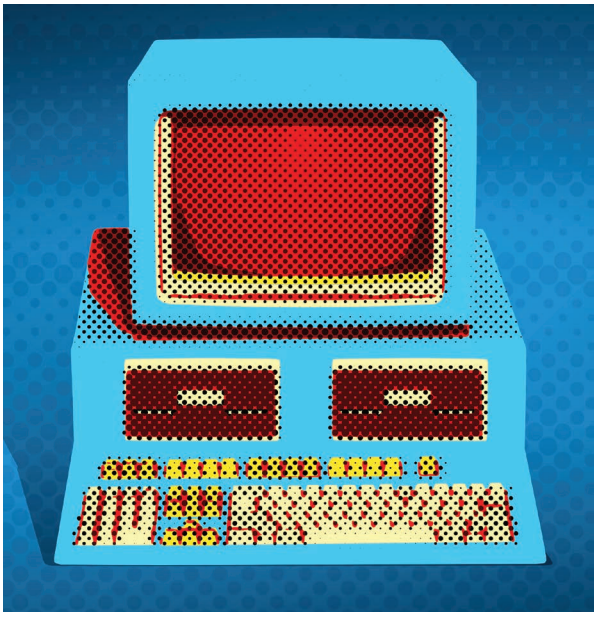

When the Cloud Goes Local: The Global Problem with Data Localization

Patrick S. Ryan, *Google and Katholieke Universiteit Leuven*
Sarah Falvey, *Google*
Ronak Merchant, *Level 3 Communications*

Ongoing efforts to legally define cloud computing and regulate separate parts of the Internet are unlikely to address underlying concerns about data security and privacy. Data localization initiatives, led primarily by European countries, could actually bring the cloud to the ground and make the Internet less secure.

Policymakers across the globe continue to try to define and regulate the cloud, even as it has become evident that there is no substantive difference between cloud computing and the Internet itself. Some proposed regulations are in response to citizen outrage about US government surveillance, while others are in the name of consumer protection.

Whatever arguments that governments and their service provider partners employ to convince users of the need for such controls, attempts to legally define the cloud and regulate separate parts of the Internet are unlikely to address underlying concerns about data security and privacy. In fact, recent governmental and commercial efforts to mandate data localization, primarily in Europe, could actually make the Internet less secure.

GROSCH'S LAW

The idea of large datacenters that serve users is as old as computing itself. In 1953, Herbert Grosch theorized that computing performance increased by the square of its cost and that relatively dumb terminals would tap into the power of large datacenters: "I believe that there is a fundamental rule, which I modestly call Grosch's law, giving added economy only as the square root of the increase in speed—that is, to do a calculation 10 times as cheaply you must do it 100 times as fast."[1] He predicted that cloud-style computing would result from the need to leverage economies of scale coupled with the need to invest in massive data processing centers.

Grosch's law was partly "repealed" by other technology laws such as Moore's law, which predicted the microprocessor revolution and the development of the personal computer.[2] But while Grosch's datacenter cost model was wrong, he was nonetheless correct in asserting that significant economies of scale and efficiencies could be achieved by relying on larger, centralized datacenters rather than on storage in individual end-user systems.

DEFINING THE CLOUD

Since the early 1990s, policymakers have been trying to break the Web into different pieces with different names and then pass new rules to control those pieces. This is exemplified by ongoing efforts to apply legal meaning to

 Published by the IEEE Computer Society

"cloud computing," even though it has never been anything more than a marketing term for the Internet.

The term originated with George Favaloro, a Compaq Computer marketing executive who in 1996 described the trend toward more intra- and intercompany connectivity, e-commerce, and the "Internet as information source."[3] Favaloro simply described what businesses were doing with computers and outfitted the Internet with a fancy new name—the cloud—and ever since, lawmakers have been obsessing over how to define and apply rules to it.

In 2011, the US National Institute of Standards and Technology defined cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." The definition specifies five essential characteristics (on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service), three service models (software as a service, or SaaS; platform as a service, or PaaS; and infrastructure as a service, or IaaS), and four deployment models (private cloud, community cloud, public cloud, and hybrid cloud).[4]

The NIST definition serves as the opening salvo in almost all policy discussions about the legal meaning of cloud computing, and various standards and regulatory organizations outside the US have likewise come up with their own definitions. However, according to the Software and Information Industry Association, "Because of [the] varying models and platforms, … 'cloud computing' resists practical definition for common treatment by law and regulation. Simply stated, cloud computing is not a single, unitary thing. There is no 'the cloud.'"[5]

Current efforts to define the cloud are problematic for two related reasons. First, as there is no clear difference between the cloud and the Internet itself, any attempt to create a legal distinction among various online services will invariably lead to legal "overreach" with unintended consequences. Second, forcing such a distinction is likely to mislead the very consumers that the legislation is intended to protect because they might wrongly think that a particular rule, regulation, or practice will protect them so long as the services they are using are labeled as cloud services.

CLOUD COMPUTING SERVICES

All cloud computing offerings are essentially two-way interactions on the Internet, and thus not distinguishable in a meaningful way.

Web-based email and word processing

As early as the 1990s, consumers were using services such as AOL Mail and Hotmail to compose email messages in their browsers and send those emails to family, friends, and colleagues. Although most regulators would think of Web-based word processing services such as Google Docs as cloud computing, few would probably recognize Web-based email as such. In fact, the data produced by both services are stored and managed in virtually identical ways.

Business and personal organizational tools

Thousands of cloud-based products and services are available to help individuals and businesses aggregate their Web-based data. Are productivity suites and online Rolodexes cloud computing tools, or are they something else? Take, for example, the Evernote family of apps (http://evernote.com), which lets users "easily collect and find everything that matters." Evernote aggregates all types of information from users and businesses—photos, documents, plane tickets, and so on—on its servers and makes this accessible from any browser or mobile phone anytime, anywhere.

All cloud computing offerings are essentially two-way interactions on the Internet, and thus not distinguishable in a meaningful way.

One of the most successful cloud computing companies is Apple, which offers artists a platform to store and distribute music, videos, and books as well as sells services to iPhone users through the App Store. A multitude of apps provide opportunities for file sharing and for creating and managing discussion boards, workspaces, and other kinds of collaboration. Does Apple's product line qualify as a SaaS, PaaS, or Iaas? These labels force distinctions that are not really meaningful to users. After all, iTunes is a service, the App Store provides a developer platform, and Apple provides a cloud storage infrastructure. From the user perspective, all of these services are "on the Internet," and consumers neither understand nor care about arcane technical or regulatory classifications of services and the implications—if any exist—for data security and privacy.

Online storage and sharing

These days, people are more likely to store and organize photos in the cloud—on sites like Picasa, Flickr, and Snapfish—than to put them in a physical album or shoebox. Other services such as Dropbox, Jungle Disk, Amazon S3, and Egnyte offer complete online file storage and sharing. Is there a tangible difference between these sites and offerings like email and online word processing? Regardless of whether a site provides infrastructure (hard drive space) alone or value-added services, users access the site and manage content in a similar way.

THE CHANGING NATURE OF DATA SECURITY

Consumers have been comfortable with these kinds of services for more than 20 years, so why the desire to label them as "cloud computing services" and attach new laws to their use now? A common refrain is that more safeguards are needed because more data is online and much of it is personal. Recent news accounts about large-scale hacking incidents and US government surveillance have exacerbated data security and privacy concerns.

Without question, the way that we handle information is rapidly changing, making data security more challenging than in the past.

Traditionally, users of computer-based word processing and spreadsheet programs like Microsoft Word and Excel, Corel WordPerfect, Quattro Pro, and VisiCalc saved the output of their work to a floppy disk, hard drive, thumb drive, or CD/DVD. The user—or the user's company—owned the computer and external storage media, meaning that data could be protected simply by locking up these devices. Physical assets needed only physical protection.

Perhaps the greatest advantage of storing data in the cloud is that information is typically sliced up and distributed among multiple systems rather than kept on a single machine or set of machines.

Perceptions about data safety began to change with the emergence of laptop computers, long before they were connected to the Internet. Almost overnight, employees began working on the go, away from their desks and often outside their offices. At first, synchronizing data between laptops and corporate servers was clunky and inefficient, requiring a manual connection in the office either through a cable or docking station. Consequently, data on an employee's laptop—including emails, contacts, and other intellectual property—often resided only on that device without any backup whatsoever and often without encryption or passwords.

According to a 2009 Ponemon Institute study, 60 percent of a typical corporation's data resides not in any kind of data vault but unprotected on PC desktops and laptops—giving hackers, thieves, and people who find lost computers relatively easy access to this information. A laptop is stolen every 53 seconds, and 1 out of 10 is stolen within a year of purchase. Even if data on these devices is encrypted or otherwise protected, it often contains confidential information and trade secrets that skilled thieves could extract. Moreover, at some point 66 percent of thumb drive owners lose their drives, 60 percent of which contain corporate data; these devices are also increasingly being used to infect networks by introducing worms or viruses into secure systems.[6]

The economic ramifications of lost and stolen laptops and portable storage media far exceed the price of the devices themselves—as high as $50,000 per device once forensic analysis, lost intellectual property and productivity, legal and consulting fees, and other costs are taken into account.[5] In short, data is expensive to store, maintain, and secure—and those who rely on physical storage devices are likely to lose data or leave it unprotected and vulnerable.

A major advantage of moving computing to the cloud is that outsourcing data security is much less expensive. Netbooks, tablets, and other mobile computing devices that access the cloud often do not even store most information locally. For example, a typical Chromebook has a 16-Gbyte solid-state drive—it has no hard-disk drive or CD/DVD drive; while it does have a USB port, core data is typically saved through the Internet using a service such as Google Drive. While there might be some loss of productivity associated with a lost or stolen Chromebook, the only monetary cost is that of the device itself, which runs between $200 and $300.

Perhaps the greatest advantage of storing data in the cloud is that information is typically sliced up and distributed among multiple systems rather than kept on a single machine or set of machines—a process known as data sharding. No single datacenter has all the information required to reassemble a given document, so if a datacenter is breached or destroyed in a natural disaster, the information itself is not compromised. Further, cloud providers can "obfuscate" data such that, even if it is not encrypted with keys, it can be impossible to read.

DATA SECURITY IN THE CLOUD: A SHARED RESPONSIBILITY

Data losses still occur in the Internet, of course, and they can be significant. The most common threats to data in the cloud involve breaches by hackers against inadequately protected systems, user carelessness or lack of caution, and engineering errors. In 2011, for instance, hackers penetrated Sony's servers in a series of attacks and absconded with data from tens of millions of its customers,[7] a phishing scam induced some of Google's Gmail users to inadvertently share their emails with other unauthorized persons,[8] and a security glitch at Dropbox caused by a code update enabled anyone to access customers' accounts for several hours by typing in any password.[9]

Obviously, Internet users must rely on service providers to combat most threats. However, even the most secure companies falter at times. Users thus have a responsibility to meet service providers partway and do what they can to be safe.

There will always be hackers, just as there will always be criminals who break into homes and steal jewelry and

other valuables. Homeowners determine what level of security they need based on many factors, including the prevalence of crime where they live—in some areas locking the doors is adequate, while in others only a sophisticated security system provides peace of mind. As Internet users attain higher levels of digital literacy, they will be able to make more informed choices about protecting their data just as they do about their physical belongings.

"Locking the door" on data can in most cases be as simple as using stronger passwords. However, users are also coming to recognize the importance of more complicated mechanisms to access Internet services, even if this means some loss of convenience. Two-factor authentication, for example, provides more assurance that hackers will not be able to acquire data using stolen passwords, but this requires some training and getting accustomed to new routines.

THE PUSH FOR DATA LOCALIZATION

Many Internet users and policymakers, particularly in Europe, have come to believe that data would be safer if it was stored locally or regionally.[10] While they are right to be concerned about data security and privacy, particularly in the wake of numerous high-profile hacking attacks and revelations by former National Security Agency (NSA) contractor Edward Snowden about the unprecedented extent of US government snooping, data localization is no panacea.

Requirements to localize data do nothing on their own to make data safer; in fact, they will only make it impossible for cloud service providers to take advantage of the Internet's distributed infrastructure and use sharding and obfuscation on a global scale. A recent paper from the International Trade Commission makes this point clearly: "Localization requirements are problematic for cloud providers, as 'location independence' is a core aspect of the cloud delivery model. Policies that require providers to locate facilities in a given location may leave them with the choice of selecting a suboptimal location or not serving the target market at all."[11]

The original local cloud: Minitel

France and Germany are leading the charge for development of local clouds in Europe, and have been doing so for at least two years. The roots are actually much deeper than that. Lest anyone claim that the cloud is a US invention, one need look no further than France's Minitel system to see the world's largest pre-Internet cloud computing deployment.

Deployed in the late 1970s and early 1980s, Minitel let the French public access numerous interactive databases through the country's telephone lines. Users could also make plane and train reservations, buy some retail products, and even send messages to each other. Because Minitel was a "pay for use" cloud-based system, France was wired much earlier than the rest of the world. In addition, the system's terminals had little or no computing power and were inexpensive to manufacture and maintain, so consumers were able to use them with little capital expense. Although the Internet has largely replaced Minitel, some of its terminals are still in use.

Thus, as long as 40 years ago, French consumers were using cloud services. Minitel brought tremendous value to the country's economy by enabling small and medium-size businesses to provide and exchange information. Notably, the government chose to leverage existing laws regarding print publications and audiovisual communications rather than create new regulations to deal with civil and criminal liability within Minitel, ensuring consumer protection while allowing the service to flourish. Of course, some hiccups occurred—for example, Minitel's "killer app" was erotic chat, which made the French government uncomfortable as the system's sponsor.

Requirements to localize data do nothing on their own to make data safer; in fact, they will only make it impossible for cloud service providers to take advantage of the Internet's distributed infrastructure and use sharding and obfuscation on a global scale.

Recent data localization initiatives

In December 2011, the German government announced development of a "Bundescloud" (federal cloud) based on the proposition that it would be safer than a cloud created exclusively through private industry.[12] In early 2012, the French government likewise announced a significant investment of €75 million in a public/private consortium to build Andromède, a French federal cloud.[13] And in August of this year, a former French finance minister called for a "Shengen for data"—a reference to the Shengen Agreement, which currently allows for the free flow of people within most European states. If enacted, such an agreement would mandate storage of consumer data within a "safe" geographical zone determined not by any particular legal distinction but by regulations designed to enable travel by car or foot without a passport.[14] (Presumably, data would not be safe in the UK or Ireland, which are not parties to the Shengen Agreement.)

This and other related proposals indicate a growing desire in Europe to create government-run or -subsidized clouds like the old Minitel system. However, while government investment in local clouds might be economically beneficial, it will not make data inherently safer. Moreover, all such ventures will undoubtedly involve commercial

activity in the US, and that raises a different kind of concern with which users outside the US should be familiar.

GOVERNMENT SURVEILLANCE

Not surprisingly, French and German plans for data localization are in large part a reaction to US surveillance activities, which were considerably expanded by the 2001 USA PATRIOT Act and associated laws. The expectation is that local clouds will protect citizens' data from the US government's prying eyes, but this is not necessarily true.

The question of international jurisdictional authority is complicated, and the US legal system's propensity to obtain information—computerized or not—located outside US borders in defiance of local laws long predates the Internet. In a 1976 federal case, *United States v. Field*, the court stated that it "simply cannot acquiesce in the proposition that United States criminal investigations must be thwarted whenever there is conflict with the interest of other states." Eight years later, in *United States v. Bank of Nova Scotia*, the court ruled that the US government could request information of any kind from a company as long as it had a subsidiary on US soil. In this particular case, a Canadian bank was forced to turn over a customer's records because it had a branch in the US. None of the records were stored in the US, and providing the information even violated the laws of the Cayman Islands and the Bahamas, where the records were actually kept.[15]

In June 2011, in the first admission of its kind from a senior representative of a cloud computing service provider, the head of Microsoft UK stated that any data stored by a US-based company, no matter where it is physically located, is subject to interception and inspection by US authorities. In fact, he said, no cloud computing service provider can even guarantee that customers would be informed of such action.[16] In a blog post two months later, Microsoft's chief counsel in Australia confirmed that this was the "reality" of the PATRIOT Act.[15]

However, the problem of unchecked government surveillance—in particular, the lack of due process protection for users—is not limited to the US. While publicly expressing outrage about NSA eavesdropping, European governments are just as interested in monitoring their own citizens' online activity, and their laws are equally problematic. The UK's Regulation of Investigatory Powers Act 2000 provides similar government access rights to data as the PATRIOT Act. In France, the Code of Criminal Procedure's Article 100 allows the government to engage in "special investigation techniques" under any circumstances if "the requirements of the investigation call for it" in a process that cannot be appealed. Italy has numerous similar laws, and, as revealed during the Amanda Knox case, between November 2007 and May 2008, Italian police reportedly conducted an astonishing 39,952 wiretaps for just one investigation.[17] Chapter 27 of the Swedish Code of Judicial Procedure allows for secret Internet wiretapping of any cross-border communication without a warrant or any court review.

Given the dearth of data on how governments use information obtained through surveillance methods, it is premature to assume that other governments—including European democracies with relatively good civil rights records—would be any more trustworthy or use any more restraint than the US government.[18,19] In fact, storing data in a particular jurisdiction might do nothing more than increase the number of state actors interested in accessing that data.

The bottom line is that governments, not private companies, are now emerging as the principal threat to data privacy. To be sure, businesses have a responsibility to protect their customers' data, but citizens also must hold their governments accountable.

Until recently, computing data was primarily stored in the machines themselves or on external media. Users and businesses felt confident that as long as they physically secured these devices, their data was also safe. However, as computing moves to the cloud—and data with it—there are growing concerns about data security and privacy. High-profile hacking attacks and reports of widespread government surveillance, primarily by US authorities, have motivated users and policymakers around the globe to call for data localization as a solution.

Yet, such an approach could actually make data less secure. Without the advantages provided by a distributed infrastructure, such as data sharding, data is more vulnerable to breaches and natural disasters. In addition, many local clouds are federally subsidized, and there is no evidence that providing data to a cloud service operated with government support is any safer than doing do with a private entity that has no such arrangement.

As cloud computing evolves, service providers—who have no desire to lose customers—will continue to develop more sophisticated data security and privacy technologies. These efforts must be complemented with greater due diligence on the part of users themselves and new laws that more aggressively curb government surveillance powers.

References

1. H.R.J. Grosch, "High-Speed Arithmetic: The Digital Computer as a Research Tool," *J. Optical Society of America*, vol. 43, no. 4, 1953, pp. 306-310.
2. C.W. Adams, "Grosch's Law Repealed," *Datamation*, vol. 8, no. 7, 1962, pp. 38-39.
3. A. Regalado, "Who Coined 'Cloud Computing'?," *MIT Technology Rev.*, 31 Oct. 2011; http://goo.gl/m4stls.
4. P. Mell and T. Grance, *The NIST Definition of Cloud Computing*, NIST Special Publication 800-145, Nat'l Inst. of Standards and Technology, 2011; http://goo.gl/0xWKdV.
5. "Guide to Cloud Computing for Policymakers," white paper,

Software & Information Industry Assoc., 2011; http://goo.gl/cLqu0.
6. "Business Risk of a Lost Laptop: A Study of IT Practitioners in the United States, United Kingdom, Germany, France, Mexico & Brazil," Ponemon Inst., 2009; http://goo.gl/3m739q.
7. J. Pepitone, "Massive Hack Blows Crater in Sony Brand," *CNNMoney*, 10 May 2011; http://goo.gl/07Kkf4.
8. A. Efrati and S. Gorman, "Google Mail Hack Blamed on China," *The Wall Street J.*, 2 June 2011; http://goo.gl/Nd0vcd.
9. B. Bosker, "Dropbox Bug Made Passwords Unnecessary, Left Data at Risk for Hours," *The Huffington Post*, 21 June 2011; http:/goo.gl/vX3u2H.
10. D. Hakim, "Europe Aims to Regulate the Cloud," *The New York Times*, 6 Oct. 2013; http://goo.gl/TZU0J3.
11. R. Berry and M. Reisman, "Policy Challenges of Cross-Border Computing," *J. Int'l Commerce and Economics*, vol. 4, no. 2, 2012, pp. 1-38.
12. J. Burke, "Innenminister Friedrich Will Bundes-Cloud Aufbauen," *WirtschaftsWoche*, 17 Dec. 2011 (in German); http://goo.gl/6AyDh0.
13. C. Deprez, "Cloud Andromède: 75 Millions d'Euros pour le Projet Orange/Thalès," *Génération Nouvelles Technologies*, 20 Apr. 2011 (in French); http://goo.gl/RehGo0.
14. "Atos CEO Calls for 'Schengen for Data,'" *Telecompaper*, 28 Aug. 2013; http://goo.gl/xKPVFO.
15. J. Bullwinkel, "The USA Patriot Act: Myth v. Reality," blog, 25 Aug. 2011; http://goo.gl/HgWtw.
16. Z. Whittacker, "Microsoft Admits Patriot Act Can Access EU-Based Cloud Data," *ZDNet*, 28 June 2011; http://goo.gl/0N48Tt.
17. D. Longhini, "We'll Be Listening: Amanda Knox Case Reveals Extent of Italian Wiretapping," *CBSNews*, 23 Nov. 2011; http://goo.gl/zrYaep.
18. W. Maxwell and C. Wolf, "A Global Reality: Governmental Access to Data in the Cloud," white paper, Hogan Lovells, 18 July 2012; http://goo.gl/FYgqWd.
19. J. McCallion, "Report: 'European Clouds Are Not Safe from Government Snoopers,'" *CloudPro*, 24 May 2012; http://goo.gl/oGLD4.

Patrick S. Ryan *is Public Policy and Government Relations Senior Counsel, Free Expression and International Relations, at Google and a senior affiliated researcher at Katholieke Universiteit Leuven, Belgium, from which he received a PhD in law. Ryan is a board member of the Broadband Internet Technical Advisory Group and was a founding board member of the International Telecommunications Research and Education Association. Contact him at pryan@pryan.net.*

Sarah Falvey *is Public Policy and Government Relations Manager, Free Expression and International Relations, at Google. She received an MA in international relations and an MPA from Syracuse University. Contact her at sfalvey@gmail.com.*

Ronak Merchant *is a network access planner at Level 3 Communications. He received an MS in telecommunications from the University of Colorado Boulder. Contact him at ronakmerchant@gmail.com.*

The authors have written this article in their individual capacities, and nothing herein should be interpreted as the opinions of their employers.