2008

Competition and Privacy in the Web 2.0 and the Cloud

Randal C. Picker

Follow this and additional works at: http://chicagounbound.uchicago.edu/law_and_economics

Part of the Law Commons

Recommended Citation
Competition and Privacy in Web 2.0 and the Cloud

Randal C. Picker

THE LAW SCHOOL
THE UNIVERSITY OF CHICAGO

June 2008

This paper can be downloaded without charge at:
and at the Social Science Research Network Electronic Paper Collection:
We are once again changing how we use computers. In the past, we moved from mainframes to mini computers to freestanding personal computers. That was a powerful shift in control and organizational structure. Mainframes were rare and were treated as such, tended to with loving care and serviced by a small caste of computing priests. PCs, in contrast, were everywhere: on every knowledge worker’s desk and eventually in the family room of many homes. Full decentralization. In the PC age, the computer desktop was the most valuable real estate around, and, for most people, that meant Microsoft Windows.

Microsoft Windows was—and is—both product and delivery system. Product in the sense that Windows performed certain functions that all operating systems perform. Windows tracks files, sends data through ports for printing and tells your computer screen how to display fonts and images. Basic stuff that we expect of our operating systems. But Windows is more than that: Window delivers software. Software delivery, especially before the Internet, was difficult. A consumer might find the software pre-installed on a new PC. Or the consumer might go to a computer store—remember those?—and plunk down her credit card and walk out with a large, almost empty box, that had, buried within it somewhere, a plastic CD with new software.

But Microsoft could guarantee software delivery by just incorporating the new software into Windows. With each new release of Windows—as Windows moved down the development path from
Windows 3.1 to Windows 95 to 98 and on towards Vista—Microsoft expanded the footprint of Windows. This was not just a question of more megabytes; Windows got bigger because it expanded its functionality and in doing so killed off what had been separate markets in freestanding functions. Including a product in the next version of Windows insured its widespread distribution as each version of Windows quickly expanded its market share.

In some basic sense, Windows was fundamentally unbounded. That is, there was no obvious boundary for the scope of functions that might be embraced in Windows.¹ This mattered most when we introduced ubiquitous networks to link computers together to create the Internet and the Web. The move to networked devices created a possible inflection point, a point of churn and competition as different firms sought the upperhand in the new computing space. In his May, 1995 Internet Tidal Wave memo, Bill Gates famously feared that Netscape would “commoditize the underlying operating system.”² Windows was going to become plumbing, important to be sure, but fundamentally anonymous and only noticed when it wasn’t working right. Microsoft moved aggressively against Netscape and relied heavily on its ability to bundle Internet Explorer with Windows at no additional charge to defeat Netscape. Microsoft won its battle against Netscape, though it did so in ways found to be illegal by competition authorities in the United States.³


³ United States v Microsoft Corp., 253 F3d 34, 58 (DC Cir 2001) (en banc). The European Union also found that Microsoft had abused its dominant position in operating systems, though the focus of the EU case was on interoperability with servers and the bundling of Windows Media Player with Windows. See Microsoft Corp. v Commission of the European Communities, Case T-201/04, Court of First Instance.
But Microsoft seems to be losing the larger war suggested by the tidal wave. We are in the midst of two large related shifts in our computing platform. The first shift, which often travels under the name Web 2.0, is fundamentally about what we use computers to do. We have moved from creating documents in Microsoft Office to living life online: searching on Google, buying and selling on eBay, hanging out with our friends on mySpace and Facebook, watching the newest viral video on YouTube. The second shift, often called cloud computing, is more about a change in the organization of the fundamental processes of computing—computation and storage—with some overlap with the Web 2.0 shift. Instead of storing my email on my laptop, I will just outsource storage and store it with Google. I won’t have an email product resident on my computer; instead, Google will provide an email service through a Web browser.

These shifts have one key point in common: the possibility of creating prodigious amounts of data about end users. The new web intermediaries at the heart of Web 2.0 have access to an enormous datastream about their users. Google can learn a great deal about my interests with every search that I run. Facebook learns about me as I build my profile and link to my friends. Imagine how much a cloud storage provider might learn about me if the provider could read all of my stored email and documents. These data are the lifeblood of Web 2.0 and could play a similarly important role as a cloud infrastructure emerges. The advertising that supports much of the content on the Internet is more valuable if it can be matched to my actual interests, and the flexibility of the web in delivering content means that web advertising is increasingly tailored advertising, or so-called behavioral advertising.

How we choose to regulate these datastreams is the central regulatory issue of the emerging computer infrastructure. Our choices here obviously have privacy consequences but also for how much competition will emerge. These are tightly linked. We have frequently regulated how intermediaries can use the information that
passes through their hands. Banks, cable companies, phone companies—even your local video store—face strong restrictions on how they can use the information seen by them as they process many of our transactions. Law disables them from using that information. In contrast, the emerging financial infrastructure for financing Web 2.0—free content paid for by online advertising supported by rich databases—is largely unregulated. To be sure, we have to assess whether the privacy issues are different in the online space, but we also need consider how different regulation of this transactional information implicates competition.

The current Web 2.0 market is dominated by Google. And like Windows, Google’s infrastructure has no obvious boundaries. Indeed, the Google engine is in many ways more powerful than Microsoft’s. It isn’t obvious how the size of Windows or how its functionality affected the price that Microsoft could charge for Windows. When Microsoft added browser functionality to Windows, it didn’t necessarily increase the price of Windows. In contrast, Google’s “price” scales up directly with each added service that it finances through advertising. Google’s expansion model results in additional revenue with each ad that is clicked. Like Windows, Google’s business has no obvious boundaries. The limit seems to be the content or services that can be supported by advertising and might be as large as anything mediated by a display screen, but, unlike Windows, Google’s revenue scales as more services are added.

1. Finding Data or Throwing It Away?

We should start with desktop computing before the emergence of the Internet. Microsoft Office—Excel, Outlook, PowerPoint and Word—set the standard for desktop productivity tools. These were the tools that we used to create documents that resided on the hard disks in our desktops or laptops. Outlook was used to manage calendar, contacts and email. The CPUs in our computers churned away to do the calculations in an Excel spreadsheet or to format a
document in Word. These documents were then distributed, on paper or via email, to be read by the recipients.

Now think about what you use your computer for today. In this new era, we might think of matching and coordination as being the defining tasks we expect software to perform. eBay is explicitly about creating a marketplace to match buyers and sellers. Craigslist matches everything under the sun: buyers and sellers to be sure and job seekers galore, but also personals and house swaps, lost and found items and rideshares. Social networking sites like mySpace and Facebook match individuals to define new groups. And Google matches people looking for content with the websites where that content is stored.

This is the emergence of a new class of online intermediaries. They typically operate over the Internet through a Web browser. They can charge transaction fees like eBay or charge for a job posting like Craigslist. Given the number of pageviews that take place, the intermediary can support all of the content with advertising as Google does. Given the ready ability to match advertising with content, a platform that generates pageviews is a valuable media property.

But there is more. The intermediary has the ability to see what is happening with every click and this creates an incredibly rich clickstream.4 eBay may be able to figure out whether I am more of a Cubs fan than a White Sox fan and how much I like Pokémon. Google has an even deeper knowledge of my interests, as I search far more often than I buy or sell on eBay. This datastream arises organically as part of the services that that website performs. We can choose to limit the use or disclosure of this information, but we almost have to engineer throwing away the information. It would otherwise emerge naturally from the role played by the intermediary.

---

The emergence of these Web intermediaries is one of the defining aspects of Web 2.0. The idea of cloud computing is related but a little different. Computing power was highly centralized with mainframes, and then we decentralized through minicomputers and PCs. With the cloud, content and computing power will increasingly be managed centrally. The problem with owning a PC is that you are your own tech support and most of us are getting lousy service. Computers are complicated. Badly-run computers inflect harm on all of us, when their power is harvested in botnets and computer spam is sent across the globe. And PCs are lumpy: you buy computing power at one time and not just when you need it.

It doesn’t have to work that way. Most people wouldn’t consider for a second rolling their own electricity; they expect to get it from a socket and want to rely on the local electricity company to do the hard work that lies behind that. We may be headed that direction on computing power, both for calculation and storage. Some content may be stored locally on your machine, while other content—content that you in some powerful sense think belongs to you—will be stored remotely. Where actually? You won’t have a clue.

Most people probably don’t have strong feelings about where their computer calculations are done. Whether most of the processing power exists locally in the device on your desktop or in your hand is a detail. If communications costs have dropped sufficiently such that we won’t notice when most of the computing is done remotely and then delivered rapidly to our local devices, we can return computing power to the center. This is really just an engineering problem that turns on the relative cost of central and local processing power and of distant and local—really local meaning on the bus inside your

---


computer—communication. Important, to be sure, but not something most end-users will care about.

But that analysis is crucially dependent on an implicit assumption, namely, that changing the location of processing or storage doesn’t change anything about how the datastream associated with processing or storage is used. Think of this as a version of cloud neutrality: where processing or storage is done should be irrelevant—neutral—for outcomes, legal or otherwise. If instead my cloud provider monitored all of my spreadsheet calculations and then tailored advertising to match what it had gleaned from the calculations—“Looks like he’s looking for a mortgage”—that would be a big change. The move to centralized processing and storage creates communications traffic that can be monitored.

Email is a good example. I have a separate email application (Microsoft Outlook) that I use to download email messages and store those on my laptop. I also have an email account through Google (Gmail) which I manage through a web browser. That email is stored remotely with Google and Google dutifully notifies me of how much of “my” storage space at Google that I have filled up. I paid cash for Outlook—more exactly, the University of Chicago has a site license with Microsoft—but I “pay” for Gmail by being exposed to the advertisements that it places on the far right edge of the screen. Google runs my email through a filter to determine which ads I should see.

How we use the rich datastreams that have emerged under Web 2.0 and that may emerge under cloud computing is a point of design, both technical design and legal design. Nothing about the change in the organization of computing need bring with it a change in information revelation. We could choose to limit how these new datastreams are used, as we have often done in the past with new datastreams. And we could have downloadable products supported by advertising, though, to be sure, to maximize the value of the advertising, the product would need to go online periodically to
communicate information about me back to the mothership and to
download new ads based on my locally-stored email. Of course,
usually if that happens, we call the software spyware, that is, software
that sits locally on a computer with the mission of secretly collecting
and reporting back information.

II. Designing Stickiness and Data Portability

The question is what happens to the datastream flowing through the
chokepoint? How is that information used and controlled? How does
that effect both privacy and competition? Take an early Web 2.0
example, eBay. eBay creates stickiness with its user reputation and
feedback scores. eBay mediates transactions between strangers. As a
purchaser, how can I determine whether my prospective seller will
deliver the listed item? Transactions between strangers at a distance is
a long-standing problem in commercial law. eBay users build up a
reputation score transaction by transaction and that reputation is the
key way in which eBay mitigates the problem of transactions at a
distance between strangers.

But the eBay reputation system also has important competitive
consequences. Since the reputation accumulates prior transactions, a
competing auctions entrant starts with an immediate disadvantage.
eBay’s reputation system is sticky, or, put differently, it creates
switching costs. A long-time seller on eBay has a reputation that she
has built up carefully. But if she switches to the entrant, she will be a
newbie again and buyers will naturally be reluctant to transact with
her. But there is a ready solution: make the eBay identity and
reputation portable. If I am a good seller on eBay as HotDVDBuysNow, I should be just as good on another site.

The consequences of stickiness through user ratings and identities
is not lost on eBay’s competitors. They understand the way in which
those scores creates entry barrier for auction competitors. Take the
case of ReverseAuction.com. eBay’s original business relied on
ascending price auctions. Users would bid against each other for a
fixed period of time, and at the end of the auction, the high bid won.
ReverseAuction entered with a declining-price auction website. Sellers offered items for sale, and the offered price declined until a buyer jumped in to buy at the current price. Once that happened, the auction was over.

ReverseAuction understood the competitive disadvantage it faced against eBay. To solve that, at least according to the eventual complaint filed by the Federal Trade Commission, ReverseAuction registered as an eBay user and agreed to the eBay user agreement. ReverseAuction then harvested information from eBay’s website by acquiring eBay user IDs, email addresses and feedback ratings. ReverseAuction then sent an email to eBay’s users suggesting that they could reserve their eBay identities at ReverseAuction and that they should do quickly lest they lose that opportunity.

The FTC found much troubling in RevereAuction’s actions but there was a recognition of the way in which eBay’s control over user reputations—the accumulated results of many transactions—blocked competition in online auctions. The critical point is that portability—or the absence thereof—is a design point. eBay’s user agreement bars users from “importing or exporting feedback information off of the Sites or for using it for purposes unrelated to eBay.” eBay understandably wants to lock-in its users and hopes to do that by restricting the extent to which the valuable eBay-based reputations can be used elsewhere. Reputation and feedback ratings are a tool which allows the auction house to make past transactions relevant today. For law, the question is whether we should limit user agreements that block reputation portability, whether that portability is sought by users directly or by competitors.

Consider another example of portability. I use RSS—Real Simple Syndication—to manage lots of the information that flows through

---

my computer and I use Google Reader to manage that content. If you want to know what I am interested in right now, you want to look at my Google Reader tag cloud. But the right question for law is: as we move from products and local storage to services and centralized storage, who owns the data and what establishes rights to access and use the data? Suppose, for example, that I wanted to drop Google Reader and switch to another tool for managing RSS, say FeedDemon. I can obviously just starting running FeedDemon, but would I have to re-type or re-link to the feeds to get them into FeedDemon? And what of my tagged items? I don’t know how to tell how many items I have in Google Reader with tags, but I suspect that the relevant order of magnitude is in the 1000s.

The answer on the feeds is OPML. Yes, I have no idea what that is either (actually, OPML is the Outline Processor Markup Language), and it is used precisely to create an XML file that should be readable by another RSS program. This isn’t as easy as switching from Diet Pepsi to Diet Coke—the contrary choice is inconceivable—but, assuming that your RSS reader supports importing and exporting OPML, it is doable. But moving the list of feeds over is just one small piece of my information. The detailed matching of news stories and tags represents a much greater share of the value. I don’t see any particularly easy way to export that information into another RSS program.

We can count on competitors to help lower these switching costs. We saw that above with ReverseAuction. In another classic case,9 Borland did this when it sold the spreadsheet Quattro Pro with an alternative interface that emulated that of Lotus 1-2-3, the dominant spreadsheet of the day. Lotus tried to rely on copyright law to defeat Borland and failed though do remember that the vote in the Supreme Court was 4-4 and ties go to the lower court winner, in this case Borland. When I switched my main browsing program from Internet

---

Explorer to Firefox, Firefox looked on my hard disk to find the links that I had stored as IE Favorites, again reducing the transaction costs of switching.

But we see how design matters when we return to my tagged stories. I don’t know for sure, but I don’t think much if any of my Google Reader info is stored locally on my machine. I don’t think that there is any locally-stored info for FeedDemon to examine were I trying to switch over both my feeds list and my tagged stories. And the question is whether FeedDemon could write something that would burrow through my Google Reader “subscription”—that seems like a fair description—to extract my tagged stories.

As the Lotus/Borland saga makes clear, law matters for switching costs and portability. Sometimes that law will be copyright law as it was there. Other times it will be antitrust, as the European Union is trying now in forcing Microsoft to disclose more so as to increase interoperability between operating systems and servers. In other cases, we will legislate portability and interoperability, as we have done with telephone number portability\(^\text{10}\) and as parts of Europe may push Apple on iTunes and the iPod.

III. Controlling How Data Are Used

We should return to Google and consider how Google might use the datastreams that arise in search. Search is an exercise in relevance: for any search term presented, the search engine wants to return the “best” matches. How should we assess best? Brin and Page’s original search patent is for their PageRank algorithm. That algorithm looks to the link structure of the web to measure importance and therefore relevance.\(^\text{11}\) If we are looking for information about the Chicago Cubs, if many pages link to a particular page about the Cubs, we might conclude that that is a particularly relevant page. The


\(^{11}\) For a basic description, see Our Search: Google Technology (online at http://www.google.com/technology/).
PageRank approach emphasizes information that is available publicly. Any entrant could do the same, at least if they could do so consistent with the original patent.

That version of search doesn’t rely on the datastreams that arise in search. But we might imagine an approach which does so and which relies more directly on collective intelligence. Focus on how searchers respond to the presented search results. If searchers routinely reject the first listed item for the second, we would be learning something about the perceived relevance of the results. That approach, multiplied over many users and an almost infinite number of searches, would create a system that learns and evolves in response to what users are doing.\(^\text{12}\) If that learning improved relevance, more searchers would seek to rely on the system, and that in turn would generate more learning. This is a positive feedback loop and should operate as a barrier to entry. Unlike the page-link information at the heart of PageRank which relies on publicly observable data, learning through search results relies on private information available only to the search engine.

These are very different approaches to the use of the datastreams available to Google and that is just in framing how relevance is assessed, the core function of search. The datastream could also be used to match the ads presented next to the organic search results with the searcher, so-called behavioral advertising. John Wanamaker, the department store magnate, famously observed that he wasted half of the money that he spent on advertising, but “I don’t know which half.”\(^\text{13}\) And Wanamaker may have been optimistic. Think about TV advertising and how many ads that you see for products that you never consume. Those ads are almost all wasted. Behavioral


\(^\text{13}\) See John Wanamaker (1838-1922), The Advertising Century (online at http://adage.com/century/people006.html).
advertising offers the promise of tailoring ads to individual consumers greatly increasing the efficiency of each ad dollar spent.

In the past, we have placed extensive controls on how intermediaries can use the information that flows through their hands. For example, the Cable Communications Policy Act of 1984 added privacy protections for consumers. The current version of that statute requires written or electronic consent of cable customers before the cable operator can use the cable system to collect personally identifiable information about its customers. But the statute also creates an exception to that rule allowing collection of such information to detect cable theft and, more generally, “to obtain information necessary to render a cable service or other service provided by the cable operator to the subscriber.” Whether collecting information to implement behavioral advertising will qualify under this safe harbor is an open question. The cable statute also bars disclosure of personally identifiable information to third parties, though, again, the statute exempts disclosures “necessary to render, or conduct a legitimate business activity related to, a cable service or other service provided by the cable operator to the subscriber.”

How we implement privacy restrictions matters enormously and indeed the limits can have perverse consequences. For example, a disclosure limit of the sort seen in the cable statute artificially pushes towards vertical integration. As most disclosure limits don’t prevent disclosure within a particular firm but only bar disclosure across firm boundaries, a firm will have an artificial incentive to expand the size and scope of the firm so as to use the information fully. Vertical integration renders the disclosure limit ineffective. We might see

---

14 PL 98-549, codified at 47 USC 551.
15 47 USC 551(b)(2)(A).
16 47 USC 551(c)(2)(A).
mergers that would otherwise be unattractive as a way to end-run the across-firm disclosure limits.

Note also that disclosure may not be the act of relevance here. If Google runs an ad placement service—of course it does: AdSense\textsuperscript{17}—it need not disclose any information to facilitate matches between content and consumers. For Google’s customers—its advertisers—the information will be in a black box. These customers will be able to evaluate the click-through rates that they are seeing from the use of the information controlled by Google, but they need never see the information itself. No disclosure, just use on their behalf. Indeed, as suggested before, Google would almost certainly prefer not to disclose the information, since disclosing the information gives up the control that Google has from its exclusive access to the information.

\textbf{IV. Conclusion}

With Web 2.0, we have once again changed how we use computers. That change has brought with it new intermediaries who sit at the crossroads of the matching and coordination that define how we use the Internet today. Those intermediaries—Google first and foremost—have access to extraordinarily detailed information about their customers. That information arises naturally from the very services provided. We will see a similar pattern as cloud computing becomes more important, and cloud service providers will also have available to them a rich datastream that arises from their customer’s activities.

To date, these intermediaries have faced few limitations in how they use the information that they see. That information can be used to improve their core businesses—adding collective intelligence to search to increase relevance—and to finance—through advertising backed by rich databases that allows ads to be matched to individual customers—virtually any content or service that can be provided

\textsuperscript{17} \url{https://www.google.com/adsense/login/en_US/}.
through a screen. To focus on Google as the largest player in this space, there is no obvious limit to its scale and an advertising-supported business adds revenue with each additional screen that is viewed.

In the past, we have regulated intermediaries at these transactional bottlenecks—banks, cable companies, phone companies and the like—and limited the ways in which they can use the information that they see. Presumably the same forces that animated those rules—fundamental concerns about customer privacy—need to be assessed for our new information intermediaries. In doing that, we need to be acutely aware of how our choices influence competition. An uneven playing field—allowing one firm to use the information that it sees while blocking others from doing the same thing—creates market power through limiting competition. We rarely want to do that. And privacy rules that limit how information can be used and shared across firms will artificially push towards greater consolidation, something which again usually works against maintaining robust competition.
Readers with comments should address them to:

Professor Randal C. Picker  
University of Chicago Law School  
1111 East 60th Street  
Chicago, IL  60637  
 rpicker@uchicago.edu
300. Adam B. Cox, The Temporal Dimension of Voting Rights (July 2006)
301. Adam B. Cox, Designing Redistricting Institutions (July 2006)
303. Kenneth W. Dam, Legal Institutions, Legal Origins, and Governance (August 2006)
305. Douglas Lichtman, Irreparable Benefits (September 2006)
306. M. Todd Henderson, Paying CEOs in Bankruptcy: Executive Compensation when Agency Costs Are Low (September 2006)
310. David Gilo and Ariel Porat, The Unconventional Uses of Transaction Costs (October 2006)
312. Dennis W. Carlton and Randal C. Picker, Antitrust and Regulation (October 2006)
316. Ariel Porat, Offsetting Risks (November 2006)
324. Wayne Hsiung and Cass R. Sunstein, Climate Change and Animals (January 2007)
<table>
<thead>
<tr>
<th>Issue</th>
<th>Article Title</th>
<th>Authors</th>
<th>Published in</th>
</tr>
</thead>
<tbody>
<tr>
<td>334.</td>
<td>Eugene Kontorovich, Inefficient Customs in International Law</td>
<td>March 2007</td>
<td></td>
</tr>
<tr>
<td>335.</td>
<td>Bernard E. Harcourt, From the Asylum to the Prison: Rethinking the Incarceration Revolution. Part II: State Level Analysis</td>
<td>March 2007</td>
<td></td>
</tr>
<tr>
<td>337.</td>
<td>Adam B. Cox and Thomas J. Miles, Judging the Voting Rights Act</td>
<td>March 2007</td>
<td></td>
</tr>
<tr>
<td>341.</td>
<td>Anup Malani, Valuing Laws as Local Amenities</td>
<td>June 2007</td>
<td></td>
</tr>
<tr>
<td>344.</td>
<td>Christopher R. Berry and Jacob E. Gersen, The Fiscal Consequences of Electoral Institutions</td>
<td>June 2007</td>
<td></td>
</tr>
<tr>
<td>349.</td>
<td>Lior Jacob Strahilevitz, Privacy versus Antidiscrimination</td>
<td>July 2007</td>
<td></td>
</tr>
<tr>
<td>351.</td>
<td>Lior Jacob Strahilevitz, “Don’t Try This at Home”: Posner as Political Economist</td>
<td>July 2007</td>
<td></td>
</tr>
<tr>
<td>361.</td>
<td>removed by author</td>
<td></td>
<td></td>
</tr>
<tr>
<td>366.</td>
<td>Lee Anne Fennell, Homeownership 2.0</td>
<td>October 2007</td>
<td></td>
</tr>
</tbody>
</table>
371. Lior Jacob Strahilevitz, Reputation Nation: Law in an Era of Ubiquitous Personal Information (November 2007)
375. Edward L. Glaeser and Cass R. Sunstein, Extremism and Social Learning (December 2007)
383. Richard A. Epstein, Decentralized Responses to Good Fortune and Bad Luck (January 2008)
384. Richard A. Epstein, How to Create—or Destroy—Wealth in Real Property (January 2008)
387. Cass R. Sunstein and Adrian Vermeule, Conspiracy Theories (January 2008)
395. Lee Anne Fennell, Slices and Lumps, 2008 Coase Lecture (March 2008)
398. Randal C. Picker, Take Two: Stare Decisis in Antitrust/The Per Se Rule against Horizontal Price-Fixing (March 2008)
400. Shyam Balganeshe, Foreseeability and Copyright Incentives (April 2008)
401. Cass R. Sunstein and Reid Hastie, Four Failures of Deliberating Groups (April 2008)
407. Cass R. Sunstein, Two Conceptions of Irreversible Environmental Harm (May 2008)
408. Richard A. Epstein, Public Use in a Post-Kelo World (June 2008)
410. Adam B. Cox and Thomas J. Miles, Documenting Discrimination? (June 2008)
412. Cass R. Sunstein, Trimming (June 2008)
413. Thomas J. Miles and Cass R. Sunstein, Depoliticizing Administrative Law (June 2008)
414. Randal C. Picker, Competition and Privacy in Web 2.0 and the Cloud (June 2008)