

COMPETITION AND PRIVACY IN WEB 2.0 AND THE CLOUD

*Randal C. Picker**

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, as such, were tended to with loving care and serviced by a small caste of computing priests. In contrast, PCs were everywhere: on every knowledge worker's desk and eventually in the family room of many homes. 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 performs 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—all things that we expect of our operating systems. But Windows is more than that: Windows delivers software. Before the advent of the Internet, software delivery was difficult. A consumer might find the software was pre-installed on a new PC. Alternately, the consumer could 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, a CD with new software.

Microsoft had a special role in software delivery because it could guarantee delivery by just incorporating the new software into Windows. With each new release of Windows—from Windows 3.1 to Windows 95 to 98 and on towards Vista—Microsoft expanded the footprint of Windows. This expanded footprint was not just a question of taking up more hard drive space; Windows got bigger because it expanded its functionality. In doing so, it killed off what had been separate markets in freestanding functions provided by other companies. Disk fragmentation was once a separate product category, but it wasn't anymore once Microsoft added that function

* Copyright © 2008, Randal C. Picker. All Rights Reserved. Paul and Theo Leffmann Professor of Commercial Law, The University of Chicago Law School and Senior Fellow, The Computation Institute of the University of Chicago and Argonne National Laboratory. I thank the Paul H. Leffmann Fund, the Sarah Scaife Foundation, and the Milton and Miriam Handler Foundation for their generous research support. I have done some consulting work in connection with some of the issues addressed in this Essay, but all of the views expressed are my own.

to Windows itself. Including a product in the next version of Windows insured its widespread distribution as each version of Windows quickly expanded its market share.

In a 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 boundlessness 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 upper-hand 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 Gates feared that users would no longer care what operating system ran on their computers; instead, consumers would care only about the browser that sat on top of the operating system. 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, although it did so in ways found to be illegal by competition authorities in the United States.³

But Microsoft seems to be losing the larger war suggested in the *Internet Tidal Wave*. We are in the midst of two large related shifts in our computing platform.⁴ The first shift, often called “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, watching the newest viral video on YouTube, and hanging out with our friends on mySpace and Facebook. The second shift, often called cloud computing, is a change in the organization of the fundamental processes of computing—computation and storage. These shifts are not fully independent; the cloud computing shift has some overlap with the

¹ See Randal C. Picker, *Pursuing a Remedy in Microsoft: The Declining Need for Centralized Coordination in a Networked World*, 158 J. INSTITUTIONAL & THEORETICAL ECON. 113 (2002); Randal C. Picker, *Unbundling Scope-of-Permission Goods: When Should We Invest in Reducing Entry Barriers?*, 72 U. CHI. L. REV. 189 (2005).

² Memorandum from Bill Gates to Executive Staff and Direct Reports at Microsoft 4 (May 26, 1995), available at <http://www.usdoj.gov/atr/cases/exhibits/20.pdf> (also commonly referenced by the title of the memorandum—*The Internet Tidal Wave*) (link).

³ *United States v. Microsoft Corp.*, 253 F.3d 34, 58–78 (D.C. 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 Case T-201/04, *Microsoft v. Comm’n*, 2007 ECJ CELEX LEXIS 554, 2007 WL 2693858 (Sept. 14, 2007), available at <http://curia.europa.eu/jurisp/cgi-bin/form.pl?lang=EN&Submit=rechercher&numaff=T-201/04> (under the “Cases” column and in the row with “Judgment,” follow the “T-201/04” link) (link).

⁴ See *infra* notes 5–7 and accompanying text.

Web 2.0 shift. Instead of storing my email on my laptop, I will just out-source 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.

Google dominates the current Web 2.0 market. And like Windows, Google's infrastructure has no obvious boundaries. Indeed, the Google engine is in many ways more powerful than Microsoft's. It is not obvious how the size of Windows or its functionality affected the price that Microsoft could charge for Windows. When Microsoft added browser functionality to Windows, there was no direct increase in 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. However, unlike Windows, Google's revenue scales as more services are added.

These two 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 it 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 much more valuable if it can be matched to my actual interests. The flexibility of the Web in delivering content means that Web advertising is increasingly taking the form of tailored advertising, or so-called behavioral advertising.

Determining how to regulate these datastreams is the central regulatory issue of the emerging computer infrastructure. We already frequently regulate the way traditional 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 as they process many of our transactions. Laws disable them—wholly or partially—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. We must consider how regulation of this transactional information affects competition, and if the privacy issues are different in the online space. Our choices here will not only have personal privacy consequences but also consequences on how much competition will emerge. Both consequences are tightly linked.

NORTHWESTERN UNIVERSITY LAW REVIEW COLLOQUY

Section I of this Essay describes the shift in the computing platform away from desktop towards Web 2.0 and cloud computing. Section II looks at two examples of “sticky” data and the way in which that stickiness operates as a barrier to entry. Section III focuses on how datastreams are used now, taking search as a prominent example. As Google moves from relying on publicly-available information for assessing relevance towards relying on collective intelligence, it will increasingly emphasize the rich information available to it in the datastream. So far, at least, Google is able to do that because it doesn’t face the same legal barriers on data use that others do. That difference in regulation limits competition that might otherwise emerge in how these rich datastreams are used. And we will need to tread carefully as we rethink how competition and privacy interact. Limits on disclosure of information across firms can have perverse effects as firms can circumvent those limits through horizontal or vertical integration.

I. FINDING DATA OR THROWING DATA AWAY?

We should start with desktop computing before the emergence of the Web. Microsoft Office 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 calendars, 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, matching and coordination are 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. The emergence of these Web intermediaries is one of the defining aspects of Web 2.0.⁵ 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, or given the number of pageviews that take place, the intermediary can support all of the content with advertising as Google does. With the ready ability to match advertising with content, a platform that generates pageviews is a valuable media property.

⁵ See Tim O’Reilly, *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*, O’Reilly, Sept. 30, 2005, <http://www.oreilynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (link).

But there is more. Intermediaries have the ability to see what is happening with every click, and this creates an incredibly rich clickstream.⁶ 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 because I search far more often than I buy or sell on eBay. This datastream arises organically as part of the services performed by the website. As a society, we can choose to limit the use or disclosure of this information, but throwing away the information requires a deliberate engineering design choice by the intermediaries. The information otherwise emerges naturally from the role played by the intermediary.

The idea of cloud computing is related to the emergence of Web intermediaries, but a little different. Computing power was first highly centralized with mainframes, and then decentralized through the switch to minicomputers and PCs. With the cloud, content and computing power will increasingly be managed centrally.⁷ The main 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 inflict harm on all of us when their power is harvested in botnets and computer spam is sent across the globe. Finally, 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 generating 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. We may be headed in 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 is a detail. If communications costs have dropped sufficiently such that we won't notice when the computing is done remotely, we can return computing power to the center. This is really just an engineering problem that turns on the relative costs of central as opposed to local processing power and on inter-computer as opposed to intra-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 how the datastream is used. Think of this as a version of cloud neutrality: where processing or storage is done should be irrelevant, or 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

⁶ See JOHN BATTLE, *THE SEARCH* 9–12 (Portfolio, 2005).

⁷ See generally NICHOLAS CARR, *THE BIG SWITCH: REWIRING THE WORLD, FROM EDISON TO GOOGLE* (2008).

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) that 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 but I “pay” for Gmail by being exposed to the advertisements that it places on the far right edge of the screen. Google filters my email to determine which ads to match with it. Send yourself an email on an obscure subject and see what ads Google serves up.

How we use the rich datastreams that have emerged under Web 2.0 and that may emerge under cloud computing is a point of both technical and legal design. Nothing about the change in the organization of computing requires a change in information revelation, that is, how much information a third-party sees from someone using its product or service. But, the change raises several questions about the use of such data: What happens to the datastream flowing through the chokepoint? How is that information used and controlled? How does that affect both competition and privacy?

II. DESIGNING STICKINESS AND DATA PORTABILITY

Control over the the datastream affects competition. Take an early Web 2.0 example: eBay. 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 are a long-standing problem in commercial law. eBay solves this problem by creating stickiness with its user reputation and feedback scores. 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 from 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 are not lost on eBay’s competitors. They understand the way in which those scores create entry barriers for auction competitors. Take the case of ReverseAuction.com. eBay’s original business relied on ascending price auc-

tions. 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.⁸

ReverseAuction almost certainly understood the competitive disadvantage it faced against eBay. According to the complaint filed by the Federal Trade Commission, to solve that problem ReverseAuction registered as an eBay user, agreeing to the eBay user agreement.⁹ ReverseAuction was then able to harvest 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.¹⁰

The FTC found ReverseAuction's actions troubling, but there was also a kernel of virtue in ReverseAuction's actions as eBay's control over user reputations 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.”¹¹ The natural take-away from this is that 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 tools that allow 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—and Google Reader to manage much of the information that flows through my computer. If you want to know what I am interested in right now, you could 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

⁸ Complaint, ¶7, *FTC v. ReverseAuction.com, Inc.*, Civ. Action No. 000032 (D.D.C. Jan. 6, 2000) [hereinafter *ReverseAuction Complaint*]. The complaint and other case documents for *Federal Trade Commission v. ReverseAuction.com, Inc.*, including the stipulated consent agreement and statements from FTC commissioners, are available at <http://www.ftc.gov/os/caselist/reverseauction/index.shtml> (link).

⁹ *Id.* at ¶8.

¹⁰ *Id.* at ¶12 (quoting the email sent by ReverseAuction to eBay users).

¹¹ eBay, Your User Agreement, <http://pages.ebay.com/help/policies/user-agreement.html> (last visited June 23, 2008) (link).

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 thousands.

The answer on the feeds is OPML. 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, it is possible, assuming that your RSS reader supports importing and exporting OPML.

We can count on competitors to help lower these switching costs. We saw that above with ReverseAuction. In another classic case, *Lotus Development Corp. v. Borland International, Inc.*,¹² 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.¹³ 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 there are limits on the ability of competitors to lower switching costs. Moving the list of feeds over through OPML is just one small piece of my information. The detailed matching of news stories and tags represents a much greater share of the value, and I don't see any particularly easy way to export that information into another RSS program. Indeed, we see how design matters when we look at transferring my tagged stories. I don't think much, if any, of my Google Reader info is stored locally on my machine. Therefore, there would be no locally stored information for the alternate RSS-tool 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" 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. Other times it will be antitrust. For example, the European Union has tried to force Microsoft to disclose more information to increase interoperability between operating systems and servers.¹⁴ In other cases, we will legislate portability

¹² 49 F.3d 807 (1st Cir. 1995), *aff'd* 516 U.S. 233 (1996) (per curiam). Justice Stevens took no part in the consideration or decision of *Lotus*, and the Court's vote was equally divided among the remaining eight justices. *Lotus Dev. Corp.*, 516 U.S. 233 (link).

¹³ See *Lotus Dev. Corp.*, 49 F.3d at 819 ("[T]he Lotus menu command hierarchy is uncopyrightable subject matter, [and therefore] Borland did not infringe Lotus's copyright by copying it.").

¹⁴ See William H. Page & Seldon J. Childers, *Bargaining in the Shadow of the European Microsoft Decision: The Microsoft-Samba Protocol License*, 102 NW. U. L. REV. COLLOQUY 332, 332-34 (2008) (link).

and interoperability, as we have done with telephone number portability¹⁵ and as parts of Europe may push Apple on iTunes and the iPod.¹⁶

III. CONTROLLING HOW DATA ARE USED

To see further how control of users' datastreams can implicate privacy and competition concerns, 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. This algorithm looks to the link structure of the Web to measure importance and therefore relevance.¹⁷ If we are looking for information about the Chicago Cubs and many pages link to a particular page about the Cubs, we might conclude that that is a particularly relevant page. The PageRank approach emphasizes information that is available publicly. Any entrant could do the same, assuming they could do so consistent with the original patent.

PageRank doesn't rely on the datastreams that arise in search. But we might imagine an approach that does so and 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.¹⁸ If that learning improved relevance, more searchers would seek to rely on the system, and that in turn would generate more learning. This positive feedback loop would operate as a barrier to entry because, unlike the publicly observable page-link information at the heart of PageRank, learning through search results relies on private information available only to the search engine.

These very different approaches to the use of the datastreams are available to Google and that is just in framing how relevance is assessed, the core function of search. The datastream could also be used in behavioral advertising to match the ads presented next to the search results with the searcher rather than just matching the ads to the text content of a page as

¹⁵ FCC, Wireless Local Number Portability (Wireless LNP) Frequently Asked Questions, <http://www.fcc.gov/cgb/NumberPortability/> (last visited July 25, 2008) (link).

¹⁶ See David S. Evans, *Antitrust Issues Raised by the Emerging Global Internet Economy*, 102 NW. U. L. REV. COLLOQUY 285, 285 ("Multiple jurisdictions in the European Community have claimed that Apple has violated [] competition laws . . .") (link).

¹⁷ See Our Search: Google Technology, <http://www.google.com/technology/> (last visited June 23, 2008) (link).

¹⁸ See Posting of Hal Varian to The Official Google Blog (Mar. 4, 2008, 7:04 PST) <http://googleblog.blogspot.com/2008/03/why-data-matters.html> (link).

AdSense typically does now.¹⁹ 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.”²⁰ 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 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 (CCPA) of 1984 added privacy protections for consumers.²¹ The current version of the CCPA 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.²² However, it 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.”²³ 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.²⁴ However, the statute also 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.”²⁵

How we implement privacy restrictions matters enormously, and indeed, the limits can sometimes have perverse consequences like undermining competition policy. For example, a disclosure limit of the sort seen in the cable statute artificially pushes towards vertical integration. A firm will have an artificial incentive to expand the size and scope of the firm so as to use the information fully because most disclosure limits do not prevent disclosure within a particular firm, but only bar disclosure across firm boundaries.²⁶ Vertical integration renders the disclosure limit ineffective. We might see mergers that would otherwise be unattractive as a way to circumvent the disclosure limits.

¹⁹ See Saul Hansell, *Google Tries Tighter Aim for Web Ads*, N.Y. TIMES, June 27, 2008, C1, available at <http://bits.blogs.nytimes.com/2008/06/26/google-tests-using-your-search-data-to-tailor-ads-to-you/> (titled online as *Google Tests Using Your Search Data to Tailor Ads to You*) (link).

²⁰ See Advertising Age, *The Advertising Century: John Wanamaker*, <http://adage.com/century/people006.html> (last visited Jul. 21, 2008) (link).

²¹ Cable Communications Act of 1984, Pub. L. No. 98-549, 98 Stat. 2780 (codified as amended at 47 U.S.C. § 551).

²² 47 U.S.C. § 551(b)(1) (2006) (link).

²³ *Id.* § 551(b)(2)(A).

²⁴ *Id.* § 551(c)(1).

²⁵ *Id.* § 551(c)(2)(A).

²⁶ See, e.g., *id.* § 551(c)(1).

Note also that disclosure may not be the act of relevance here. Google's ad placement service, AdSense, does not disclose any information to facilitate matches between content and consumers.²⁷ For Google's advertisers the information will be in a black box. These advertisers 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 of the datastream, just use on the advertiser's 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.

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 defines 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 they provide. 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. These intermediaries can use this information 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 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 such as banks, cable companies, phone companies 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 that allows 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

²⁷ See Google, Welcome to AdSense, https://www.google.com/adsense/login/en_US/ (last visited June 23, 2008) (AdSense instead crawls the content of subscriber webpages and posts ads related to the content of individual subscriber websites.) (link).

NORTHWESTERN UNIVERSITY LAW REVIEW COLLOQUY

across firms will artificially push towards greater consolidation, something that usually works against maintaining robust competition.