

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)

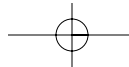
¹There are numerous definitions for privacy. Our chief interest here is understanding privacy rights as they relate to the collection and use of personal information, as opposed to other privacy protections that seek to preserve control over, say, one's bodily **one's physical??** integrity.

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

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

12 For too long, our approach to information-protec-
13 tion policy has been to seek ways to prevent informa-
14 tion from "escaping" beyond appropriate boundaries,
15 then wring our hands when it inevitably does. This
16 hide-it-or-lose-it perspective dominates technical and
17 public-policy approaches to fundamental social ques-
18 tions of online privacy, copyright, and surveillance.
19 Yet it is increasingly inadequate for a connected world
20 where information is easily copied and aggregated
21 and where automated correlations and inferences
22 across multiple databases uncover **routinely expose??**
23 information even when it is not explicitly revealed. As
24 an alternative, accountability must become a primary
25 **means** through which society addresses issues of
26 appropriate use. Information accountability means
27 that such use should be **must be made??** transparent.
28 Assessing whether it is appropriate under a set of rules
29 should be computable **so our data systems?? are**
30 **designed to respond?? automatically??**. And individ-
31 uals and institutions alike should be held accountable
32 for **misuse of the information in their care,?? even if**
33 **such care is only temporary??**. Information account-
34 ability means that information usage should be trans-
35 parent so it is possible to determine whether a use is
36 appropriate under a given set of rules and that the sys-
37 tem enables individuals and institutions to be held





38 accountable for misuse.

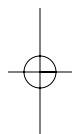
39 Transparency and accountability make bad acts
40 visible to all concerned. However, visibility alone
41 does not guarantee compliance. Then again, the vast
42 majority of legal and social rules that form the fabric
43 of our societies are not enforced perfectly or auto-
44 matically, yet somehow most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

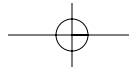
52 Personal privacy, copyright protection, and gov-
53 ernment surveillance are among the more intractable
54 policy challenges in our information society. In each
55 of these policy areas, excessive reliance on secrecy and
56 up-front control over information has resulted in
57 policies that fail to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

61 Information privacy rights aim to safeguard indi-
62 vidual autonomy against the power that institutions
63 or individuals gain over others through the use of
64 personal information.¹ Sensitive, and possibly inaccur-
65 ate, information may be used against people in
66 financial, **political??**, employment, and health-care
67 settings. In democratic societies, citizens' behavior is
68 unduly restrained if they fear they are being watched
69 at every turn. They may not read **may deliberately**
70 **avoid reading??** controversial material or feel inhib-
71 ited from associating with certain communities, **peo-**
72 **ple,?? and ideas??** for fear of adverse **social,??**
73 **financial,?? political,?? legal??** consequences.

74 Protecting privacy is more challenging than ever
75 due to the proliferation of personal information on
76 the Web and the increasing analytical power available
77 to large institutions and to everyone else through
78 Web search engines and other facilities.² Access con-
79 trol and collection limits over a single instance of per-
80 sonal data are insufficient to guarantee the protection
81 of privacy when either the same information is pub-
82 licly available elsewhere on the Web or it is possible
83 to infer private details to a high degree of accuracy
84 from other information that itself is public [8, 10].
85 Worse, many privacy protections (such as lengthy
86 online privacy-policy statements and in the context
87 of health care and financial services) are mere fig
88 leaves over the increasing exposure of our social and
89 commercial interactions. In the case of publicly avail-
90 able personal information, people often intentionally
91 make the data available, not always by accident [9].
92 They may not intend for it to be used for every con-
93 ceivable purpose but are willing for it to be public
94 nonetheless.

95 Even technological tools that help individuals
96 make informed choices about data-collection prac-
97 tices they are prepared to permit are no longer suffi-
98 cient to protect privacy in the age of the Web. As a
99 case in point, the growth of e-commerce over the sec-
100 ond half of the 1990s sparked concern among Web
101 users worldwide about consumer **about their per-**
102 **sonal??** privacy that led to an emphasis **by e-busi-**
103 **nesses??** on Web-site privacy policies and





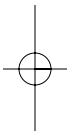
104 infrastructure (such as the World Wide Web Consor-
105 tium's Platform for Privacy Preferences, or P3P,
106 www.w3.org/P3P/). A fully implemented P3P envi-
107 ronment gives Web users the ability to make privacy
108 choices about every single request by business **orga-**
109 **nizations?? and government agencies??** to collect
110 information about them. However, the number, fre-
111 quency, and specificity of these choices would be
112 overwhelming, especially if they cover all possible
113 future uses by the data collector and by third parties.
114 Individuals should not have to agree in advance to
115 complex policies with unpredictable outcomes.
116 Moreover, they should be confident that there will be
117 redress if they are harmed by the improper use of the
118 information they provide. Otherwise, individuals
119 cannot be expected to see any reason to attend to pri-
120 vacy choices.

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

131 Such tales are offered to support the argument for
132 Web privacy. Did the bookstore **the online book-**
133 **stores??** assert that the titles of Alice's purchases
134 would be kept confidential? Did AOL promise never
135 to release information about her online searches? Did
136 the chat service guard against lurkers in the chat
137 room, recording the names of participants? A policy
138 regime based on information hiding would focus on
139 these potential acts of data release, perhaps even tak-
140 ing the position that it is Alice's **own personal??**
141 responsibility to inform herself about the privacy
142 policies of Web sites before using their services. This
143 focus is misplaced. The actual harm was caused not
144 by the disclosure of information by the bookseller,
145 AOL, or chat service, but by the decision to deny
146 Alice the job, that is, by the inappropriate, discrimi-
147 natory, and possibly illegal use of the information. It
148 is quite conceivable that Alice wants to be publicly
149 identified as someone with an interest in her child's
150 particular illness. Forcing her to hide in order to pro-
151 tect herself against improper information use signifi-
152 cantly limits her ability to exercise her right to
153 freedom of association. Rather, we **Alice?? and every-**
154 **one else??** should be able to live in an online envi-
155 ronment that provides transparent information use
156 and accountability to rules that limit the harmful use
157 of personal information.

158 **Copyright**

159 Looking into copyright and government surveil-
160 lance reveals deficiencies in the reliance on informa-
161 tion hiding as a policy tool. In the copyright
162 context, information hiding commonly takes the
163 form of digital rights management (DRM). As with
164 personal privacy, locking up information is
165 extremely difficult, and efforts at up-front control
166 over the information flow results in user frustration
167 and substantially imperfect security. This is a lesson
168 that even the most ambitious online businesses have
169



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

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

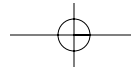
188 **Government Data Mining**

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

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

211 **Legal Framework**

212 The information-accountability framework more
213 closely mirrors the relationship between the law and
214 human behavior than do the various efforts to
215 enforce policy compliance through access control
216 over information. As an early illustration of informa-
217 tion accountability at work today, consider credit
218 bureaus and their large collections of personal infor-
219 mation. When these databases came on the scene in
220 the consumer financial markets of the 1960s, policy-
221 makers recognized the public imperative to protect
222 individual privacy and assure data accuracy, all while
223 maintaining enough flexibility to allow analysis of
224 consumer credit data based on the maximum
225 amount of useful information possible. Under the
226 Fair Credit Reporting Act (enacted 1970) [3], privacy
227 is protected **in the U.S.??** not by limiting the collec-
228 tion of data, but by placing strict rules on how it may
229 be used. Analysis for the purpose of developing a
230 credit score is essentially unconstrained, but the
231 resulting information can be used only for credit or
232 employment purposes. It cannot be used for market-
233 ing or other profiling. Strict penalties are imposed by
234
235



236 the FCRA for the breach of these use limitations.
 237 Data quality is protected by giving all consumers the
 238 right to see the data held about them (transparency).
 239 If a user of the data makes a decision adverse to the
 240 consumer (such as denial of a loan or rejection of an
 241 employment application) the decision must be justifi-
 242 fied with reference to the specific data in the credit
 243 report on which the decision was based (accountabil-
 244 ity). If the consumer discovers that the data is inac-
 245 curate, he/she may demand that it be corrected. Stiff
 246 financial penalties are imposed by the FCRA against
 247 the credit bureau if it fails to make the appropriate
 248 corrections.

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

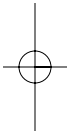
261 262 **Technical Architectures**

263 What technical architecture might be required to
 264 support information accountability? Our goal in
 265 promoting accountability systems is to build into
 266 our information infrastructures the technology nec-
 267 essary to make acts of information usage more
 268 transparent in order to hold the individuals and
 269 institutions who misuses data accountable for their
 270 acts. Systems supporting information accountability
 271 require three basic architectural features:

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

277 *Policy-language framework.* Assessing policy com-
 278 pliance over a set of transactions logged at a hetero-
 279 geneous set of endpoints by diverse human actors
 280 requires a common framework for describing policy
 281 rules. Drawing on semantic Web techniques, larger
 282 and larger overlapping communities on the Web can
 283 develop shared policy vocabularies in a bottom-up
 284 fashion. Perfect global interoperability of these poli-
 285 cies is unlikely but not a fatal flaw. Just as human
 286 societies learn to cope with overlapping and some-
 287 times contradictory rules, so too will policy-aware
 288 systems be able **be likely??** to develop at least partial
 289 interoperability [1].

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



302 data transfers. They could also present accountability
 303 reasoning in human-readable ways and allow anno-
 304 tation, editing, and publishing of the data and rea-
 305 soning being presented [6]. This aspect of the
 306 accountability and transparency perspective is
 307 closely related to the issue of maintaining prove-
 308 nance for scientific data [4, 11].

309 **Conclusion**

310 Alan Westin published a landmark study *Privacy*
 311 *and Freedom* in 1967 [12]. Still in the age of main-
 312 frame computers, it set the stage for thinking about
 313 privacy over the next three decades. Westin pre-
 314 sented what has become a classic definition of pri-
 315 vacy, emphasizing the individual's right to control
 316 how personal information "is communicated to
 317 others." An information-accountability perspective
 318 on privacy would reframe this definition, shifting
 319 toward how **personal??** information is used. So, fol-
 320 lowing Westin, we would say that privacy is the
 321 claim of individuals, groups, or institutions to deter-
 322 mine for themselves when, how, and to what extent
 323 information about them is used lawfully and appro-
 324 priately by others.

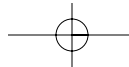
325 Westin's work is essential today for identifying the
 326 role of privacy in a free society. However, advances in
 327 communications and information technology and
 328 the ease of data searching and aggregation have ren-
 329 dered his definition incomplete as a framework for
 330 information policy and information architectures
 331 that seek to be policy aware.

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

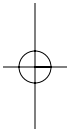
347 Contemporary information systems depart from
 348 the norm of social systems in the way they seek to
 349 enforce rules up front by precluding any possibility
 350 of violation, generally through the application of
 351 strong cryptographic techniques. In contrast, we fol-
 352 low rules because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

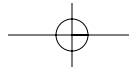
358 **References**

- 359 1. Barth, A., Mitchell, J., and Rosenstein, J. Con-
 360 flict and combination in privacy policy languages.
 361 In *Proceedings of the 2004 ACM Workshop on Pri-
 362 vacy in the Electronic Society* (Washington, D.C.,
 363 Oct. 28). ACM, New York, 2004, 45–46.
- 364 2. Dempsey, J. and Flint, L. Commercial data and
 365 national security. *The George Washington Law*
 366 *Journal*, 2004, 103–110.



- 368 *Review 72*, 6 (Aug. 2004).
 369
 370 3. Fair Credit Reporting Act, 15 U.S.C.†§†1681;
 371 **how about a URL??**.
 372
 373 4. Golbeck, G. and Hendler, J. A semantic Web
 374 approach to the provenance challenge. In?? Concur-
 375 rency and Computation: Practice and Experience is
 376 **this a book title?? a journal title?? 2007 if a book,**
 377 **need publisher name?? publisher's city?? 2007 if a**
 378 **journal, need vol.??, no?? (month?? quarter??**
 379 **2007), pages??**.
 380
 381 5. Jobs, S. *Thoughts on Music* (Feb. 6, 2007);
 382 www.apple.com/hotnews/thoughtsonmusic/.
 383
 384 6. Kagal, L., Hanson, C., and Weitzner, D. Inte-
 385 grated policy explanations via dependency tracking.
 386 In *Proceedings of the IEEE Workshop on Policies for*
 387 *Distributed Systems and Networks* (June 2–4, 2008).
 388
 389 7. Lunt, T. *Protecting Privacy in Terrorist-Tracking*
 390 *Applications*. Presentation to the Department of
 391 Defense Technology and Privacy Advisory Commit-
 392 tee (Arlington, VA?? Washington, DC??, Sept. 29,
 393 2003); [www.sainc.com/tapac/library/Sept29/Lunt-](http://www.sainc.com/tapac/library/Sept29/Lunt-Presentation.pdf)
 394 [Presentation.pdf](http://www.sainc.com/tapac/library/Sept29/Lunt-Presentation.pdf). **beware dead link??**
 395
 396 8. Samarati, P. Protecting respondent's privacy in
 397 microdata release. *IEEE Transactions on Knowledge*
 398 *and Data Engineering 13*, 6 (Nov./Dec. 2001),
 399 1010–1027.
 400
 401 9. Solove, D. *The Digital Person*. New York Univer-
 402 sity Press, New York, 2004.
 403
 404 10. Sweeney, L. K-anonymity: A model for protect-
 405 ing privacy. *International Journal on Uncertainty,*
 406 *Fuzziness, and Knowledge-based Systems 10*, 5
 407 (month?? quarter?? 2002), 557–570.
 408
 409 11. Szomszor, M. and Moreau, L. Recording and
 410 reasoning over data provenance in Web and grid ser-
 411 vices. In *Proceedings of the International Conference*
 412 *on Ontologies, Databases, and Applications of Seman-*
 413 *tics 2888* (Catania, Sicily, Italy, date(s)?? 2003),
 414 603–620.
 415
 416 12. Westin, A. *Privacy and Freedom*. Atheneum
 417 Press, New York, 1967.
-
- 419 **DANIEL J. WEITZNER** (djweitzner@csail.mit.edu) is Director of the
 420 MIT Decentralized Information Group, principal research scientist at
 421 the MIT Computer Science and Artificial Intelligence Laboratory,
 422 Cambridge, MA, and Technology and Society Policy Director of the
 423 the World Wide Web Consortium.
 424
-
- 425 **HAROLD ABELSON** (hal@mit.edu) is the Class of 1922 Professor of
 426 Computer Science and Engineering at MIT, Cambridge, MA.
 427
-
- 428 **TIM BERNERS-LEE** (timbl@csail.mit.edu) is Director of the World
 429 Wide Web Consortium and holds the 3Com Founders chair and is a
 430 senior research scientist at the Laboratory for Computer Science and
 431 Artificial Intelligence at the Massachusetts Institute of Technology,
 432 Cambridge, MA.
 433





434
435 **JOAN FEIGENBAUM** (joan.feigenbaum@yale.edu) is the Grace Mur-
436 ray Hopper Professor of Computer Science at Yale University, New
437 Haven, CT.

438
439 **JAMES HENDLER** (hendler@cs.rpi.edu) is a?? professor of computer
440 science and the Tetherless World Constellation Chair at Rensselaer
441 Polytechnic Institute, Troy, NY.

442
443 **GERALD JAY SUSSMAN** (gjs@mit.edu) is the Panasonic Professor of
444 Electrical Engineering at MIT, Cambridge, MA.

445
446
447 **(acknowledgement)**

448
449 The work reported here was conducted at MIT,
450 RPI, and Yale with support from the National Sci-
451 ence Foundation Cybertrust Grant (award #04281)
452 and IARPA (award #FA8750-07-2-0031).

453
454
455 **Pull Quotes**

456
457 In democratic societies, citizens' behavior is unduly
458 restrained if they fear being watched at every turn.

459
460
461 Information accountability means that information
462 usage should be transparent so it is possible to deter-
463 mine whether a use is appropriate under a given set
464 of rules.

465
466
467 Contemporary information systems depart from the
468 norm of social systems in the way they seek to
469 enforce rules up front by precluding any possibility
470 of violation.

471
472
473 We should ask how to build systems that encourage
474 compliance and maximize the possibility of
475 accountability for violations.

476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499

