Data Tagging for New Information Governance Models

Data Tagging for New Information Governance Models

Paula J. Bruening
Centre for Information Policy Leadership, Hunton & Williams LLP

K. Krasnow Waterman
Massachusetts Institute of Technology

The ubiquitous collection, use, and flow of data challenge existing frameworks for data protection and management. Organizations collect and derive data from myriad sources and use it for a wide variety of purposes, so that the rules that apply to their data holdings vary. A company might use data for internal processes such as product development and accounting in one instance, and in another transfer that same data for processing by a vendor or business partner halfway around the world.

Although geography and national borders place few inherent limitations on where organizations can transfer data, such boundaries demarcate different and very real requirements and obligations for handling personal information. For owners and processors, moving data across these boundaries presents practical challenges in administering and implementing the rules and laws by which individuals maintain their rights to data protection and privacy.

Here, we describe data governance in this complex and dynamic environment, where the rules and obligations that govern how organizations use and protect information attach to the data and must be met wherever or by whomever it is collected, processed, or stored. We can facilitate such an approach via "tagging" data with sufficient information that its recipients and users can understand their specific obligations for its appropriate use and safeguarding.

Emerging Approaches to Data Governance

The emergence of nearly instantaneous collection, analysis, use, and sharing of data has prompted policymakers, privacy experts, businesses, and regulators to call for new approaches to securing and governing it. Various forums have highlighted current governance models' limitations. In its December 2009 "Opinion on the Future of Privacy," the Article 29 Data Protection Working Party expressed the view that the present legal framework hasn't been fully successful in ensuring that data protection requirements translate into effective mechanisms that deliver real privacy protection.[1] Its 13 July 2010 release proposes a legal system architecture that would integrate an accountability approach to data protection.[2] Organizations participating in the US Federal Trade Commission's (FTC's) "Exploring Privacy" workshop series emphasized current approaches' limitations in effectively protecting individuals in the emerging data environment.

Approaches to privacy protection that rely exclusively on "notice and choice" have come under significant criticism as being impractical and ineffective. In a notice-and-choice model, consumers receive information about how an organization will collect, use, and share data about them. On the basis of this notification, consumers choose whether to allow its use. Such a model breaks down in an environment in which organizations can analyze and process information instantaneously at the collection point, and where data collection has become so ubiquitous that individuals could receive privacy notices every time they connect to the Web, are monitored by surveillance cameras, use a mobile communications device, or visit a building that uses sensors. In many cases, notices are lengthy and complex, and don't inform any meaningful choice. Choice itself might now be illusory—at worst, inappropriate, and at best, giving the data custodian or controller helpful parameters for data use only in limited circumstances. Acknowledging this reality, commenters at the FTC "Exploring Privacy" workshops urged policymakers to look beyond notice and choice as the starting point for privacy protection. (For example, in response to the failure of fair information practices, Fred H. Cate argues for a more tailored,

less procedure-based privacy protection that includes "substantive restrictions on data privacy processing designed to prevent specific harms."[3] The Center for Democracy & Technology, in contrast, argues for grounding privacy protection in a more comprehensive iteration of fair information practices that incorporates principles beyond notice and choice.[4])

New models proposed for information protection and privacy reflect and respond to the realities of 21st century data collection, analytics, use, and storage. These approaches realistically take into account where notice is effective and where individual choice and control are appropriate and real. They reflect information's role as a critical business asset and the challenge of responsibly managing data within organizations. Such models include accountability;[5] the application of fair information practices based on data use, rather than its collection;[6] and a comprehensive system of securing and managing data referred to as *strategic information management*.[7]

These approaches recognize that if data protection and management are to be effective, the obligations to protect and secure data attach to the data itself and must be met wherever it's stored or processed. They also rely on the ability to tag data with information about those obligations, so that all relevant parties can understand and meet them. Such obligations might arise from law and regulation, self-regulatory guidelines and best practices, and the promises organizations make to individuals about how they will protect and responsibly use those individuals' data. For example, when the fictional online retailer BuyWeb collects data from customers to fill an order, deliver goods, facilitate internal processes such as billing and accounting, and provide customer service, this data collection might be governed by one or more laws, self-regulatory guidelines, and privacy promises. BuyWeb is committed to fulfilling those governance obligations. When it makes data available to an outside vendor—for instance, to process billing or respond to customer inquiries—the requirement to meet those obligations doesn't end; the vendor must also follow the applicable rules.

Imagine that a BuyWeb customer moves from Tokyo to Los Angeles or London. BuyWeb notes the move and enters the address change into its customer database. The address change means that the individual's home jurisdiction and the laws that apply to his or her data have also changed. BuyWeb must first determine whether the new or old jurisdiction's rules apply to previously collected data and then both apply the correct rules in its own systems and ensure that its business or process partners do the same.

Organizations have also begun to appreciate data's full value as a critical business asset and to take a comprehensive approach to protecting it. Companies understand that they should safeguard and manage data in ways that not only protect individuals' privacy but also ensure data's integrity and availability for a wide range of uses within the company. BuyWeb will want to use the customer's change in address to accurately market weather- or culture-related products. Different co-branding or supply-chain partners will likewise wish to capitalize on the updated information.

Data must also be available when called for in judicial and legal proceedings, an increasingly complex problem as jurisdictions have developed apparently contradictory requirements.[8,9] For example, a customer service representative might appropriately look at a customer's address to verify a caller's identity or determine if a shipping address matches company records. That same representative might be precluded from seeing credit-card information if not taking an order. New approaches to data protection within companies involve setting rules about data access, use, storage, and retention, and ensuring that employees follow those rules as data flows throughout the organization.

To facilitate these new approaches to data protection and management, data protection obligations must *attach to and travel with the data*. Individuals must be able to rely on the law, best practices, and the company's representations about its data practices, no matter who processes that data, or when. Users and data custodians must understand and follow the rules that govern who may use data within the organization, in what ways, under what circumstances, and to further what ends. Third-party data processors must be able to understand what requirements they must meet and the specifications about how they may use data. These approaches would guarantee that individuals receive protection in a decentralized, networked data environment, where they might have no knowledge of, and little choice about, the actual party or parties handling their information.

Accountability

An accountability principle has been a feature in both the earliest major international instrument on privacy—the Organization for Economic Cooperation and Development's Privacy Guidelines[10]—and the most recent—the Asia Pacific Economic Cooperation (APEC) Privacy Framework.[11] Both require that the information owner or data controller "should be accountable for complying with measures that give effect" to the fair information practices articulated in the guidelines.[10,11]

Efforts are currently under way to define the contours of accountability and explore the conditions

that an organization must demonstrate and that regulators must measure to certify accountability. Policymakers, regulators, and experts have described an accountable organization as one that sets privacy protection goals for companies based on external criteria established in law, self-regulation, and best practices, and vests the organization with the ability and responsibility to determine appropriate, effective measures to reach those goals. Given that the complexity of data collection practices, business models, vendor relationships, and technological applications in many cases outstrips individuals' ability to make decisions through active choice about how their data is used and shared, accountability requires that organizations make disciplined decisions about data use even absent traditional consent.

Accountability's essential elements are organizational commitment to accountability and adoption of internal policies consistent with external criteria; mechanisms to put privacy policies into effect, including tools, training, and education; systems for internal, ongoing oversight and assurance reviews and external verification; transparency and mechanisms for individual participation; and means for remediation and external enforcement.

As an accountable organization, BuyWeb might establish an internal privacy and data management policy consistent with both local laws and regulations and the promises about privacy it makes to consumers. Under an accountability approach, BuyWeb would also implement mechanisms to ensure that employees adhere to those policies and systems for internal risk assessment and mitigation, including oversight and assurance reviews. Those systems would govern how the organization handles information internally. BuyWeb might also use an outside vendor located in Vietnam to provide customer service and address complaints about products or billing. In this case, the rules that govern the data apply even when the outside vendor is doing the processing. BuyWeb will have to ensure that the vendor is committed to and capable of meeting these obligations.

In another example, BuyWeb might wish to avoid addressing cross-jurisdictional legal requirements as much as possible and might thus create an internal policy to limit the receipt of customer data outside each individual's home jurisdiction. It might implement this policy in part through mechanisms that look for clues (IP address or telephone area code) about where an incoming customer request is coming from and route it to a service representative in the same jurisdiction. The organization would later provide validated reporting about its performance, perhaps including the numbers or percentage of employees trained on the policy in the prior year, or of requests successfully routed according to the policy.

Central to an accountability approach is the organization's ongoing assessment and mitigation of the risks inherent to individuals from information use. In the case of the routing-service-requests-to-matching-jurisdiction example, the retailer would also capture and analyze the incidents that didn't comply with the policy and attempt to identify modifications to the practice or technology to improve future performance.

Use-and-Obligations Model

The use-and-obligation model establishes data use rather than its collection as primarily driving users' obligations to protect and safeguard information. Collecting data and consumer consent to or choice about its use traditionally have triggered an organization's obligations. In this model, however, the mere fact that an organization collects information from a customer wouldn't typically trigger an obligation. Instead, this would occur only, for example, if the company used the customer's address to confirm his or her identity or direct a package delivery. The use-and-obligations model proposes a framework for implementing and interpreting traditional principles of fair information practices that addresses how companies can use and manage information in the 21st century. It incorporates the full complement of fair information practices, including transparency and notice, choice, access and correction, collection limitation, data use minimization, data quality and integrity, data retention, security, and accountability.

The use-and-obligations model takes into account all uses that might be necessary to fulfill the consumer's expectations and meet legal requirements. It imposes obligations on organizations based on five categories of data use:

1. fulfillment activities necessary to establish and maintain the relationship between the organization and consumer;
2. internal business operations and processes necessary to operate a business, such as accounting, product development, and personnel management;
3. marketing;
4. fraud prevention and authentication; and
5. national security and legal requirements imposed by courts and government.

In our BuyWeb example, checking a customer address to confirm identity would fall under use number 4 and to direct a package would fall under use number 1. The obligations based on these uses that apply to the data must be met even if the data is shared or processed by a third party.

Strategic Information Management

Strategic information management is an integrated approach to managing data across an enterprise to minimize risk and enhance competitive opportunities.[12] It envisions not simply protecting personally identifiable information but all information assets. It recognizes that information is a critical business resource and appropriately protects and manages data in a way that facilitates the organization's compliance with legal requirements and minimizes the risk using that information might raise to the company and its customers. Managing information strategically requires that companies make decisions about data that ensure that it's available to the appropriate personnel when needed, and fosters new and creative use that can add value for the organization and consumers.

For example, an organization might decide that to protect its data resources, it will adopt a policy-based access control system, a method that restricts access to data based on predetermined rules. Under this broad umbrella might be rules about handling information that are designed to protect trade secrets, others implementing privacy law, and still others ensuring that the organization meets fiduciary responsibilities. For instance, BuyWeb's competitiveness might be based on a cheaper cost of goods than its competitors; its company policy might treat the sources of goods as a trade secret and protect that high-value data by limiting access to its suppliers' identities to those people who negotiate acquisition terms or receive the goods at the port of entry. BuyWeb's implementation of OECD guidelines might prohibit access to individual customer data to anyone in the accounting department, except individuals directly addressing customer complaints and corrections. And, perhaps, BuyWeb has decided to centralize fulfilling its statutory obligations to file sales tax payments in all the countries where it operates, allowing only assigned workers in the corporate tax office and auditors access to the tax calculation and payment data. These access rules serve a different purpose but share a common structure: people with a particular responsibility are permitted to access particular data for a particular purpose.

Practical Considerations

Each of these new models relies on individuals' and organizations' responsibility to handle data—whether at rest, in transition, or in motion, and whether in a centralized or decentralized environment—in accordance with rules. These rules about handling information fundamentally share a common structure—they describe a policy (such as, permit, require, or prohibit) about whether an entity (a person, organization, or system) may use particular data (data type, subject, provenance, and so on) in a particular way (collect, copy, merge, share, delete, and so on) under certain circumstances. Consider some policies we've described:

- The entity called customer service is permitted to use data about a customer's address to verify identity.
- The company's computer systems are required to route customer service requests to customer service representatives in the same jurisdiction.
- The company is prohibited from addressing a package to an address not in the customer's profile.

Data custodians' ability to ensure that their organization follows all necessary rules depends entirely on their ability to identify the data, the actor, the transaction, the circumstances, and some means to associate those factors with the rules that govern them. Although we can perform such identification manually, the volume of data and transactions has made human review an impractical approach to the challenges; computer-assisted review is now required. Systems can recognize such data (about actors on the data, about the data itself, or about the actions and circumstances) if it's annotated, or tagged.

Computer systems aren't human clones. They can't consistently glean meaning from whole sentences nor independently implement complex logic. Even so, privacy rules can be incrementally implemented in digital environments by reducing the text to something that looks more like an algebra problem:

- IF (Entity called "Customer Service") AND (Data category "Customer's Address") AND (Purpose of Use is "Verify Identity"), THEN Permitted.
- IF (Data category "Shipping Address") NOT SAMEAS (Data category "Customer's Address"), THEN Prohibited.

This is how programmers write instructions that computers can understand. They identify categories of information that are relevant to the business activity (such as "entity," "data category," and "purpose of use"). Depending on the rule, the programmer might pre-define the only things that can be placed in that category or permit other people or systems to put anything in that category. If the data in a system is tagged to identify such categories, then a computer can gather the necessary information to implement policies.

If all information necessary for implementing a privacy rule existed in a single database, then tagging might not be so important. To understand why, consider a corollary from the pre-digital world: a business might have kept a customer's records in a file folder tabbed with the customer's name. Inside

a customer's file, the company might place a name, address, and account number, but the file typically wouldn't include the name or job duties of everyone who ever opened the file, put something in, or took something out. Nor would it include a list of questions the business had used the file to answer. But, even the simple rules we just described require information about the data in the database and data outside it—who's trying to use the information and why.

Typically, laws and contracts are even more complex. They have conditions and exceptions that might in turn have conditions and exceptions. They require knowledge about information sources, the date and time of acquisition, the proposed information recipients, the rules that applied to the data before the data holder received it, and many other facts not ordinarily collected in either the old-fashioned paper file folder or a typical digital data file. As entities tag these other sorts of data—data about provenance, transactions, associated rules, and so on—organizations can implement increasingly complex, automated or semi-automated rules processing. They can automate rules regulating acceptable information use, appropriate data protections, and transparency and accountability, and they can increasingly validate how consistently rules are applied, even after the data changes hands, purposes, or forms.

New approaches to governance attempt to respond to the new information environment, where data collection can occur in circumstances where traditional notice and choice might not be possible, sharing and analysis might happen in real time, and processing might take place outside the jurisdiction where information was collected. Data tagging offers a practical way to digitally attach obligations to information and reap the benefits of these new protection models. Legacy data systems raise important cost issues for organizations contemplating data tagging. While a growing market of products reduce those costs, policymakers and organizations will need to strike the appropriate cost-benefit balance as they consider this important path forward toward data protection that will serve the 21st century digital environment. □

References

1. "The Future of Privacy," *Joint Contribution to the Consultation of the European Commission on the Legal Framework for the Fundamental Right to Protection of Personal Data*, Article 29 Data Protection Working Party, 2009; http://ec.europa.eu/justice_home/fsj/privacy/docs/spdocs/2009/wp168_en.pdf.
2. "Opinion 3/2010 on the Principle of Accountability," Article 29 Data Protection Working Party, 2010; http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf.
3. F.H. Cate, "The Failure of Fair Information Practice Principles," *Consumer Protection in the Age of the Information Economy*, J.K. Winn, ed., Ashgate Publishing, 2006.
4. "Refocusing the FTC's Role in Privacy Protection," *Comments of the Center for Democracy & Technology in Regards to the FTC Consumer Privacy Roundtable*, 6 Nov. 2009.
5. "Data Protection Accountability: The Essential Elements, a Document for Discussion," *The Galway Accountability Project*, Oct. 2009; www.ftc.gov/os/comments/privacyroundtable/544506-00059.pdf.
6. "A Use and Obligations Approach to Protecting Privacy: A Discussion Document," *The Business Forum for Consumer Privacy*, 7 Dec. 2009.
7. P. Bruening et al., "Strategic Information Management," *Privacy & Security Law,* Bureau of Nat'l Affairs, vol. 7, no. 36, 2008.
8. *In re Advocat "Christopher X,"* Cour de Cassation, no. 07-83228, 12 Dec. 2007.
9. *United States v. Vetco, Federal Reporter*, 2nd Series, vol. 691, 1981, p. 1281 (US 9th Circuit Court of Appeals).
10. "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data," 1980; www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html.
11. "APEC Privacy Framework," 2005; www.ag.gov.au/www/agd/rwpattach.nsf/VAP/(03995EABC73F94816C2AF4AA2645824B)~APEC+Privacy+Framework.pdf/$file/APEC+Privacy+Framework.pdf.
12. "Strategic Information Management," *Privacy & Security Law,* Bureau of Nat'l. Affairs, Sept. 2008.

Paula J. Bruening *is Deputy Executive Director of the Centre for Information Policy Leadership at Hunton & Williams LLP in Washington, DC. Her work focuses on cross-border data flows, emerging technologies, privacy accountability, and cybersecurity issues. Bruening has a JD from Case Western Reserve University School of Law. She recently spoke at the US Federal Trade Commission's "Exploring Privacy" workshop. Contact her at pbruening@hunton.com.*

K. Krasnow Waterman *is a visiting fellow at the Massachusetts Institute of Technology's Decentralized Information Group, prototyping accountable systems and launching a course on creating linked-data ventures. She's both a technologist and lawyer, has been a systems manager at JP Morgan, the inception CIO of a large federal counterterrorism task force, in private practice with Brown & Bain, and a Special Assistant US Attorney. Waterman has a JD from Cardozo School of Law. Contact her at kkw@mit.edu.*