Taming the Wild West of Big Data
The Social Norms and Laws of Social Data

Kelly T. Slaughter, PhD
The University of Texas at Dallas

Gavin D. George
Haynes and Boone, LLP

A white paper sponsored by the Metroplex Technology Business Council and written for the Dallas Business Journal

June 2013
As each new invention is discovered, it is necessary for the courts and the legal profession to consider to what extent the existing laws will suffice to develop the new legal relations brought into existence by the use of the invention. Whenever it is possible, the conservatism of jurisprudence prefers the application by analogy of existing rules to the formation of new ones. This may, however, be carried too far, and injustice done through the unwillingness to establish new rules specifically adopted to the circumstances arising out of the new invention.

– C.G. Tiedman in The Central Law Journal, writing in 1881 about the commercial transition from mail to the telegraph

The Opportunity and the Uncertainty

The introduction to society of new technologies often results in an initial Wild West period of opportunity and uncertainty. Consider the rapidly evolving technologies of social media and personal digital devices. These technologies are facilitating the creation of “Big Data,” an immense volume and variety of dynamic social data. Organizations and entrepreneurs are anxious to “stake their claims” to this Big Data to more deeply engage with their existing consumers and connect with potential consumers. For instance, imagine that after you muster the discipline to bypass the Blue Bell Vanilla Cherry ice cream while grocery shopping, you receive a text offering a special “just for you” price for said item on the way to the checkout lane. Or imagine that your new “digitized” golf bag identifies three other golfers to join you for a round that afternoon, and the clubhouse, alerted to the formation of your foursome, offers a free drink, via cell phone, to the player with the best score.

During the Big Data version of the Wild West period, there is significant uncertainty concerning the appropriate capture and use of social data. The resolution of this uncertainty will shape the payoffs that Big Data opportunities offer. The legal frameworks meant to address this uncertainty for Big Data are still being deliberated. Justice Holmes offered, “the life of law has not been logic; it has been experience.”¹ In other words, the law will evolve as the culture makes sense of new technologies, concurrently informing and being informed by the social norms emerging to resolve conflict. Indeed, as Robert Ellickson describes in his book Order Without Law, social norms may not only arise in advance of the law, but may serve as the preferred arbitrator after the law is established. Consider the backlash Amazon experienced when monitoring buyer behavior to enable individualized (perfect) price discrimination. Though no legal remedy was sought (and no law broken), the public perceived this act as a violation of a social norm (perhaps more a violation of the transparency of the action rather than the actual act of price discrimination).³ Amazon quickly issued a public apology and promised to discontinue this practice.

In this white paper we discuss emerging areas of contention regarding Big Data and identify the nascent remedies emerging around the acceptable capture and use of this social data.

State of the Social Data Field

Existing privacy laws fail to address certain concerns about Big Data for several reasons. Contributions to Big Data are typically not created as a result of a usual consumer-to-business transaction, but rather through the private act of an individual. The contributed
data becomes accessible to the general public by design of the data host, or indirectly through others who quote or repost from the original source, or even capture and share screenshots. Certain social data may be captured, with or without explicit permission, through use of an application (e.g., Web browsing) or by possession of a device (e.g., location tracking on a smart phone). Big Data may also include data about individuals not involved in its generation, such as information pertaining to an individual’s location and participation in an event as posted on Flickr or Instagram.

These differences suggest that former legal norms for the capture and use of data, such as the expectation of privacy, are difficult to apply to these new technologies. For instance, consider the case where law enforcement officials obtained a cell phone from a drug dealer and viewed texts to identify additional drug-dealer meetings. The officials argued that no warrant was needed for reading the texts because texts can be viewed by anyone in close proximity to the mobile phone; thus there is no expectation of privacy. The Electronic Frontier Foundation, a technology rights advocacy organization, countered that texts are more analogous to phone calls, thus requiring a warrant to view the texts. Or consider a recent lawsuit against Facebook concerning the sharing of user information with advertisers when users clicked on an ad. In this lawsuit, the affected Facebook users argued for an analogy of wiretapping to evaluate the users’ expectation of privacy. The lawsuit was dismissed, with the judge suggesting that the wiretapping analogy was flawed.

The difficulties presented by these two cases are the tip of the iceberg compared to the bewildering array of new concerns introduced by the evolving technologies associated with Big Data. A Google Glass app, named “Winky,” has been developed that enables the wearer of Google Glass to take a picture based on a wink, allowing the undetectable recording of those near the wearer. The concern in this case is the inability of those present to detect the capture of their face and information, as well as the propagation of this information immediately to those viewing the social media posts of the Google Glasses wearer.

Casinos have banned recording devices to deter cheating (e.g., card counting) and for the comfort of their patrons. Consistent with these purposes, some casinos have already taken a position against the wearing of Google Glass on their premises. Though the banning of Google Glass in a Seattle bar was later acknowledged to have been a PR stunt, it is not difficult to imagine other stakeholders who may have similar concerns. When apps using Foursquare data began offering location targets by gender (e.g., finding a bar near the app user with the most women), Foursquare changed its data feed to only allow views of others if both are checked into the same location, noting that “real life” provides the same transparency.

There is no precedent for the new combination and aggregation of social data. An organization may capture the data related to a potential car buyers shopping for cars online, combine this data with personal information from social media, and deliver the combined data to an auto dealer to prepare for the shopper’s arrival. Many Facebook apps share personal information with third parties, including information on the app user’s friends. In many cases, multiple sources of anonymous data can be assembled to identify individuals. Big Data is increasingly being used by hiring companies to screen prospective employees.
Occasionally, Big Data is explicitly shared by the original creator with the public. An intriguing example of the capture and use of public data is the website “PleaseRobMe.” Using a combination of Twitter and Foursquare feeds, the site identifies when an individual has indirectly announced that he or she is not home (e.g., by identifying the restaurant he or she is frequenting at the moment). Though the site is meant as a satire, how might the misuse of Big Data reflect back on Twitter and Foursquare? One might argue that persons implicitly announce when they are not home by going out to a public restaurant, yet the nature of Big Data makes these physical-world comparisons suspect. The existing lack of legal remedy against PleaseRobMe seems inadequate.

A natural extension of these concerns is Big Data created and shared for misleading purposes. One example of this is creating “fake” twitter accounts directed towards political ends. Another example is the use of “fake” user accounts on websites to post product reviews. Such sites are undermined when authors review their own works (or the works of others) under pseudonyms or when commercial entities pay for positive reviews.

As Big Data technologies evolve, the possibilities for created value expand, and new dilemmas are introduced. For instance, researchers at The University of Texas at Dallas are examining how social media sites can directly capture commercial value from individual likes and shares on social media sites. How might an agreeable division of value be achieved between user and host? We next share how user expectations interact with the productive evolution of Big Data technologies to the satisfaction of all stakeholders.

### A Peak Into Emerging Social Data Use Solutions

When the telegraph was a new technology, the courts sought to resolve disputes by referencing legal precedent under a post office mail framework. In fact, Samuel Morse unsuccessfully sought to have the government purchase his telegraph patent because he believed the telegraph was a logical extension of the federally managed mail service.

There were, of course, several distinctions between the telegraph and the mail that organizations and courts needed to address to give legal certainty to the propagation of the new technology. For instance, unlike the mail, the telegraph required a third party to transmit and decipher a message, introducing a new potential source of error. Consequently, some telegraph service providers offered the resending of the original message at half the price of the original to address this concern. Technological standards also arose to track lines that were down and to allow the reporting of delayed messages. These standards facilitated the adoption of the new telegraph technology.

Today, Big Data stakeholders, individually and collectively through trade associations, are similarly addressing the user expectations, potential sources of error, and new standards. Below are several examples that give a sense of the direction of the Big Data conversation.

### Vendor Self-Policing

A benefit of many commercial websites to both consumer and vendor alike is the collection of product reviews. The value of product reviews hinges on the credibility of the author. Consequently, many sites provide some form of validation of the authors of reviews. For instance, Amazon is validating both reviews and authors through badges. The “Amazon
Verified Purchase badge indicates that a product reviewer did indeed purchase the product under review (through Amazon). Amazon also offers a “Real Name” badge for reviewers who use their actual name as author of the review. Online travel sites such as Expedia are similarly validating that reviewers did indeed stay at the hotel under scrutiny. Dallas-based Traxo allows members to maintain a travel profile across websites, aggregates the reviews of its members, and provides similar “consumption” validation. Twitter protects its celebrity and commercial users through the actual verification of identity, and provides a badge to those users validated (though this form of validation is not available for the general user). In contrast to positive validation, Yelp boosts the credibility of its reviews through the identification of the beneficiaries of “fake” reviews. If an organization is found to have manipulated reviews for its benefit, Yelp will place an alert on the offending organization’s Yelp page for a ninety day period.

While seeking to continue the collection of Big Data, Google allows users to know what data is being captured (see www.google.com/ads/preferences). Some advertisers have followed Google's lead. Zappos, when displaying an ad, includes a link to an explanation of how the user was selected to see the ad, what data served as a basis for this match, and how to opt out of the ad. In an effort to influence social norms around an emerging technology, Google has created “etiquette cards” for Google Glass, indicating where the glasses may be inappropriate, how to wear the glasses, and how to indicate that the glasses are not recording.

**Technical Solutions**

The issues above are a result of technological progression, so it is not surprising that some of these issues may be solved through further technological advances. The automatic expiration, sometimes within seconds of its posting, of social data before it becomes Big Data is a relatively recent offering. Snapchat allows photos to be sent with a limited duration, and TigerText allows for the expiration and the recall of instant messages.

Technical options to avoid the collection of Big Data at the individual level are also gaining traction. For instance, Disconnect and Ghostery allow the tracking and optional blocking of third-party collection of Web browsing information. Recent versions of Web browsers are similarly enabling a “Do Not Track” feature for the same purpose.

Other technical solutions include true anonymity when interacting digitally. For instance, the Electronic Frontier Foundation supports the effort of creating true digital cash, working in the same manner as “real world” cash, in that it is recognized as legal tender for online transactions, but is not associated with a specific individual.

Technical solutions are also being applied to certain parts of Big Data already available to the public. For instance, Amazon has created proprietary algorithms to delete what are determined to be “fake” reviews of products. Similarly, Yelp applies algorithms to sort reviews, presenting suspect reviews toward the end of the list. These algorithms may flag multiple reviews from new reviewers for a single product or various products made by a single brand, reviews written by those who join shortly after a product is available, and reviews more extreme than other reviews.
Professional Associations and Third Parties
Rather than struggle alone, some organizations are looking to their trade associations to establish standards to advance the trust and credibility of Big Data. Prominent among these organizations is The Wireless Association, an international nonprofit organization supported by members such as AT&T, Verizon, and Alcatel-Lucent. Chief among its concerns is establishing standards regarding the collection of location data to assuage consumer concerns and support adoption of location services. Standards this association supports include notifications to the device owner that location data is being collected (even if it is aggregated in a manner to make individuals anonymous), destruction of any collected information after it has served its intended purpose, and anonymization of data (e.g., through aggregation).26

Similar standards have been adopted by the W3C group, the international Web standards organization with IBM, Microsoft, the Mozilla Foundation, and Sony among its several hundred members. This organization supports the “Do Not Track” option in browsers, arguing that it should be actively selected by the user rather than default to no tracking.27 Advertising associations also support this position,28 while Mozilla supports the default of no tracking.29

Professional associations outside the technical domain are also stepping up to preserve the integrity of Big Data as it pertains to their product. For instance, the Crime Writers’ Association expels members who have used fake on-line identities to praise their own work or deride the work of others in the profession.30 The Electronic Frontier Foundation publicly ranks organizations on their “Who Has Your Back” ratings website, evaluating an organization’s social data policies such as privacy legislative support and maintenance of a public policy on social data.31

The Law and Norms
As mentioned in the opening passage, the law does not linger behind new technologies for long. Instead it interacts with emerging norms, both informing and being informed by technological changes. As with trends in Big Data, frequent and important developments are arising in the law as a result of new cases, legislation, and regulatory guidance on social data.

Legal issues will become increasingly important to Big Data in light of increasing privacy lawsuits and new privacy standards from industry trade groups and government agencies. A Big Data issue that will continue drawing legal attention is how companies track users’ browsing histories, location data, and personal information. As the use of personal digital devices increases, privacy concerns and privacy regulations will play an important part in online and mobile application development.

In the United States, use of Big Data is governed by a patchwork of federal and state laws, government agencies, private lawsuits, and industry standards. Federal privacy legislation
is not comprehensive across industries, but instead focuses on Big Data within specific sectors, such as the Gramm-Leach-Bliley Act (financial), the Children’s Online Privacy Protection Act (minors), and the Health Insurance Portability and Accountability Act (medical). Multiple bills proscribing more comprehensive Big Data reform have been presented to Congress for consideration, but none has been passed yet.

The Federal Trade Commission (“FTC”) is the most active federal government regulator of Big Data and privacy. In 2009, the FTC published “Fair Information Practice Principles,” a document that encouraged companies to address five core principles of privacy: (1) notice and awareness, (2) choice and consent, (3) access and participation, (4) integrity and security, and (5) enforcement and redress. The FTC does not enforce violations of these principles directly, but pursues questionable Big Data collection and use practices under Section 5(a) of the FTC Act, which prohibits unfair or deceptive trade practices. The FTC may bring an action against an organization if its Big Data gathering and sharing practices do not match its privacy policies or are otherwise “unfair” in the eyes of the FTC.

In early 2012, the Obama administration unveiled an outline of privacy principles that the administration expects Internet companies to adopt. The outline includes seven principles related to personal data: (1) Individual Control, (2) Transparency, (3) Respect For Context, (4) Security, (5) Access and Accuracy, (6) Focused Collection, and (7) Accountability. The Obama administration expects the FTC will use these principles as a basis for its enforcement actions.

The Children’s Online Privacy Protection Act of 2003 sets privacy rules for websites and mobile applications that knowingly collect information about children under the age of 13. Health care companies with access to Big Data should know that the Health Insurance Portability and Accountability Act of 1996 covers health plans, health care clearinghouses, and health care providers that transmit any personal health information in electronic form, as well as any companies that provide services to such entities.

European Union governments generally recognize privacy as a fundamental human right and the European Union Directive on Data Protection of 1995 prohibits the transfer of personal data to non-European Union countries that do not meet an “adequate” level of privacy protection. Current U.S. thought leadership regarding personal data is similar to that in the EU Directive. Jeffrey Rayport in MIT Technology Review suggests four principles in a Big Data Code of Ethics. These principles are summarized as:

- Clarity on Practices: Let users know immediately what data is being collected
- Simplicity of Settings: Avoid fifty pages of “legalese”
- Privacy by Design: Let users easily activate features that can block third-party ads and content
- Exchange of Value: Explain how collected data benefits the user

Through emerging norms, trade association standards, technical advances, vendor actions, and the law, increasing order will eventually be brought to the Wild West of Big Data, protecting the individual while preserving the benefits of social media.
Authors

Kelly T. Slaughter is the director of the UT Dallas Jindal School of Management Center for Information Technology and Management. Contact him at kelly.slaughter@utdallas.edu.

Gavin D. George is an associate in the intellectual property practice group of Haynes and Boone, LLP, specializing in the areas of data privacy and technology transactions. Contact him at gavin.george@haynesboone.com.

References

4. https://www.eff.org/cases/washington-state-text-message-privacy-cases
8. http://www.forbes.com/sites/kellyclay/2013/03/10/seattle-bar-that-banned-google-glasses-admits-it-was-a-pr-stunt/