Abelson 06/08 4/15/08 4:11 PM Page 1

Information Accountability

By Daniel J. Weitzner, Harold Abelson,
Tim Berners-Lee, Joan Feigenbaum,
James Hendler, and Gerald Jay Sussman

With access control and encryption no longer able to protect privacy, laws and systems are needed that hold people accountable for the misuse of personal information, whether public or secret.

(footnotes)

[1]There are numerous definitions for privacy. Our chief interest here is understanding privacy rights as they relate to the collection and use of personal information, as opposed to other privacy protections that seek to preserve control over, say, one's bodily **one's physical??** integrity.

[2]See the authors' technical report; ~~dspace.mit.edu/bitstream~~/1721.1/37600~~/2MIT-CSAIL-TR-2007-034.pdf.~~

Existing legal and technical mechanisms intended to protect our privacy, copyright, and other important values have been overwhelmed by the increasingly open information environment in which we live. These threats follow from the ease of information storage, transportation, aggregation, and analysis. We must therefore rethink our approach to protecting our rights to be sure that the technical laws spelled out by Gordon Moore and Robert Metcalfe don't permanently overwhelm our values as enshrined in society's laws.

For too long, our approach to information-protection policy has been to seek ways to prevent information from "escaping" beyond appropriate boundaries, then wring our hands when it inevitably does. This hide-it-or-lose-it perspective dominates technical and public-policy approaches to fundamental social questions of online privacy, copyright, and surveillance. Yet it is increasingly inadequate for a connected world where information is easily copied and aggregated and where automated correlations and inferences across multiple databases uncover **routinely expose??** information even when it is not explicitly revealed. As an alternative, accountability must become a primary ~~means~~ through which society addresses issues of appropriate use. Information accountability means that such use should be **must be made??** transparent. Assessing whether it is appropriate under a set of rules should be computable **so our data systems?? are designed to respond?? automatically??**. And individuals and institutions alike should be held accountable for misuse **of the information in their care,?? even if such care is only temporary??**. Information accountability means that information usage should be transparent so it is possible to determine whether a use is appropriate under a given set of rules and that the system enables individuals and institutions to be held

Abelson 06/08 4/15/08 4:11 PM Page 2

accountable for misuse.

Transparency and accountability make bad acts visible to all concerned. However, visibility alone does not guarantee compliance. Then again, the vast majority of legal and social rules that form the fabric of our societies are not enforced perfectly or automatically, yet somehow most of us follow most of them most of the time. We do so because social systems built up over thousands of years encourage us, often making compliance easier than violation. For those rare cases where rules are broken, we are all aware that we may be held accountable through a process that looks back through the records of our actions and assesses them against the rules.

Personal privacy, copyright protection, and government surveillance are among the more intractable policy challenges in our information society. In each of these policy areas, excessive reliance on secrecy and up-front control over information has resulted in policies that fail to meet social needs, as well as in technologies that stifle information flow without actually resolving the problems for which they are designed.

Information privacy rights aim to safeguard individual autonomy against the power that institutions or individuals gain over others through the use of personal information.[1] Sensitive, and possibly inaccurate, information may be used against people in financial, **political??**, employment, and health-care settings. In democratic societies, citizens' behavior is unduly restrained if they fear they are being watched at every turn. They may not read **may deliberately avoid reading??** controversial material or feel inhibited from associating with certain communities**, people,?? and ideas??** for fear of adverse **social,?? financial,?? political,?? legal??** consequences.

Protecting privacy is more challenging than ever due to the proliferation of personal information on the Web and the increasing analytical power available to large institutions and to everyone else through Web search engines and other facilities.[2] Access control and collection limits over a single instance of personal data are insufficient to guarantee the protection of privacy when either the same information is publicly available elsewhere on the Web or it is possible to infer private details to a high degree of accuracy from other information that itself is public [8, 10]. Worse, many privacy protections (such as lengthy online privacy-policy statements and in the context of health care and financial services) are mere fig leaves over the increasing exposure of our social and commercial interactions. In the case of publicly available personal information, people often intentionally make the data available, not always by accident [9]. They may not intend for it to be used for every conceivable purpose but are willing for it to be public nonetheless.

Even technological tools that help individuals make informed choices about data-collection practices they are prepared to permit are no longer sufficient to protect privacy in the age of the Web. As a case in point, the growth of e-commerce over the second half of the 1990s sparked concern among Web users worldwide about consumer **about their personal??** privacy that led to an emphasis **by e-businesses??** on Web-site privacy policies and

Abelson 06/08 4/15/08 4:11 PM Page 3

infrastructure (such as the World Wide Web Consortium's Platform for Privacy Preferences, or P3P, www.w3.org/P3P/). A fully implemented P3P environment gives Web users the ability to make privacy choices about every single request by business **organizations?? and government agencies??** to collect information about them. However, the number, frequency, and specificity of these choices would be overwhelming, especially if they cover all possible future uses by the data collector and by third parties. Individuals should not have to agree in advance to complex policies with unpredictable outcomes. Moreover, they should be confident that there will be redress if they are harmed by the improper use of the information they provide. Otherwise, individuals cannot be expected to see any reason to attend to privacy choices.

Consider the complexities of protecting privacy in this scenario: Alice is the mother of a three-year-old child with a severe chronic illness. She learns all she can about it, buying books online, searching the Web, and participating in online parent-support social networks and chat rooms. She then applies for a job and is rejected, suspecting it's because a background check identified her Web activities and flagged her as high risk for expensive family health costs.

Such tales are offered to support the argument for Web privacy. Did the bookstore **the online bookstores??** assert that the titles of Alice's purchases would be kept confidential? Did AOL promise never to release information about her online searches? Did the chat service guard against lurkers in the chat room, recording the names of participants? A policy regime based on information hiding would focus on these potential acts of data release, perhaps even taking the position that it is Alice's **own personal??** responsibility to inform herself about the privacy policies of Web sites before using their services. This focus is misplaced. The actual harm was caused not by the disclosure of information by the bookseller, AOL, or chat service, but by the decision to deny Alice the job, that is, by the inappropriate, discriminatory, and possibly illegal use of the information. It is quite conceivable that Alice wants to be publicly identified as someone with an interest in her child's particular illness. Forcing her to hide in order to protect herself against improper information use significantly limits her ability to exercise her right to freedom of association. Rather, we **Alice?? and everyone else??** should be able to live in an online environment that provides transparent information use and accountability to rules that limit the harmful use of personal information.

Copyright

Looking into copyright and government surveillance reveals deficiencies in the reliance on information hiding as a policy tool. In the copyright context, information hiding commonly takes the form of digital rights management (DRM). As with personal privacy, locking up information is extremely difficult, and efforts at up-front control over the information flow results in user frustration and substantially imperfect security. This is a lesson that even the most ambitious online businesses have

Abelson 06/08 4/15/08 4:11 PM Page 4

learned. For example, in early 2007, Apple CEO Steve Jobs wrote [5] that DRM has not worked nor is it ever likely to work. Soon afterward, Apple changed the way it sells music online by offering a more expensive version of its download service unencumbered by DRM. Apple now implements a basic form of information accountability. The newly unlocked tracks include the purchaser's name and other personally identifying information. That way, if he or she shares the purchased music with, say, a hundred million closest friends through the Internet, the purchaser could be held accountable.

The Creative Commons, another approach to online copyright protection, likewise does not rely on up-front enforcement of licenses. Rather, its architecture recognizes the value of having information flow freely around the Internet but still seeks to impose certain restrictions on how the information is used.

Government Data Mining

Recent government use of advanced data mining techniques is another example of the deficiency of access-control and collection-limitation approaches to privacy compliance on the Web. Laws that limit access to information do not protect privacy here because so much of the data is publicly available. To date, neither law nor technology has developed a way to address this privacy loophole [2].

Airline passenger screening by law-enforcement and national-security agencies illustrates the growing complexity of information handling and transfer. Society may be prepared to accept and even expect national security agencies to use aggressive data mining techniques over a range of information in order to identify potential terrorism risks. However, we **most?? a substantial percentage of?? U.S. citizens??** consider it unacceptable to use the same information with the same powerful analytic tools to investigate domestic criminal activity. Therefore, we need rules **in the U.S.?? everywhere??** that address permissible uses of certain classes of information, in addition to simple access and collection limitations.

Legal Framework

The information-accountability framework more closely mirrors the relationship between the law and human behavior than do the various efforts to enforce policy compliance through access control over information. As an early illustration of information accountability at work today, consider credit bureaus and their large collections of personal information. When these databases came on the scene in the consumer financial markets of the 1960s, policymakers recognized the public imperative to protect individual privacy and assure data accuracy, all while maintaining enough flexibility to allow analysis of consumer credit data based on the maximum amount of useful information possible. Under the Fair Credit Reporting Act (enacted 1970) [3], privacy is protected **in the U.S.??** not by limiting the collection of data, but by placing strict rules on how it may be used. Analysis for the purpose of developing a credit score is essentially unconstrained, but the resulting information can be used only for credit or employment purposes. It cannot be used for marketing or other profiling. Strict penalties are imposed by

the FCRA for the breach of these use limitations. Data quality is protected by giving all consumers the right to see the data held about them (transparency). If a user of the data makes a decision adverse to the consumer (such as denial of a loan or rejection of an employment application) the decision must be justified with reference to the specific data in the credit report on which the decision was based (accountability). If the consumer discovers that the data is inaccurate, he/she may demand that it be corrected. Stiff financial penalties are imposed by the FCRA against the credit bureau if it fails to make the appropriate corrections.

The typical consumer appreciates the paradox associated with protecting privacy or other information policy values through increased transparency. As the FCRA illustrates, we achieve greater information accountability only by making better use of the information that is collected and by retaining the data that is necessary to hold data users responsible for policy compliance. The success of this accountability regime for the past 40 years over a very large set of data—credit reports on nearly every adult in the U.S.—makes it a worthy model for considering policy compliance in other large systems.

Technical Architectures

What technical architecture might be required to support information accountability? Our goal in promoting accountability systems is to build into our information infrastructures the technology necessary to make acts of information usage more transparent in order to hold the individuals and institutions who misuses data accountable for their acts. Systems supporting information accountability require three basic architectural features:

Policy-aware transaction logs. In a decentralized system each endpoint will have to **must??** assume the responsibility of recording information-use events that may be relevant to current or future assessment of accountability to some set of policies.

Policy-language framework. Assessing policy compliance over a set of transactions logged at a heterogeneous set of endpoints by diverse human actors requires a common framework for describing policy rules. Drawing on semantic Web techniques, larger and larger overlapping communities on the Web can develop shared policy vocabularies in a bottom-up fashion. Perfect global interoperability of these policies is unlikely but not a fatal flaw. Just as human societies learn to cope with overlapping and sometimes contradictory rules, so too will policy-aware systems be able **be likely??** to develop at least partial interoperability [1].

Policy-reasoning tools. Accountable systems must **be able to??** assist users in answering such questions as: Is this piece of data allowed to be used for a given purpose? and Can a given string of inferences permissible be used in a given context, depending on the provenance of the data and the applicable rules? One possible approach to designing accountable systems is to place a series of accountable appliances throughout the system that communicate using Web-based protocols [7]. Accountability appliances would serve as proxies to data sources, mediating access to the data, and maintain provenance information and logs of

Abelson 06/08 4/15/08 4:11 PM Page 6

data transfers. They could also present accountability reasoning in human-readable ways and allow annotation, editing, and publishing of the data and reasoning being presented [6]. This aspect of the accountability and transparency perspective is closely related to the issue of maintaining provenance for scientific data [4, 11].

Conclusion

Alan Westin published a landmark study *Privacy and Freedom* in 1967 [12]. Still in the age of mainframe computers, it set the stage for thinking about privacy over the next three decades. Westin presented what has become a classic definition of privacy, emphasizing the individual's right to control how personal information "is communicated to others." An information-accountability perspective on privacy would reframe this definition, shifting toward how **personal??** information is used. So, following Westin, we would say that privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is used lawfully and appropriately by others.

Westin's work is essential today for identifying the role of privacy in a free society. However, advances in communications and information technology and the ease of data searching and aggregation have rendered his definition incomplete as a framework for information policy and information architectures that seek to be policy aware.

Will the new tools and laws we've described here put an end to all privacy invasion, unfair misuse of personal information, copyright infringement, and identity theft? Of course not. Perfect compliance is not the proper standard against which to judge laws or systems that help enforce them. Rather we should ask how to build systems that encourage compliance and maximize the possibility of accountability for violations. We should see clearly that our information-policy goals can be **cannot be??** achieved by restricting the flow of information alone. While the accountability approach is a departure from contemporary computer and network systems policy techniques, it is far more consistent with the way legal rules traditionally work in democratic societies.

Contemporary information systems depart from the norm of social systems in the way they seek to enforce rules up front by precluding any possibility of violation, generally through the application of strong cryptographic techniques. In contrast, we follow rules because we are aware of what they are and because we know there will be consequences, after the fact, if we violate them. Technology will better support freedom by relying on these social compacts than by seeking to supplant them.

References

1. Barth, A., Mitchell, J., and Rosenstein, J. Conflict and combination in privacy policy languages. In *Proceedings of the 2004 ACM Workshop on Privacy in the Electronic Society* (Washington, D.C., Oct. 28). ACM, New York, 2004, 45–46.

2. Dempsey, J. and Flint, L. Commercial data and national security. *The George Washington Law*

Abelson 06/08 4/15/08 4:11 PM Page 7

Review 72, 6 (Aug. 2004).

3. Fair Credit Reporting Act, 15 U.S.C.†ß†1681; **how about a URL??.**

4. Golbeck, G. and Hendler, J. A semantic Web approach to the provenance challenge. **In??** Concurrency and Computation: Practice and Experience **is this a book title?? a journal title??** 2007 **if a book, need publisher name,?? publisher's city,??** 2007 **if a journal, need vol.??, no?? (month?? quarter??** 2007), **pages??.**

5. Jobs, S. *Thoughts on Music* (Feb. 6, 2007); www.apple.com/hotnews/thoughtsonmusic/.

6. Kagal, L., Hanson, C., and Weitzner, D. Integrated policy explanations via dependency tracking. In *Proceedings of the IEEE Workshop on Policies for Distributed Systems and Networks* (June 2–4, 2008).

7. Lunt, T. *Protecting Privacy in Terrorist-Tracking Applications.* Presentation to the Department of Defense Technology and Privacy Advisory Committee (**Arlington, VA?? Washington, DC??,** Sept. 29, 2003); www.sainc.com/tapac/library/Sept29/Lunt-Presentation.pdf. **beware dead link??**

8. Samarati, P. Protecting respondent's privacy in microdata release. *IEEE Transactions on Knowledge and Data Engineering 13,* 6 (Nov./Dec. 2001), 1010–1027.

9. Solove, D. *The Digital Person.* New York University Press, New York, 2004.

10. Sweeney, L. K-anonymity: A model for protecting privacy. *International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems 10,* 5 (**month?? quarter??** 2002), 557–570.

11. Szomszor, M. and Moreau, L. Recording and reasoning over data provenance in Web and grid services. In *Proceedings of the International Conference on Ontologies, Databases, and Applications of Semantics 2888* (Catania, Sicily, Italy, **date(s)??** 2003), 603–620.

12. Westin, A. *Privacy and Freedom.* Atheneum Press, New York, 1967.

Daniel J. Weitzner (djweitzner@csail.mit.edu) is Director of the MIT Decentralized Information Group, principal research scientist at the MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, MA, and Technology and Society Policy Director of the the World Wide Web Consortium.

Harold Abelson (hal@mit.edu) is the Class of 1922 Professor of Computer Science and Engineering at MIT, Cambridge, MA.

Tim Berners-Lee (timbl@csail.mit.edu) is Director of the World Wide Web Consortium and holds the 3Com Founders chair and is a senior research scientist at the Laboratory for Computer Science and Artificial Intelligence at the Massachusetts Institute of Technology, Cambridge, MA.

Abelson 06/08 4/15/08 4:11 PM Page 8

Joan Feigenbaum (joan.feigenbaum@yale.edu) is the Grace Murray Hopper Professor of Computer Science at Yale University, New Haven, CT.

James Hendler (hendler@cs.rpi.edu) is **a??** professor of computer science and the Tetherless World Constellation Chair at Rensselaer Polytechnic Institute, Troy, NY.

Gerald Jay Sussman (gjs@mit.edu) is the Panasonic Professor of Electrical Engineering at MIT, Cambridge, MA.

(acknowledgement)

The work reported here was conducted at MIT, RPI, and Yale with support from the National Science Foundation Cybertrust Grant (award #04281) and IARPA (award #FA8750-07-2-0031).

Pull Quotes

In democratic societies, citizens' behavior is unduly restrained if they fear being watched at every turn.

Information accountability means that information usage should be transparent so it is possible to determine whether a use is appropriate under a given set of rules.

Contemporary information systems depart from the norm of social systems in the way they seek to enforce rules up front by precluding any possibility of violation.

We should ask how to build systems that encourage compliance and maximize the possibility of accountability for violations.