December 2007

A Personal Information Auction: Measuring the Differential Value of Privacy

Daniel Rice
Loyola College in Maryland

John D'Arcy
Towson University

A PERSONAL INFORMATION AUCTION: MEASURING THE DIFFERENTIAL VALUE OF PRIVACY

Daniel O. Rice
Loyola College in Maryland,
The Sellinger School of Business and Management
drice2@loyola.edu

Abstract:

This paper proposes a research methodology to studying the differential value that individuals place on their own personal information with reference to data protection, security, and markets for privacy. Gopal et al. [2006] shows that there is a technological solution, the CVC-STAR approach, for the differential protection of numerical data in a market for personal information. Differential protection enables customized levels of privacy for individuals' personal information in lieu of the currently predominant "all-or-nothing" approaches. This research will show that differential privacy fundamentally affects individuals' willingness to contribute personal information and the value that they put on the protection of their personal information and privacy.

Keywords: privacy, data privacy, personal information, data protection, information market, value of personal information.

Introduction and Literature Review

Recent advances information technologies that enable the collection, storage and analysis of personal data have increased the ease with which businesses can make profitable use of information about individuals. While, businesses are increasingly aware of the value created through the use of accurate information to make informed business decisions about products and customers, individuals are becoming more aware of the value of their own personal information and are also concerned over how the loss of control over this information impacts their personal privacy (Harris and Westin [1995, 1996, 2002]). Today more than ever there is an imperative need to balance the collection and use of personal information with individuals' right to keep personal information private. The convergence of increased technological capability, business motives, and individual's privacy awareness has created conditions for a privacy "perfect storm" and the creation of markets for the trade of personal information.

The emergence of markets for personal information has been addressed in information systems, economics, and law literature. Laudon [1996] foretells a market where individuals' personal information is traded and where individuals maintain control over their information. The National Telecommunications and Information Administration (NTIA), of the U.S. Department of Commerce, investigate the privacy and self-regulation through markets for privacy (NTIA [1997]). Varian [2006] introduces a view of personal privacy as the voluntary exchange of individuals' personal information between parties rather than as a security issue. Varian [2006] also provides a simple example that shows how personal information could be used in economic transactions and points out that there are advantages to making personal information available. Cate [2001] lauds a self-regulated market that is "more flexible and more sensitive to specific contexts and therefore allow individual to determine a more tailored balance between information uses and privacy than privacy laws do". Gopal et al. [2005] develop and test a personal information market implemented using a flexible data protection model in where individuals can choose how much privacy protection they require (protection levels) and can also choose the level of compensation they require for the disclosure of their personal information at that protection level.

The intellectual arguments against the self-regulation of privacy through markets for private information are technological, economic, and legal in nature. The technological arguments point to current inadequate technologies, especially deficiencies in those technologies needed to manage and control the inherent complexities of such a market and businesses' current predisposition to an "all-or-nothing" approach (I take the liberty here to discuss this as a technological limitation). Economic arguments focus on the inability to accurately determine the "value of personal information" and the legal arguments are mostly concerned with the property rights of personal information and the inadequacy of contractual law to enforce the contractual agreements.

In this research, we intend to address the above issues and acknowledge that this is done with deliberate specificity. By citing Gopal et al. [2005] we introduce a candidate technology, that coupled with other existing privacy enhancing technologies, provides a potential candidate that overcomes some of the technological challenges facing privacy markets (namely, the management of complexities and a tailored solution to privacy over the "all-or-nothing" privacy solution). We intend to test a market where individuals control their privacy level and intend to determine the differential value individuals place on their personal information in a market where they individuals control how much personal information they trade for compensation.

The Trading of Personal Information

Anecdotal evidence suggests that individuals are willing to trade personal information for compensation (Chang et al. [1999]). Empirical and experimental evidence shows that there is a dollar value individuals put on the protection of their personal information (Chellappa and Sin [2005] and Hann et al. [2002]). However, these previous studies have examined only a “reveal / don’t reveal” disclosure condition for individuals’ personal privacy. I propose to extend previous work by determining the differential value of privacy; that is, the value that individuals put on differing levels of disclosure of personal information. I call this the “differential value of privacy,” and research into this area will significantly contribute to privacy protection research. This research is likely to have an impact on business because it explores a new and practical approach to the protection and availability of consumers’ personal information for use in business intelligence (BI) applications.

Individuals are becoming more aware of the value of their personal information, and they are concerned about how the loss of control over this information impacts their personal privacy (Harris and Westin [1995 and 1996] and Harris Interactive Poll [2002]). In fact, a survey found that 80% of Americans felt that "consumers have lost all control over how personal information about them is circulated and used by companies” (Harris and Westin [1995]).

Fueled by an increased concern about privacy, researchers have continued to study ways that personal information should, and can, be protected. The privacy protection literature is extensive and examines various dimensions of personal privacy including the impact of technology on privacy, the trade-off between information availability and privacy, the self-regulation of markets where private information is traded, and various techniques for the protection of privacy. (Agre and Rotenberg [1997], EPIC [2002], Laudon [1996] and Milne [2000])

In Gopal, et al. [2005] a sustainable personal information market where private information is made available and simultaneously protected. Gopal et al. [2005] assumes that subjects may be willing to share their personal information, either partially or completely, and develop protection and compensation models that allow for differential protection while providing incentives for participation in the market. This research investigates the differential value that individuals place on the protection of their personal information.

The Value of Privacy

The value of information has been described as the benefit derived from use of the information while the value of privacy refers to the value of keeping personal information private. There is undoubtedly an economic tradeoff between the availability of information and privacy. Recently, many information systems researchers have addressed this tradeoff (for a list of works see the website The Economics of Privacy at http://www.heinz.cmu.edu/~acquisti/economics-privacy.htm#new which is maintained by Alessandro Acquisti at CMU). In a market for personal information a buyer pays for information and the subject contributing the information is compensated (Laudon [1996], Varian [2000]). Chang et al. [1999] explores the willingness of individuals to disclose personal data on the Internet for compensation. They explain that there are many recent examples of firms doling out "freebies", such as free internet access, e-mail service, and even computers, in exchange for the right to collect and use consumers' personal information. More recently others study the dollar value individuals place on privacy (Chellappa and Sin [2005], Hann et al. [2002]). In Huberman et al. [2005] the researchers conduct auction experiments designed to elicit the value people place on their personal information. To date, however, researchers have studied the value of privacy from an "all or nothing" point of view (total disclosure or total protection). This paper will extend earlier research by studying the differential value of privacy, where individuals choose their own protection levels and compensation prices. This is critical research for the further development of markets for personal information.

Statistical Database Security and Interval Protection

Data protection techniques have been developed to help balance the access to individuals' personal information with security and control over the dissemination of the information. Specifically, statistical database (SDB) security techniques have been developed to provide protect individuals from disclosure of private information from analytic information disseminated to SDB users (Castana et al. [1995]). Recently, a special technique for protecting personal numerical data has been designed such that individuals gain more control over the dissemination of their personal information (Gopal et al. [2006]).

Statistical database (SDB) security is at the core of protecting personal numerical data stored in databases. The research of SDB is extensive and can be broken down into several accepted techniques including perturbation, query restriction and confidentiality via camouflage (CVC). Adam and Wortman [1989] and Gopal et al. [2002] provide excellent overviews of existing SDB protection. Interval protection is a feature of the CVC-STAR protection model that allows differential disclosure while maintaining security (Gopal et al. [2005]). For example, consider a single numerical point of personal information such as a person's weight is 119 pounds. This single datum is the personal information to be protected. CVC-STAR can allows for different levels of protection for a subject's actual weight by guaranteeing that no one will be able to determine that the value of the subject's weight is within the bounds of [90 , 125] (total protection), [104.5 , 122] (50% protection), or [119 , 119] (no protection). This is assuming that this subject assumes a range of [90,125] to be total protection. For example, at a 50% protection level a subject is confident that all a users of the system will know is that their weight is somewhere between 104.5 and 122 pounds. (See Gopal et al. [2005] for details of the mechanism that allows for the guarantee of intervals of protection).

Hypotheses, Experimental Design and Methodology

Hypotheses

The central hypothesis is that individuals will choose lower protection levels, and demand lower compensation payments, for the disclosure of personal information to a group when they perceive their information to be either typical, or extremely atypical, for that group. (e.g., if I am of "normal" weight for a group, or extremely overweight compared to the group norm, I will be more willing to share information about my weight with that group.) Huberman et al. [2005] show that individuals are more willing to disclose personal information to a group when they feel their information is either "typical" or extremely "atypical" for that group. Their experiment tests this when the individual has only one disclosure option, to reveal or not to reveal the information. This can be broken down into the following hypotheses which can be tested:

Hypothesis 1 – Price is related to subjects' perception about their own weight (perception is measured as "somewhat underweight", "average", "somewhat overweight", and "overweight", ranging 1-4, respectively).

Hypothesis 2 – Price is related to subjects' body mass index (BMI=weight/height2). BMI scales weight indicating overweight/underweight determination.

Hypothesis 3 – Protection level (%) is related to subjects' perception about their own weight.

Hypothesis 4 – Protection level is related to subjects' BMI.

Hypothesis 5 – Differential price (calculated as protection level * price) is related to subjects' perception about their own weight.

Hypothesis 6 – Differential price is related to subjects' BMI.

Naturally, other interesting questions are likely to arise and to be studied such as the affects of gender and demographics. (e.g., are males less sensitive to disclosure of weight information than females and are there differences in demanded prices (or protection levels) between subjects groups?)

Experimental Design

The population to be studied includes the membership of three local fitness centers (approximate total membership 1200 persons) and the membership of a college fitness center. The two fitness center populations are expected to be very different, adding robustness to the data set.

Subjects are informed that the will be participating in a reverse second price auction for personal data. This is in the spirit of Huberman et al. [2005]. Auction Theory by Vijay Krishna [2002] discusses details of second-price auctions. Essentially the competitive nature and dynamics of the second price auction help promote a truer revelation of value. A nominal payment ($10 in cash or a $10 gift certificate to the college book store) will be provided as incentive to participate in the study, plus the chance to earn cash in the auction. Private information is treated as a "real good" in the experiment. West [2000] and Rose [1999] discuss the consideration of private information as a "real good". We offer to buy height, weight, and age information and inform the participants that the winner of the auction will be paid the second lowest price attained in the auction in exchange for the posting of on a website or publishing it in a newsletter. A reverse second price auction is used because it is sealed bid, ensuring privacy, and because the dynamics and competitive nature of the auction have been shown to elicit true valuation (Krishna [2002]). The subjects will access a PC based auction site graphic user interface (GUI) to enter information including their height, weight, age, price to disclose the information, and protection level. They will also respond to a short questionnaire.

Methodology

Gender, weight, height, age, questionnaire responses, the required level of protection and price information will be collected. Weight and age are considered to be the private personal information. Subjects will be given a randomly generated identifier and we will keep no records that link the individuals to their numbers. All of the private data involved in this study is considered to be non-threatening in terms of risk for identity theft crimes. However, weight and age information should be private enough that subjects may not want disclose it publicly. Input data will be verified using a scale for weight and the subject drivers' licenses for height and age information.

The above hypotheses will be tested using standard data analysis techniques. Price and protection level are treated as dependent variables (DVs). Age, weight, and height are treated as independent variables (IVs). BMI will be calculated. I will specifically determine the relationship and significance of (1) price and subjects' perception of their own weight, (2) protection level and subjects' perception of their weight, (3) price and subjects' BMI, (4) protection level and BMI, (5) differential price and subjects' perception of weight, and (6) differential price and subjects' BMI. Standard multiple (or sequential) regression techniques will be used to test for multiple correlation, R. Canonical regression may be used if the DVs are tested together. Differences between groups (between males and females and between fitness center subject groups) will be tested using a standard one-way ANOVA and *t*-test (or, a factorial ANCOVA or MANCOVA test). The ultimate decision on how to analyze the data will depend on how the DVs and IVs are defined in the model.

Data Analysis and Results

The results for this research in progress will be based on primary data collection and the auction experiment. Sample data has been simulated and is presented to illustrate some analysis concepts presented in the previous section. First, Figure 1 below shows sample data collected from the Healthy Weight Loss Clinic which is part of a leading general healthcare system weight management program. These individuals volunteered their information freely as participants of the program. (from http://www.mgh.org/weight/lose_it.html).

Name	Weight	BMI	Height	Age
Anna G.	259.25	40.6	5ft 7in	38
Steen W.	27	31.1	5ft 10in	49
Pam C.	223.5	39.6	5ft 3in	50
Bill M.	312.75	47.5	5ft 8in	67
Sandra L.	19	28.5	5ft 9in	35
Pam E.	255.5	45.2	5ft 3in	37
Sandra M.	16	31.2	5ft 1.5in	53
Jackie J.	19	33	5ft 4in	47
Bill D.	21	32.2	5ft 11in	57
Doug R.	28	38.1	6 ft	50
Lynn M.	19	31.9	5ft 5 in	49

Figure 1. Sample Confidential Information

Based on this sample data, we have simulated 720 observations, in order to simulate some of the relationships we will be tracking in this research. The first 14 simulated observations are given as Figure 2.

Observation	Weight	Height	BMI	Sex	Protection	Price
1	222	76.7	26.53	F	23%	9
2	188	64.5	31.77	M	39%	17
3	203	67.7	31.14	F	38%	12
4	207	60.9	39.24	F	33%	17
5	202	70.7	28.41	F	31%	16
6	173	54.8	40.5	F	40%	15
7	198	58.8	40.26	M	54%	16
8	216	70.6	30.46	F	0%	9
9	176	57	38.08	F	36%	19
10	220	74.9	27.57	M	28%	12
11	194	74.7	24.44	F	42%	13
12	201	70	28.84	M	45%	15
13	215	62.7	38.45	M	59%	16
14	178	52.1	46.1	F	45%	13

Figure 2. The first 14 of 720 simulated observations.

Relationships we will be tracking in this experiment to include the following:

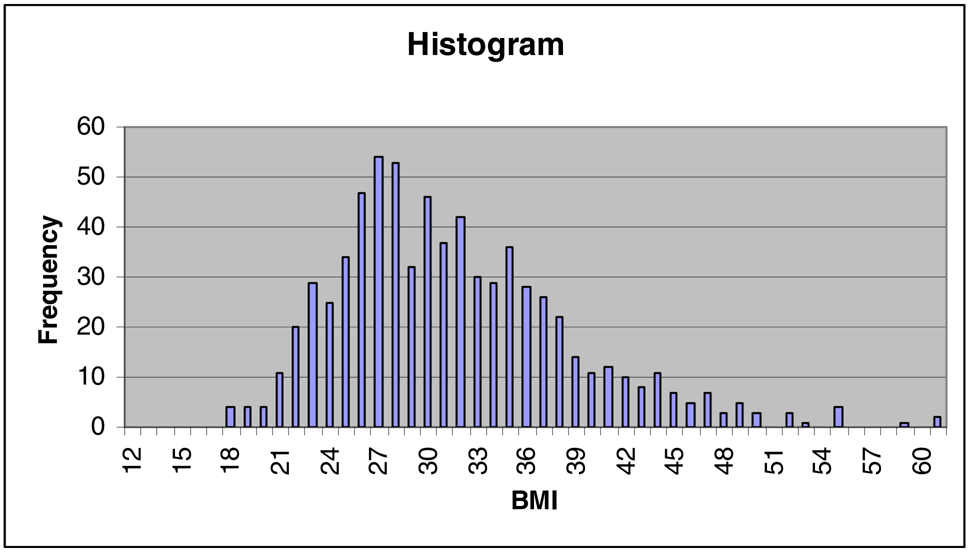

Figure 3. BMI Distribution Histogram (from simulated data)

The distribution of BMI, as shown in Figure 3, as well as other independent variables is of interest for the testing normality assumptions as part of the statistical analysis.

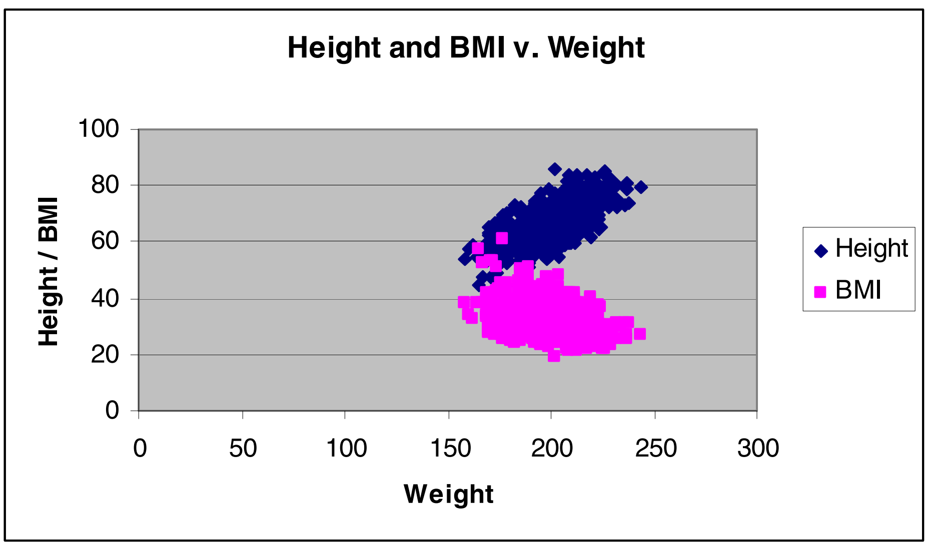

Figure 4. Height and BMI v. Weight

Figure 4 shows the relationships between height, weight and BMI.

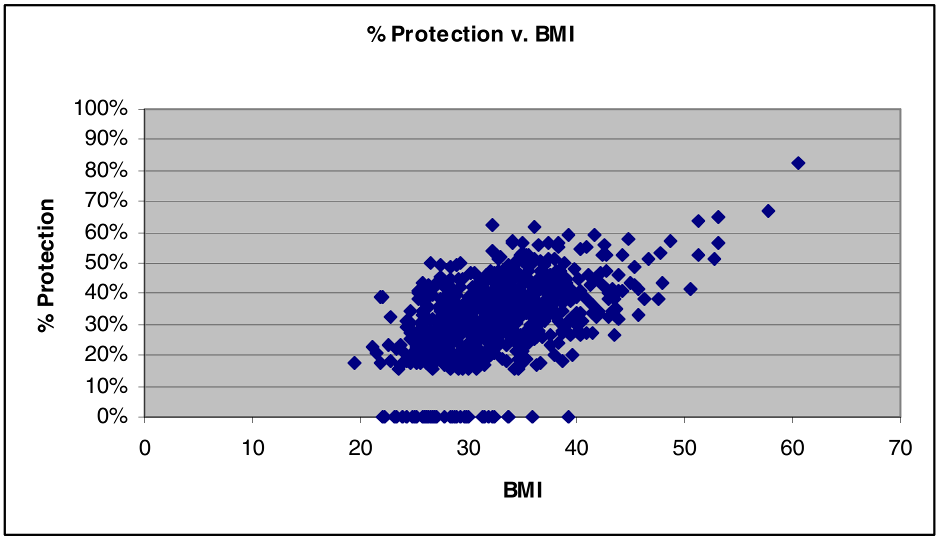

Figure 5. % Protection v. BMI

The percent protection chosen versus BMI shows the relationship between observations' BMI and their chosen level of protection.

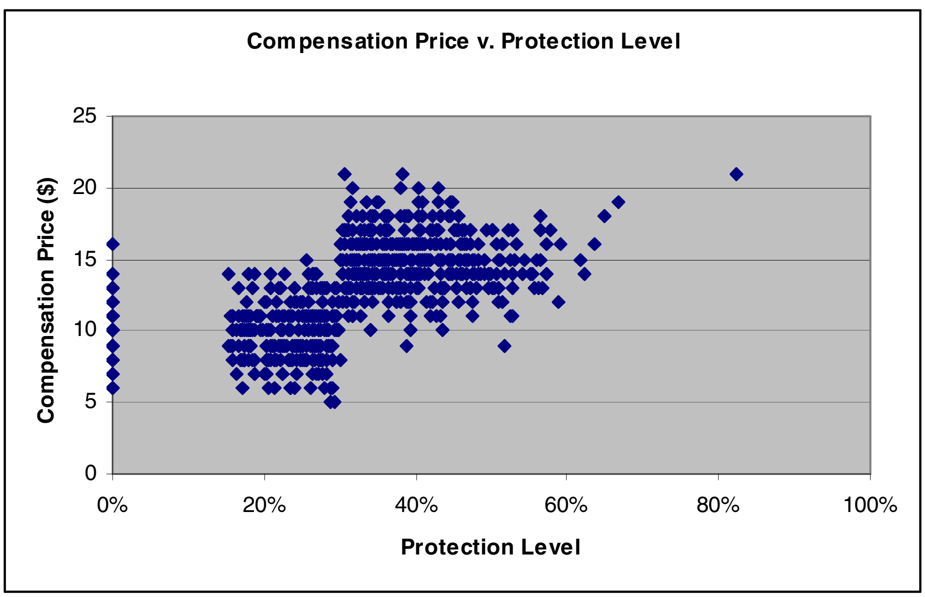

Figure 6. Compensation Price v. Protection Level

The relationship between compensation price and protection level is important because this will show the experimental value of privacy. In Fig. 6 above, it appears that individuals require more compensation as they require higher levels of protection. This will have important pricing implications for personal information markets.

Conclusions and Continuing Research

This paper proposes a research methodology to studying the differential value that individuals place on their own personal information with reference to data protection, security, and markets for privacy. Gopal et al. [2006] shows that there is a technological solution, the CVC-STAR approach, that enables differential protection of numerical data in a market for personal information. The possibility of differential protection allows for a different look via "tailored" levels of privacy versus previous the predominant "all-or-nothing" approaches used today. We intend to show that differential privacy fundamentally affects an individual's willingness to contribute personal information and the value that they put on their privacy.

REFERENCES

N. Adam and J. Wortman, "Security control mechanisms for Statistical databases: A comparative study," ACM Computing Surveys, 21(4):515--556, December 1989.

P. Agre and M. Rotenberg, editors, Technology and Privacy: The New Landscape, MIT Press, United States, 1997

S. Castana, G.F. Fugini, G. Martella, and P. Samarati, , Database Security, ACM Press published by Addison Wesley Longman Limited, England (1995)

F. Cate, Privacy in Perspective, the American Enterprise Institute, February 25 2001

R. Chellappa and R. Sin, "Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma", Information Technology and Management, 6, 181-202, 2005

A. Chang,P. K. Kannan, and A. Whinston,"The economics of freebies in exchange for consumer information on the Internet: An exploratory study," International Journal of Electronic Commerce, 4(1):85-102 , Fall 1999.

Electronic Privacy Information Center and Privacy International, Privacy and Human Rights 2004: An International Survey of Privacy Laws and Developments, 1st Edition, 2004.

R. Gopal, R. Garfinkel, D. Rice, and M. Nunez, "Secure Electronic Markets for Private Information", IEEE Transactions on Systems, Man, and Cybernetics, Part A, IEEE Transactions on Systems Man and Cybernetics, Part A - Special Issue on Secure Knowledge Management, Vol. 36, Issue 3, May 2006, 461-472

R. Gopal, P. Goes, and R. Garfinkel, "Confidentiality via camouflage: The CVC approach to database security," Operations Research, 50(3):501--516, May-June 2002.

I. Hann, K. Hui, T. Lee, and I.P.L. Png, "The Value of Online Information Privacy: Evidence from the U.S.A. and Singapore", accessed online at Harris Interactive Poll, "Privacy on and Off the Internet: What consumers want," February 20, 2002. [Online] Available: http://www.harrisinteractive.com/news

L. Harris and A. Westin, Consumer Privacy Survey, Harris-Equifax, Atlanta, GA, 1995.

L. Harris and A. Westin, Consumer Privacy Survey, Harris-Equifax, Atlanta, GA, 1996.

B. Huberman, E. Adar, and L. Fine, "Valuating Privacy", IEEE Security & Privacy, 3(5):22-25, September / October 2004

K. Laudon, "Markets and privacy" Communications of the ACM, 39(9):92--104, September 1996.

G. Milne, "Privacy and ethical issues in database/interactive marketing and public policy: A research framework and overview of the special issues," Journal of Public Policy & Marketing, American Marketing Association, 19(1):1--6, Spring 2000.

Privacy and Self-regulation in the Information Age, U.S. Department of Commerce, http:// www.ntia.doc.gov/ reports/ privacy/ privacy_rpt.htm accessed on 8 April 2004.

F. Rose, The Economics, Concept, and Design of Information Intermediaries: A Theoretic Approach, Physica-Verlag Heidelberg, Germany (1999)

D. Solove, The Digital Person: Technology and Privacy in the Information Age, New York University Press, 2004

H.R. Varian, Economic Aspects of Personal Privacy, from Privacy and Self-Regulation in the Information Age, Department of Commerce, accessed at http://www.sims.berkeley.edu/~hal/Papers/privacy/, accessed 23 October 2005.

L. West, Private markets for public goods: pricing strategies of online database vendors. Journal of Management Information Systems, 17, 1 (Summer 2000) 59-86