches.

This paper has three main contributions:

- We present a framework for comparing and evaluating models seeking to achieve information privacy and eliminate the negative externality that exists today.
- We use this framework to understand the shortcomings of existing models and identify the open, crosscutting technology and policy research questions that must be solved for the current models to ameliorate the negative externality associated with information privacy.
- And finally, we offer an explanation for why marketbased approaches may need to follow the development of regulatory approaches.

We begin in Section 2 with a brief look at information privacy, what constitutes violations of it, how individuals feel about it, and what is occurring in the growing information industry. Section 3 presents our framework and discusses how existing models fare with respect to the fundamental issues underlying current approaches to information privacy. Section 4 looks at information markets in more detail and discusses the problem of getting them started. Section 5 presents our conclusions and an option for moving forward.

## **INFORMATIONAL PRIVACY**

To understand how to achieve information privacy, we must understand how it is violated. For the purposes of this paper, information privacy is violated when personal information is collected unbeknownst to individuals and when personal information, which may have been given freely and knowingly, is later used or disclosed in a manner outside the original agreement or understanding.

Awareness of the collection of personal information is a first step in achieving information privacy. In other words, controlling the flow of one’s personal information starts at the point of collection. With awareness, a choice can be made whether to proceed with some action that will result in the release of personal information.

Once an individual chooses to release some portion of his or her personal information, the individual must then rely on laws or mechanisms to control the further distribution and subsequent use of that information. We believe that the level of information privacy perceived by individuals depends heavily on the successful implementation of these controls.

Consider an online purchase of an item where an individual provides her name, address, and credit card number in

order to pay for and receive a desired item. Personal information is given in order to complete the transaction, and we say that completing the transaction is the primary use of the personal information. Without (implicit or explicit) agreements for other uses, privacy is violated if the merchant later uses that personal information in a manner outside of this primary use (e.g., the merchant sells his customer list) or allows the information to be disclosed to a party not involved in the primary use (e.g., the merchant allows a hacker to steal the personal information of his customers). We refer to any use of the personal information outside of its primary use as a secondary use<sup>1</sup>.

The concern over secondary uses has driven much of the design of the existing models for information privacy. These models, however, if expected to be widely adopted, cannot drastically alter the primary use of personal information by individuals and businesses. Most people are comfortable with primary uses of personal information, as such uses improve the efficiency of our economic transactions [40]. It is the disclosure and use of personal information outside of its primary use that people fear. Many surveys during the Internet boom clearly support this position and indicate that individuals felt like they incurred a cost from this practice. For example, a Business Week / Harris poll in March 2000 [6] found that 78% of individuals that shopped online were concerned that their personal information would be used to send them unwanted information, an increase of 13% over a similar survey in February 1998.

A more recent Harris poll [37] shows that, though fewer consumers in 2003 than in 1999 feel that they have lost total control over how their personal information is collected and used by companies, more than half feel that existing laws and organizational practices do not provide a reasonable level of information privacy protection. This is an increase of 15 percentage points from 1999 [37].

Certainly, the lack of adequate information privacy protection has not kept businesses and governments from stepping up their efforts at collecting and using personal information [29]. In fact, in recognition of the economic, law enforcement, and antiterrorism uses of personal information, corporations in the information industry have actively looked for ways over the past decade to increase the size and scope of their databases of personal information. Governments and their law enforcement agencies have also begun to delegate the collecting and warehousing of personal information to these information service providers [29]. ChoicePoint, based in Alpharetta, Georgia, is an example of one of the leaders in the information industry [26]. Over the past seven years, ChoicePoint has acquired more than 50 other information companies [27]. The personal information in their database is sold to law enforcement and other governmental agencies, Fortune 1000 companies, and even individuals [9]. Unfortunately, these large databases are also tempting targets for criminals [27].

The overall result of all of these commercial and criminal trends is an ever growing cost on individuals.

## 3 A FRAMEWORK

Our framework is built on the argument that the practicality and utility of any approach to materially improving privacy by increasing controls over personal information depends critically on how the approach addresses the issues of decisionmaking, negotiation, and enforcement. We begin in Section 3.1 with a general look at each of our critical issues. Throughout this section and the next, we use the term data subject to represent the individual whose personal information is being collected and used, and the term data user to represent the party that collects or uses that information. Sections 3.2 through 3.5 consider how each of these issues are handled by the four models of information privacy mentioned in the Introduction.

### 3.1 Crosscutting Issues

**Decisionmaking** asks who are the decisionmakers, and do they have the information and incentives to make good decisions about what personal information is worth protecting and controlling. This can be more complex than it sounds as many people have very different priorities with regard to what personal information they view as deserving of protection. Even where there is a reasonable amount of consensus (e.g., with health or financial data), subtleties may arise with personal information “owned” by a group of data subjects, such as those in a household or family. The issue is further complicated by the fact that innocuous data can sometimes imply more sensitive information. For instance, 87% of the US population is uniquely identified by birth date, gender and 5digit ZIP [33].

---

<sup>1</sup> We understand that it is difficult to draw a bright line between primary and secondary uses. We’ll revisit this in section 3.

**Negotiation** is the process by which data subjects and data users reach agreement on the rights and responsibilities of the data users with respect to the data subjects' personal information<sup>2</sup>.

Today, information is collected for myriad reasons as we go about our daily lives, but almost always for some specific primary use (e.g., to pay for a commercial product or to receive appropriate medical attention). This leads to several complications. First, it can be very hard to determine a priori when primary uses end and secondary uses begin. Second, it is natural to provide some amount of personal information during a commercial transaction and to collect that information in a manner separate from primary use would lead to inefficiencies in many of our daily activities. Finally, it is important to keep in mind that if personal information and policy choices are bundled with some other good or service—for instance, agree to this privacy policy or don't enter this hospital—then individuals have no choice and will feel that they have no control over their personal information.

The result of negotiation should be a clear statement of the personal information collected during a primary use and an enforceable contract describing the rights and responsibilities of the data user with respect to this personal information.

**Enforcement** is the mechanism or set of mechanisms that provide a guarantee that the data user abides by the negotiated rights.

We evaluate different enforcement mechanisms in terms of strength and transparency. By transparency we mean that data subjects should be able to see that an enforcement mechanism is effective. This could be something as simple as allowing individuals to see who had or still has access to their released personal information.

When considering information privacy alone, stronger guarantees are always better. In other words, the more difficult it is to circumvent the technology behind the enforcement or the more costly it is to get caught violating the negotiated rights, the more certain individuals can be that their rights are not being abused. We include auditing mechanisms in enforcement, as auditing helps catch and prosecute those violating the negotiated rights.

Overall, however, an implementer of a particular model for information privacy may not choose enforcement mechanisms with the strongest guarantees. From the point of view of society as a whole, the strongest mechanisms may too harshly diminish economic efficiencies, or they may require society to sacrifice other important goals. From an individual's point of view, one goal of information privacy is risk management, and weaker mechanisms may sufficiently mitigate the likely risks and thus provide reasonable levels of protection to the data subject.

Each of the models in the following sections can often be improved by the adoption of one or more Privacy Enhancing Technologies, which limit the amount of personal information disclosed during a transaction. Examples include anonymization tools [13], which allow individuals to engage in transactions without releasing as much information as they would otherwise, and database sanitation techniques [14, 32, 23], which ensure that the private information of an individual is not released from databases used for marketing or research purposes.

It is important, however, to recognize that these technologies are incomplete solutions to the problem of information privacy. Some data cannot be aggregated, sanitized, or anonymized, and at that point, trust is required. For that trust to exist, there needs to be mechanisms in place for decisionmaking, negotiation, and enforcement of privacy. It is the building of these mechanisms that concern us here. We consider Privacy Enhancing Technologies to be orthogonal to the models we discuss.

### 3.2 SelfRegulation

The model for improving privacy most heavily promoted by industry is selfregulation. Firms and industry groups argue that if consumers truly care about privacy, then privacy invasive practices will offend them and cause the reputations of offending firms to suffer. This sort of thinking has led to the observed "privacy paradox," in which consumers claim in surveys to care deeply about privacy, but their actions suggest otherwise.

In this view of the world, data belongs to those who collect and aggregate it—there is no true negotiation. The data is collected by a firm, usually for some primary use other than merely data collection and the rights to use the information for secondary purposes is bundled with the decision to engage in business with that firm. Proponents of self-regulation would argue that businesses can and should give notice of their business practices via a privacy policy and if individuals are dissatisfied with this, they can take their business elsewhere.

---

<sup>2</sup> For ease of exposition and without loss of generality, we will assume that the decision-making process has determined that all personal information is worth protecting and controlling.

Enforcement under this system is primarily through the reputation hits a company suffers when their poor data practices are exposed. A classic example of success in this area is in AT&T's advertisements stating that it would not use calling records to contact potential new customers, as MCI had done under its "Friends and Family" program [34]. In order to avoid a reputation hit, it is presumed that a firm must comply with its privacy policy. To this end, there is a growing industry of online risk management systems that help companies manage their databases and information transactions according to a stated policy.<sup>3</sup>

The problem with this is that, in practice, privacy policies make extremely poor signals [41], causing good decisionmaking on the part of the consumer to be almost impossible. First of all, due to liability concerns as well as the direct financial incentives in exploiting the personal information of individuals, privacy policies tend to be difficult and timeconsuming to read and understand. Secondly, these policies are subject to change at any time, particularly if the company in question changes ownership or has its information assets sold. Thirdly, the periodic exposure of bad practices—necessary as an enforcement policy—has led to a jaded public that is unlikely to believe claims of privacy protection from firms with any secondaryuse interest in the data. Lastly, because data practices are bundled with some good or service, data protection is not the primary concern in the buyer's mind—they are not buying privacy, they're buying some other thing. Often there are no good privacy choices that allow the consumer to acquire the good or service they want, so they settle for bad ones. These facts lead to a lemons market for privacy, in which no consumer will reward good practices—since consumers don't know about nor believe in them—so no firm has an incentive to create good privacy practices and all practices are bad.

### 3.3 Government Regulation

In this model, the government makes laws specifying how data holders can use personal information and when the consent of the data subject is necessary. There are numerous examples of government regulation. For example, in 1995, the European Union (EU) adopted a comprehensive set of regulations called the Data Protection Directive aimed at providing its citizens with consistent levels of protection and enabling the free flow of personal information within its member countries [29]. The Directive is focused mainly on the private sector, with exceptions for law enforcement. In the United States, on the other hand, regulation is more sectoral. Medical data, for example, is regulated under HIPAA, while data collected on children's web sites is regulated under COPA.

Under a model of government regulation, especially when such regulation is Draconian in nature, the need for negotiation is reduced. The government simply decrees how data can flow and what sort of consent from the data subject is needed.

As far as enforcement goes, the government makes use of law enforcement agencies, using traditional investigation methods to catch offenders and traditional legal punishments to deter them. There is no example of any governments using active enforcements, meaning there is nothing to stop anyone from breaking these laws except the fear of getting caught.

This sort of enforcement has many problems. One problem—as seen in the EU directive—is that information travels easily across borders and governmental data protection agencies can only protect data subjects and their personal information from abuse when such abuse occurs within the agencies' jurisdiction. Another complication is that the EU Directive is enforced differently by each of its member countries. Finally, any investigative, auditing data collected to enforce privacy regulations will end up in the hands of law enforcement—often the very institutions that most concern privacy advocates.

The government is the decision maker here, for better or worse. It decides what sort of data should be protected and how. The benefit of this is that the government ideally has the ability to gain a global picture of what is being done with data and make appropriate choices about when it needs to flow and when it does not. On the flip side, privacy choices are very individual and data that is personal to one individual may not matter at all to another. Choices made by the government for all of its citizens are bound to produce some inefficiency. In addition, the government may simply choose to subordinate the needs of individuals for privacy for the needs of businesses for consumer data or law enforcement for investigative data. Governments are not disinterested third parties, but are increasingly among the most avid consumers of data. Particularly in a post 9/11 world, governments encourage data collection in the private sector

---

<sup>3</sup> Though compliance concerns and reputation hits may be cited as reasons for purchasing online risk management systems, getting individuals to use online banking and other online services may turn out to be a more meaningful motivator in getting companies to create and abide by their own privacy policies [18].

so that law enforcement may access it in the course of an investigation.

### 3.4 Thirdparty Regulation

If governments are not appropriate as the disinterested third party and their ability to provide boundaryless protection is hampered by existing jurisdictions, an obvious alternative is to create a truly independent regulatory entity. The sole purpose of this entity is to act as an unbiased broker between data subjects and data users and as a facilitator of global policies for the protection of personal information. We consider two different examples of this model.

The first is online privacy seals, such as those offered by TRUSTe, WebTrust or BBBOnline [7]. Broadly stated, privacy seals are meant to raise the awareness of privacy issues in both data subjects and data users, and ideally to inform Internet users when a data user abides by an internationally recognized code of fair information practices [7].

Our second example of thirdparty regulation adapts the technology found in Digital Rights Management (DRM) systems, which is used to protect intellectual property, to the problem of information privacy protection. The particular proposal we discuss is called Privacy Rights Management (PRM) [20]. It uses various technical means to prevent data, once transferred, from being transferred further.

Online policy seals. The idea behind online privacy seals is that independent organizations assess the privacy practices of data users (e.g., commercial companies) that paid them to do so, and then award a distinctive Seal of Approval to those data users that meet a wellknown set of privacy standards. By prominently displaying the seal, companies could send a simple signal to consumers. Consumers would use this seal to make appropriate privacy choices and thus better protect their personal information.

Though online privacy seals seem to eliminate the need for negotiation and address the issue of decisionmaking directly, they have a serious problem as currently implemented. This problem is described by economists as capture. The third parties obtain their income from the firms that they approve; they thus have an incentive to make it easy for firms to use them. TRUSTe came under criticism for granting its seal to any firm that had a privacy policy and adhered to it, no matter how privacy invasive that firm's practices might be, and for never revoking the seal [4]. Since then TRUSTe has both created a privacy standard that a firm must meet in order to be awarded a seal, and TRUSTe has revoked its seal, though this is something they still do only very rarely [38, 5]. Until the incentives and business models of those entities offering online privacy seals better balance the rights of data subjects against the needs of the data users, these entities will not be truly independent third parties.

In terms of enforcement, a provider of an online privacy seal should perform regular audits of the privacy practices of the companies displaying its seal, and it should aid individuals in redressing misuses of personal information by those companies. The former is a potentially timeconsuming and laborintensive activity, though some technology is now available to help in this process. TRUSTe, for example, uses Watchfire [42], a tool that scans websites for privacy compliance issues, to do auditing and compliance testing on its customers' websites. In support of the latter obligation, a seal provider could launch an investigation and bring a lawsuit against a company displaying its seal, if it finds that the company breached the contract represented by the seal. To our knowledge, the most drastic action taken against a company that fails an audit is revocation of the seal.

Privacy Rights Management. PRM [20] is a comprehensive and technologydirected approach to thirdparty regulation. It proposes the creation of a data controller as a middle man between the data subjects and the data users (called data processors in their terminology). The data controller gathers, manages, and stores all personal information released by the data subjects. Data processors strike contracts with the data controller for use of the stored personal information. The data controller is responsible for ensuring that the data processors abide by the contracts and that the contracts reflect the privacy wishes of the data subjects.

By positioning the data controller between the data subjects and the data processors, negotiation occurs twice: once between the data subjects and the data controller, and then again between the data controller and the data processors. If we assume a mostly technologybased enforcement system, as is common in the DRM systems on which PRM was modeled, all three parties must be using the same or interoperable software systems, and the capability of the enforcement system will directly influence the type and richness of the negotiation. In existing DRM systems, content owners and content users select among predefined rights described using policy languages such as XrML, an extensible rights markup language for expressing the rights and conditions associated with digital content or services [11]. As they stand today, it is not practical for anyone but technical experts to negotiate directly in these policy languages.

The PRM system depends on technology for not only creating an audit stream, which could be used as in the other regulatory approaches to monitor and prosecute data processors that violate the assigned rights, but also for active enforcement. Rights management systems are built to prevent the data user from being able to abuse the protected information. This contrasts with the other regulatory approaches, which depend on legal means alone to prevent abuse. On the other hand, there are no perfectly secure rights management systems, and thus this approach should be backed by legal protections similar in nature to what is found in the other regulatory approaches. And, of course, there is the open research question of how best to preserve individual privacy when mining large databases [3].

In this model, decisionmaking is distributed. Designers of the PRM system decide the kinds of personal information that the system will manage and store. The designers also determine the protection choices available to the data subjects; one could imagine a system such as the AT&T Privacy Bird, used in P3P [12], where individuals select among a few general choices (e.g., high, medium, or low). The data subjects then choose the protection level associated with their personal information or with each category of their personal information.

### 3.5 Information Markets

As issues in electronic privacy began to garner significant attention, several economists proposed enshrining privacy rights in the hands of individuals [1, 21, 28, 40]. They reasoned that the problem with privacy in electronic markets today is that there exist third parties who collate and sell private information. Individuals suffer a cost from these transactions, but they do not get to participate in the market. This negative externality results in the privacy problems witnessed. Many individuals might be willing to give up certain information about themselves if they were compensated. In turn, the cost of using personal data would rise and there would be more privacy in society as a whole as firms would likely reduce the amount of information they collect. In order to facilitate the fair and legal transfer of information, economists proposed information markets.

The term “property rights” generally refers to two types of rights: possessory rights and rights of transfer [31]. Possessory rights are rights to use things and to prevent others from using them. The right of transfer is the right to give a possessory right to someone else. A distinct but similar right is that of the recipient to then transfer the possessory right further.

Kenneth Laudon was among the first to describe a market for private information [21]. He envisioned a national information market overseen by a Federal Information Commission (FIC) much like the U.S. Securities and Exchange Commission. People would retain property rights tied to an account number that could be managed by a local bank. The banks could aggregate information according to similar demographics to both aid in privacy and provide greater value. When anyone wanted to use the information for a secondary purpose (outside of the normal information required to do business) they would need to acquire the approval of the owner of that information (the identity it was tied to) and pay them a fee. The FIC would track the market and make sure that information was not transferred further or used for unapproved purposes.

Unfortunately, information markets as currently conceived have significant hurdles to overcome when viewed on the issues of negotiation, enforcement, and decisionmaking.

Information rights are negotiated in contracts between data users and data subjects, possibly with the help of a bank or information broker [21]. Questions arise, however, when one realizes that proponents of information markets generally make the assumption that all exchanges involving personal information flow through the market. How does this assumption affect personal information collected for primary uses—for example, a shipping address—or incidentally by an observer? To what extent can this data be used for secondary uses? Not at all? With the subject’s consent at the time of the primary use? And now we’re back to the question of whether information property rights can be bundled with some service, as is effectively done today with ecommerce web sites and privacy policies.

One might assume that the situation would be greatly simplified if the secondary uses were clearly decoupled from the purchase of products and services. One might provide financial remuneration to individuals who opt in and allow their personal information to be used for secondary purposes. Unfortunately, this is not straightforward to regulate, since it is difficult to distinguish between compensation for sharing information and pricing models that make it infeasible to maintain privacy.

The markets approach also assumes a change in the legal structure to give individuals property rights over their personal information. In particular, negotiated contracts ought to be legally enforceable. Furthermore, since each data

exchange is governed by a separate contract, the result is more complex than the previously described regulatory models.

Generally, individuals want to give away possessory rights to their personal information but not transfer rights, as in the case of intellectual property which is often licensed, but rarely sold outright. Most proponents of information markets assume some sort of agency, like Laudon's FIC, will provide a reasonable level of auditing to discourage and help prosecute abuses. No one discusses active enforcement mechanisms. In addition, because the information property rights are backed by governments, information markets share with government regulation the problem of national borders and how to pass information over them when protection regimes do not line up.

To create an information market, a government must first decide over which pieces of personal data a subject has property rights. Then, individuals must decide whether to sell or license those rights. This decision can be especially difficult when individuals lack the information needed to make appropriate choices about their privacy. Individuals often do not have a thorough picture of the costs of releasing certain pieces of information and how those pieces can be correlated with other previously released information. They may also have difficulty valuing privacy for psychological reasons [2]. These problems may be mitigated by information banks or brokers that would help individuals make better privacy choices. However, similar institutions would be of use to today, and they have not emerged. Those that have, such as TRUSTe, have the problem that they have been captured by the data users—the ones with the money [4]. It is not clear how such institutions would be more easily or effectively established in a markets approach.

## **4 INFORMATION MARKETS: WHY AND WHEN**

Taken pure and alone, all of the models in the previous section have severe limitations. We understand that most real world examples are—and real solutions are going to be—hybrids of these different approaches. The EU Data Protection Directive, often viewed as blanket regulation, leaves most data selfregulated. Firms are simply required to give notice of their practices and data subjects the right to object in certain situations. The marketbased solutions require government declaration of the property rights data subjects have in their personal information, quite possibly supported by international treaties. Rights management approaches would become stronger with legal support, and they require either selfregulation—firms decide to adopt the technology—or regulation—they are forced to.

Still, none of the existing models solve all of the problems involved in negotiation and enforcement. Clearly, new ideas and more research is needed. The question is on what model should we focus our efforts.

We are attracted to a marketbased solution not simply because we think it will improve information privacy by reducing the amount of personal information collected, but also because markets have the best potential of permitting individuals to choose what personal information is considered private (i.e., worthy of being controlled and protected). Though attaching monetary values to personal information may be a sufficient incentive for changing the behavior of businesses, we don't think it is the only incentive that matters for individuals.

Unfortunately, our framework has also lead us to believe that the obstacles facing information markets are more significant than those facing the various regulatory approaches. Though selfregulation has failed to create efficient and fair markets for personal information that respect the interests of individual consumers, the case study in Section 4.1 shows that markets cannot merely be legislated into existence either. Ambiguity in an information market will only cause such efforts to be ignored or resisted. We believe that an effective market requires institutions that trade information and actively enforce information rights. As Sections 4.2 and 4.3 discuss, building these institutions so that they are strong enough to supplant current datamining practices is an extremely difficult challenge.

### **4.1 Case Study: Oregon**

There has really only been one law that attempted to give individuals property rights to their personal information. This was an Oregon state law (ORS 677.097) passed in 1995. During the years that the law was on the books, it was never challenged in court. As a result, it did not provide any precedent regarding the legal issues discussed in the previous section. The relevant clause read, "An individual's genetic information and DNA sample are the property of the individual except when the information or sample is used in anonymous research (the identity of the person from whom the sample is derived cannot be determined) [16]." In 2001, Oregon Senate Bill 114 repealed the law, replacing it with penalties for misuse of DNA information. This decision was based on the recommendation of the Genetic Research Advisory Committee (GRAC), which consisted of representatives from the legislature, health care industry, pharma-

ceutical industry, and business and consumer affairs [30]. Genetic privacy is an interesting case to study, because there are clearly important privacy concerns involved, but also because a great public good can be accomplished by allowing this information to flow appropriately.

The GRAC cited three reasons why the property clause was included in the first place:

1. 1. It's a simple concept.
2. 2. It gives families ownership of the genetic material of a descendant.
3. 3. It provides families with protection from discrimination by providing them with standing for legal action.

The majority of the opposition to the law came from the drug industry. It is illustrative to look at the criticisms of the law.

- Obtaining consent from individuals for the use of their DNA is slow, costly, and sometimes impossible. As such it inhibits important genetic research.
- The law makes genetic privacy an alienable right which can be sold, leaving the individual with no recourse and no control over their information. Perhaps it is possible to license DNA to firms but not sell one's property rights completely, but there was no discussion of this and no mechanism for it in the law.
- No one knew how property rights should be obtained. Did a firm need to obtain consent from individuals to use their DNA? Did individuals whose DNA was used maintain an interest in any drug or treatment developed based on their genetic material?
- Genetic information cannot clearly belong to a single person, as blood relatives (especially identical twins) share that genetic information and therefore are each entitled to some control over it.
- The committee members felt that individuals were more interested in control of their DNA and protections against misuse than in monetary gain. As a result, data protection laws might provide adequate protection.

An important lesson from Oregon's experience is that property rights alone are not enough. Most of the criticism of the law amounted to opposition to its ambiguity. Data users did not know what their liability was and there was no market structure in which they could easily and unambiguously obtain the rights or licenses needed to legally do their research. Furthermore, there were no clear choices given to data subjects (e.g., to license their DNA anonymously for clinical research). The situation in Oregon demonstrates the need for a market structure in which property rights for DNA or any other regulated personal information can be meaningful.

As we discussed in Section 3, it must be clear what data are to be protected, what rights options are available, and how one may obtain these rights. The enforcement mechanisms and consequences must be known to both data subjects and data users.

## 4.2 Negotiation

As we said earlier, privacy means different things to different people. Individuals care about different pieces of information and have different levels of tolerance. Choice is a key component of negotiation.

Do we have to create an information market to provide such choice? Why not simply have the same rules (regulations) for everyone and provide choice through competition in the marketplace? For example, different firms could offer different policies for personal information and advertise this. Consumers would then select the firms that best match their preferences. Unfortunately, this is the status quo. Most companies do have a privacy policy listed, but it is usually unintelligible by the average user and subject to arbitrary change [41]. In addition, there is not often a good range of privacy policies available in a given good or service industry.

Another pressing problem is the bundling of personal information with the goods or services sold. From a privacy perspective, you would want the licensing of personal information to be entirely separate from the purchasing of goods and services. However, separating these transactions would be inefficient since merchants need to collect personal information in order to conduct business and much of the information today focuses on purchasing habits. Merchants will want to collect this information for their own internal records and for liability reasons. Given this information, it is only natural that they will want the right to do data mining on it and possibly sell it to third parties. So we must find a way for data users to negotiate these rights with data subjects that is efficient and acceptable to both parties.

Clearly, bundling is not a problem when pure information buyers want to negotiate information rights with individuals. These merchants could easily provide individuals with a rich set licensing choices, and may benefit in the information marketplace from creative compensation packages for more liberal uses of the personal information.

For both primary and secondary uses, it would be useful for the negotiated rights to be expressed in a machine readable language, such as P3P or EPAL. This would force the licensing choices to be explicit and have a universally understood meaning that could be interpreted by software on the individual's side. It would enable automatic negotiation between software that understood the individual's preferences with the firm purchasing that information, potentially solving a piece of the difficult primaryuse dilemma. By combining (directly or indirectly) the personal information with this machinereadable license in the data user's database, enforcement becomes much easier.

A system where privacy choices are encoded in machinereadable form, negotiated automatically between firms and consumers, and then built into the firm's database system is not beyond the reach of today's technology. However, this capability is not built into the offtheshelf components that most firms use for their online interactions. There are steps that have been taken: IBM's EPAL system helps companies enforce their privacy policies in their databases, and the P3P standard is a machine readable language for specifying privacy policies that can be interpreted by some browsers. However, no complete solution yet exists, and there isn't currently the motivation for one. While there aren't any aspects of the problem that make in intractable from a research perspective, it would require a lot of good engineering to get right. Interestingly, this problem is similar to the information sharing problems faced by law enforcement and intelligence communities that wish to share information, but are constrained by laws intended to protect privacy or separate powers.

## **4.3 Enforcement**

Perhaps a greater problem is the enforcement of the many unique agreements produced by a marketbased system. Generally, people want to give away possessory rights to their data but not transfer rights, as in the case of intellectual property which is often licensed, but rarely sold outright. We currently lack the technology to audit information flow and determine its misuse. And to ensure that we do not make the situation worse, all attempts to create an infrastructure to do this sort of auditing should avoid concentrating huge amounts of personal data in the hands of the auditors.

### **4.3.1 Technology Directions in Enforcement**

Information in the digital age is cheap and easy to copy. Controlling information property is analogous to controlling intellectual property, otherwise known as digital rights management (DRM) in the entertainment industry. One of the principal tools being researched by that industry is watermarking, along with fingerprinting and other information hiding techniques. Watermarking is the practice of embedding a mark into a piece of content that is difficult to remove and that can be used to track it [10]. Recent advances in watermarking technology have enabled traitor tracing techniques that customize information to the holder, thus enabling discovery of the "traitor" (i.e., data user) who illicitly leaked the material [19].

The DRM industry is also investigating the use of secure hardware that can certify the software on top of it and protect storage from the user. Information could be given to a data user with such a machine. The machine could then certify that the software would only use the protected information as licensed [24, 39]. With DRM, consumers fear that secure hardware may reduce their ability to use their computers as they please and to share copyrighted content within fairuse laws. It is not clear whether similar concerns with surface among data users.

As mentioned earlier, Korba and Kenny [20] discussed the direct application of DRM technology to privacy. This work considers data subjects in the role of rights holders and data holders as consumers. There are two main challenges in "porting" these technologies to privacy. First of all, there is their general immaturity—existing technology will not deter powerful adversaries, and data users almost certainly have more resources than most consumers. Secondly, there is the problem of scalability and heterogeneity. There are many more data subjects than rights holders in conglomerates like the RIAA or MPAA. And there is a question of whether data subjects would trust all their information to a single DRM company or system. Finally, since personal information has the potential to be more heterogeneous than current media files, the resulting protection system might need to be equally more complex.

However, despite their immaturity, these imperfect methods might effectively raise the cost of illicit transactions that violate privacy.

### 4.3.2 Policy Approaches to Enforcement

Not all enforcement concerns are amenable to technological solutions. For example, sensitive and personally identifiable items like social security and credit card numbers are surely easy to copy, as anyone who sees them can write them down on a piece of paper. One possible answer to this is to require information holders to have a license to possess that information [8]. Offenders can then be sued if they are found to illegally possess unlicensed information.

However, in order to pursue legal action, individuals will need to identify the offender. This is likely to prove difficult. The majority of victims of identity theft do not know how their information was lost. Often they do not even realize they are victims. If there is a strong probability of getting away with abuse, an information market will not be an effective mechanism for giving individuals control over their personal information. Strong auditing mechanisms are absolutely essential for the functioning of this market.

In general, it is probably impossible to prevent all abuses in a marketbased solution, and thus society might choose to limit them through other legal action. For example, a firm that abuses the private information of individuals may be subject to a class action lawsuit and forced to pay penalties. If these penalties were sufficiently high, Draconian measures to prevent abuse might be unnecessary.

## CONCLUSIONS AND FUTURE DIRECTIONS

As people progress through their lives, in this the 21st century, their actions, on or offline, are being recorded and stored in networked databases. While privacy enhancing technologies may staunch the flow somewhat [15, 17], ultimately, the data is not going to go away. The selfregulatory regime that has so far prevailed has left behind a confused and disgruntled public. As Harris Polls have discovered, we are quickly becoming a nation of “privacy pragmatists, who have strong feelings about privacy and are very concerned to protect themselves from the abuse or misuse of their personal information by companies or government agencies” [37]. If we as a society are serious about improving privacy, then something needs to be done and soon.

Towards this goal, we have presented a framework that can be used to evaluate different models for information privacy and better understand their utility. The framework stresses the need to address the issues of negotiation, enforcement, and decisionmaking that come with any data protection effort.

Adhering to the theory of second best [22], government regulation seems the most feasible interim approach. While governments may not be truly disinterested third parties, every model requires or directly benefits from some amount of government regulation. There doesn't seem to be any better place to start.

Implementing government regulation is not without challenges. Figuring out what sort of information should be protected and how much bundling of secondary uses with primary uses should be allowed is difficult. Auditing and enforcing these regulations even trickier. However, a compliance industry will likely spring up, as has happened in response to other regulations like HIPAA and the EU Data Protection Directive that might help in this matter.

While a marketbased approach might be more efficient and flexible than a regulatory one, it is significantly more complex. The Oregon case demonstrated that an ambiguous and immature dataaspropertyrights approach was less preferable to industry than straightforward regulation. Any effective market solution would require legislative support, definitions of protected information, and enforcement mechanisms—all of which must also be developed for a regulatory solution. However, the negotiation and enforcement mechanisms are significantly more complex in the markets case, mainly because of the many individual contracts that must be negotiated and enforced. Institutions need to be built where information rights can be traded under rational privacy choices. As realists, we must admit that this sort of approach is not feasible in the near future.

In the meantime, there are a number of research problems—really hard problems—that we as the academic community ought to work towards solving. In the technical realm, we can work towards improving our ability to enforce privacy rules and audit privacy practices, especially across administrative domains. In the economic realm, we can look at issues of bundling and figure out how to allow individuals to make real choices about their personal information as they go about their lives.

If we can manage to make significant advances in these fields, then perhaps, in time, we will be able to revisit the idea of more flexible, marketbased approaches and come to a more positive conclusion. In the interim, we suggest research and regulation.

## ACKNOWLEDGEMENTS

Thanks to Roger Dingledine, Geoff Goodell, Allen Friedman, Glenn Holloway, H.T. Kung, Greg Morrisett, David Malan, David Parkes, and Stuart Schechter for their helpful suggestions and support. This work has been funded in part by NSF Trusted Computing grant CCR-0310877 and by gifts from Microsoft.

Rachel Greenstadt was supported in part by a U.S. Department of Homeland Security (DHS) Fellowship, a program administered by the Oak Ridge Institute for Science and Education (ORISE). ORISE is managed by Oak Ridge Associated Universities under DOE contract number DEAC0500OR22750.

## References

- [1] E. Adar and B. Huberman. "A Market for Secrets," *First Monday* 6(8), August 6, 2001.  
<http://www.firstmonday.org/issues/issue6.8/adar/>
- [2] A. Acquisti and J. Grossklags. "Privacy and Rationality in Individual Decision Making," *Proceedings of the Third Annual Workshop on the Economics of Information Security (WEIS04)*, May 2004.
- [3] R. Agrawal and R. Srikant. "PrivacyPreserving Data Mining," *Proceedings of the ACM SIGMOD Conference on Management of Data*, pp. 439450, May 2000.
- [4] P. Boutin. "Just How Trusty Is Truste?" *Wired News*, April 9, 2002.  
[http://www.wired.com/news/exec/0,1370,51624,00.html?tw=wn\\_story\\_related](http://www.wired.com/news/exec/0,1370,51624,00.html?tw=wn_story_related)
- [5] T. Bridis. "Privacyassurance Seal Yanked from FreeiPods.com Web Site," *The Detroit News*, February 10, 2005. <http://www.detroitnews.com/2005/technology/0502/11/tech85504.htm>
- [6] BusinessWeek Online. "Privacy on the Net," *Business Week/Harris Poll: A Growing Threat*, March 20, 2000.  
<http://businessweek.com/2000/00.12/b3673006.htm>
- [7] A. Cavoukian and M. Crompton. "Web Seals: A Review of Online Privacy Programs," *22nd International Conference on Privacy and Personal Data Protection*, Venice, September 2000.  
<http://www.privacy.gov.au/publications/seals.pdf>
- [8] S. Cha and Y. Joung. "From P3P to Data Licenses," *Proceedings of the Workshop on Privacy Enhancing Technologies (PET 2003)*, March 2003. <http://petworkshop.org/2003/preproc/14preproc.ps>
- [9] ChoicePoint home page. <http://www.choicepoint.com>
- [10] C. Collberg and C. Thomborson. "Watermarking, TamperProofing, and Obfuscation: Tools for Software Protectionx," *IEEE Transactions on Software Engineering* 28(8):735746, August 2002.
- [11] ContentGuard. "XrML The Digital Rights Language for Trusted Content and Services," *XrML web site*, 2005.  
<http://www.xrml.org/about.asp>
- [12] L. Cranor, M. Arjula, and P. Guduru. "Use of a P3P User Agent by Early Adopters," *Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society*, pp. 110, November 2002.
- [13] R. Dingledine and N. Mathewson. "Anonymity bibliography," *The Freehaven Project*.  
<http://www.freehaven.net/anonbib/date.html>
- [14] S. Chawla, et. al. "Toward Privacy in Public Databases," *EU Workshop on Multiparty Protocols*, 2004, pp. 363-385.
- [15] J. Feigenbaum, M. Freedman, T. Sander, and A. Shostack. "Economic Barriers to the Deployment of Existing Privacy Technologies," *Proceedings of the Workshop on Economics and Information Security*, May 1617,

2002.

- [16] geneforum.org. "Genetic Privacy: Oregon Genetic Privacy Statutes."  
[http://www.geneforum.org/learnmore/gp/or\\_gps.cfm](http://www.geneforum.org/learnmore/gp/or_gps.cfm)
- [17] I. Goldberg. "Privacyenhancing technologies for the Internet, II, Five years later," Proceedings of the Workshop on Privacy Enhancing Technologies, LNCS 2009, April 2002.
- [18] D.Kawamoto. "Analyst: Hidden Costs in Security Breaches," CNET News.com, March 1, 2005.  
[http://news.com.com/Analyst+Hidden+costs+in+security+breaches/21001029\\_35595312.html](http://news.com.com/Analyst+Hidden+costs+in+security+breaches/21001029_35595312.html)
- [19] A. Kiayias and M. Yung. "Breaking and Repairing Asymmetric PublicKey Traitor Tracing," Security and Privacy in Digital Rights Management, ACM CCS9 Workshop (DRM 2002), pp. 3250, November 2002.
- [20] L. Korba and S. Kenny. "Towards Meeting the Privacy Challenge: Adapting DRM," Security and Privacy in Digital Rights Management, ACM CCS9 Workshop (DRM 2002), pp. 118136, November 2002.
- [21] K. Laudon. "Markets and Privacy," Communications of the ACM, 39(9):92104, Sept. 1996.
- [22] R. G. Lipsey and K. Lancaster. "The General Theory of Second Best," Review of Economic Studies, vol. XXIV (October, 1956), pp. 1132.
- [23] A. Meyerson and R. Williams, General kanonymization is Hard. Carnegie Mellon School of Computer Science Tech Report,2003; 03113.
- [24] Microsoft, "Microsoft NextGeneration Secure Computing Base Technical FAQ," July 2003.  
<http://www.microsoft.com/technet/archive/security/news/ngscb.mspix>
- [25] D. Mulligan, A. Cavoukian, A. Schwartz, and M. Gurski. "P3P and Privacy: An Update for the Privacy Community," March 28, 2000. <http://www.cdt.org/privacy/pet/p3pprivacy.shtml>
- [26] R. O'Harrow, Jr. "In Age of Security, Firm Mines Wealth of Personal Data," The Washington Post, Jan. 20, 2005. <http://www.washingtonpost.com/wpdyn/articles/A222692005Jan19.html>
- [27] R. O'Harrow, Jr. "ID Data Conned from Firm," The Washington Post, Feb. 17, 2005.  
<http://www.washingtonpost.com/wpdyn/articles/A308972005Feb16.html>
- [28] D. Parkes. "Challenge Problem: AgentMediated Decentralized Information Mechanisms," Agentcities: Challenges in Open Agent Environments, Burg et al. (eds.), SpringerVerlag, 2003.
- [29] M. Rotenberg and C. Laurant. Privacy and Human Rights 2004: An International Survey of Privacy Laws and Developments, published by Privacy International (London, UK) and Electronic Privacy Information Center (Washington, DC), November 17, 2004. <http://www.privacyinternational.org/survey/phr2004>
- [30] J. Santa and B. Speight. "Assuring Genetic Privacy in Oregon," A 2001 Report of the Genetic Research Advisory Committee (GRAC), November 15, 2000. <http://www.dhs.state.or.us/publichealth/genetics/docs/gracrpt.pdf>
- [31] S. Shavell. Foundations of Economic Analysis of Law (Chapter 7: Property Rights in Information), Belknap Press of Harvard University Press, 2004.
- [32] L. Sweeney, "kanonymity: a model for protecting privacy." International Journal on Uncertainty, Fuzziness and Knowledgebased Systems, 10 (5), 2002; 557570.
- [33] L. Sweeney, "Weaving technology and policy together to maintain confidentiality." Journal of Law, Medicine and Ethics. 1997. 25:98110.
- [34] P. Swire. "Markets, SelfRegulation, and Government Enforcement in Protection of Personal Information," Privacy and SelfRegulation in the Information Age (Chapter 1.A), Report of the National Telecommunications and Information Administration (NTIA), June 1997. [http://www.ntia.doc.gov/reports/privacy/privacy\\_rpt.htm](http://www.ntia.doc.gov/reports/privacy/privacy_rpt.htm)

- [35] Z. Tang, Y. Hu and M. Smith, "Protecting Online Privacy: SelfRegulation, Mandatory Standards, or Caveat Emptor," Workshop on Economics and Information Security, June 2005.
- [36] H. Tavani. Ethics & Technology, Wiley and Sons, 2004.
- [37] H. Taylor. "Most People Are 'Privacy Pragmatists' Who, While Concerned about Privacy, Will Sometimes Trade It Off for Other Benefits," Harris Poll #17, March 19, 2003. [http://www.harrisinteractive.com/harris\\_poll/index.asp?PID=365](http://www.harrisinteractive.com/harris_poll/index.asp?PID=365)
- [38] P. Thibodeau. "Truste Tightens Requirements for Its Seal of Approval," ComputerWorld, December 11, 2002. <http://www.computerworld.com/securitytopics/security/privacy/story/0,10801,76650,00.html>
- [39] Trusted Computing Group home page. <http://www.trustedcomputinggroup.org>
- [40] H. Varian. "Economic Aspects of Personal Privacy," Privacy and SelfRegulation in the Information Age (Chapter 1.C), Report of the National Telecommunications and Information Administration (NTIA), June 1997. [http://www.ntia.doc.gov/reports/privacy/privacy\\_rpt.htm](http://www.ntia.doc.gov/reports/privacy/privacy_rpt.htm)
- [41] T. Vila, R. Greenstadt, and D. Molnar. "Why We Can't Be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market," Workshop on Economics and Information Security, 2003. <http://www.cpppe.umd.edu/rhsmith3/>
- [42] Watchfire. "TRUSTe Bolsters Privacy Certification and Seal Program Enforcement with Watchfire(R) WebXM(TM)," August 27, 2002. <http://www.watchfire.com/news/releases/82702.aspx>