User Data on the Social Web: Authorship, Agency, and Appropriation

Jessica Reyman

The Internet of the second decade of the twenty-first century promises no less than the opportunity for democratic discourse, access to the world’s knowledge, ever-present connection to social networks, and the ubiquity of tools for creating and sharing online. Those who deliver popular technology services assert that such services will revolutionize communication, information, and the world. Facebook, for instance, states that its mission is “to give people the power to share and make the world more open and connected” (About), and Google aims “to organize the world’s information and make it universally accessible and useful” (Company Overview). These statements claim that social Web technologies serve all people, everywhere, around the globe, with the goals of providing access to information, enabling connections with others, empowering individuals, and facilitating participation in powerful and meaningful ways. And, in fact, many people are using the social Web with these goals in mind: lifelong learners use the Google search engine to find useful information; amateur videographers use YouTube to share their creations; everyday critics use Tumblr or WordPress to compose and share content; and citizens worldwide use Twitter or Facebook to network with others professionally and personally, and to engage citizens and even organize revolutions. Users are frequently amazed at the power and ease of social and participatory Web services to deliver content and manage information, and, for the most part, to do so free of cost to users.

Currently, college and university students are among the most prolific social media users on the Web. According to a 2010 Pew Internet & American Life Project

Jessica Reyman is associate professor of rhetoric and professional writing in the Department of English at Northern Illinois University. Her teaching and research interests include technical communication, digital rhetoric, social media, and authorship, copyright, and intellectual property. Her work has appeared in CCC, Technical Communication, Technical Communication Quarterly, and several edited collections. She is the author of The Rhetoric of Intellectual Property: Copyright Law and the Regulation of Digital Culture (Routledge, 2010).

Copyright © 2013 by the National Council of Teachers of English. All rights reserved.
study, 72 percent of those online aged eighteen to twenty-nine use social networking sites, with 45 percent doing so daily (Lenhart, Purcell, Smith, and Zickuhr). We cannot ignore that our students rely on social Web technologies when writing, researching, and performing other academic tasks. A student can rarely complete a class assignment without conducting a search on Google, using Yahoo! to send an email to a professor, soliciting feedback from friends on Facebook, reading excerpts from texts on Google Scholar, or drafting notes on a WordPress site. In fact, in some cases we teach with select social and participatory Web technologies in the classroom, requiring that students join social networks or write to class blogs or wikis that are hosted by corporate entities and over which instructors and university officials have little control. These technologies offer much in the way of free or inexpensive tools for communicating, composing, and learning, but students, professors, and university officials often have limited understanding of the hidden practices surrounding the management of user data on the social Web.

For access to many social Web technologies there are trade-offs: tacit agreements that users enter into, and a set of unspoken assumptions that governs who owns what is created and how it circulates. We should not forget that when users access, read, network, post, or compose within many online spaces, they are simultaneously giving up information about a wide range of their online and offline activities and, ultimately, giving up control and ownership of their contributions. Social Web services catalog users’ individual and collective activities across the Internet—aggregating, analyzing, and selling a vast array of data in a practice known as data mining—to be used largely for consumer profiling and target marketing. Further, user data that is applied toward commercial ends is often not intentionally or voluntarily given. Although users are aware of the content they are contributing online—when sharing a photo, writing a blog post, updating a status, or entering a 140-character tweet—many are unaware of the additional, hidden contributions of data made with each act of participation. Yet these additional contributions of user data are inextricable from each and every act of composing via social and participatory Web technologies. With every click and path followed, every status update and tweet entered, every photo and post contributed, every comment, every item tagged, users are collectively producing both the visible and the invisible social Web.

In some instances, user data mining can be conducted by responsible parties in ways that benefit a community of users. California’s community college system, for instance, has introduced a new social technology platform, Sherpa, which relies on user data, data aggregation, and user feedback to recommend courses, university services, and other personalized information to students via email, text message, and the Web (Bramucci and Gaston). A 2011 Chronicle of Higher Education article written by a California State University–Channel Islands university police lieutenant, however, sheds light on the potential dark side of user data mining for university
Lieutenant Michael Morris argues that university officials should mine students’ online data in efforts to mitigate violence on campus. He posits that officials should collect information from students’ Google searches (such as “online research about the personal life and daily activities of a particular faculty member”); Facebook activity (namely, “angry and threatening comments on his Facebook wall about that professor”); Amazon purchases (like the purchase of “high-powered firearms and ammunition”); and file management (for example, a saved “draft version of a suicide note on his personal network drive”). Although most universities are capable of tracking students’ online activities, as many students access the Internet with a unique identifier via a university network, the suggestion of collecting and analyzing students’ private emails, Web searches, Facebook pages, and Amazon accounts—especially without students’ awareness—will strike many as an overreaching form of surveillance, an invasion of privacy in the name of campus security. But a similar type of data collection is what many private, corporate entities are doing every day, and with less altruistic goals. Students who use social and participatory Web services—for both academic and personal reasons—often do so without knowing that data about them is being appropriated for a wide variety of purposes.

In seeking to understand the complexities surrounding the use of social Web technologies for writing, teaching, and learning, we need to recognize the full range of productive activities that occur in social and participatory Web environments, including the production of content but also the often-overlooked generation of data. In the field of English studies, we have explored in some depth the ways in which social and participatory technologies affect authorship of content online (for example, DeVoss and Porter; Herrington; Laquintano; Logie; Reyman; Sano-Franchini), but we have not fully examined the related, inextricable act of generating data as part of composing processes. An exception is Heidi McKee’s “Policy Matters Now and in the Future,” in which she argues that corporate data mining and government surveillance online are two of the most important issues shaping the future of writing on the Internet. While noting that there is still hope for a different future, McKee warns, “By 2020 the Web could be completely controlled by corporations with what user-generated content there is being packaged by commercial sites and closely watched by government agents” (277).

Johndan Johnson-Eilola agrees with McKee that the increasingly complex relationships between humans and social Web technologies—and the records of those relationships in the form of aggregate data—are important for critical study. He notes that as a field, we should reflect on the “different, still-emerging set of possibilities in order to identify the positive and negative aspects” of such relationships (49). If users’ online activities generate data about our every move online, and, further, create connections among data generated by other users, we should consider ways in which digital writers and readers are affected by and perhaps complicit in the
agreements governing how that data is appropriated and applied. These discussions introduce space for exploration of the connections among technological structures that compel users to give up control of their contributions, the policies that dictate ownership of user data, and the nature of composing online. As users produce data through interactions within social and participatory Web services, they are acting neither as autonomous composing agents nor as mere bystanders in the generation of a technology by-product. Rather, they are working collaboratively with other users, texts, and technologies to create content, form networks, and, ultimately, write the social Web.

This article explores the implications of practices of data mining for intellectual property on the social Web. This exploration seeks to push the field toward new inquiries about what it means to author online: How do we write on the social Web? How do we write ourselves through the social Web? And how do we write the social Web? Although the social Web appears to be a space in which Internet users are empowered to create, share, and connect, current data-mining practices have the potential to transform our composing activities into data to be appropriated and used for privatized and commercial ends, commodifying each social, creative, and intellectual act. As social media platforms and technology companies record our activities and discover our preferences, our interests, and our desires, they simultaneously transform these contributions into data products to be appropriated for commercial purposes. User data, I argue, is not merely a technology by-product to be bought and sold; rather, it forms a dynamic, discursive narrative about the paths we have taken as users, the technologies we have used, how we have composed in such spaces, and with whom we have participated.

The debate surrounding user data is, in large part, a debate about the ownership status of a particular kind of online information. Social and participatory Web technology companies that collect, use, and sell user data understand it as valuable property to which they have justifiable legal rights. They point to their investment in offering a free service for user participation and creating the technological means for generating data and mining it for meaning. User data, then, is not seen as created by an author-user, but rather as a by-product of technological algorithms and aggregation formulas. The means, terms, and applications of user contributions are not controlled by the authors of those contributions, but by technology companies that seek to harness them for commercial ends. Many current data-mining practices and policies, which equate user data with corporate-owned property, are problematic because they favor a structure of information flow that discourages users’ control over what they contribute and threatens the future of the social Web as “co-created by and for the community of connected users” (O’Reilly and Battelle 1). Current data-mining practices prevent author-users from controlling the creation, distribution, and uses of their work. In fact, social and participatory Web users may not even be
consistently aware of their data contributions, which are the central good produced by their more overt compositions in social media spaces. Many users remain largely unaware that they are contributing these invisible compositions of data trails, and fail to realize the power and control technology giants have over their work. Social Web companies have managed to quietly wrest control from users without a battle having ever been waged.

The mismanagement of user data can have real and often severe consequences for individual users (including many university students), for the future of the Internet, and for the future of citizenship and culture. Data-mining practices, for instance, can lead to social implications, such as revealing personal purchasing and viewing habits to unintended audiences (boyd; Cheng; Duhigg; Perez; Singer). It can lead to a culture of surveillance on college campuses, in which the ethics of monitoring students’ online activities seem to go unquestioned (Hurley). In a global context, monitoring users’ searches, content creation, and activity in social media spaces can have political implications: restrictions on access to information and services, punishment and imprisonment of dissenters, and violent crackdowns by authoritarian governments (Ghonim; MacKinnon; Shirky). The data-mining practices involved in the recording, collecting, and monitoring of users’ activities can have real consequences within personal, social, political, and cultural spheres. Considering data mining in terms of authorship and ownership illuminates important implications for the field of English studies in particular, revealing both practical concerns for users (including ourselves, our colleagues, and our students) and theoretical implications for the concept of intellectual property online.

**Data Use Policies in Social Media**

The generation, collection, and use of user data occur on many levels on the social and participatory Web. On one level, data generated by individual users enables user tracking: the activities of individual users can be traced so that companies are able to deliver customized goods, services, and advertisements. Consider Amazon.com, which offers personalized recommendations for books, music, and many other products based on users’ browsing histories and purchasing habits. Data produced by tracking individual users can also be compiled into massive data sets, along with data from other users of the same service, to mean more to companies and advertisers. For instance, Target Corporation uses data gathered from large numbers of its consumers to predict, with a surprising degree of accuracy, not only who might be expecting a baby but also the due date of the child (Duhigg). In turn, Target is able to advertise to parents-to-be at a critical point in their consumer lives. Such predictions about consumer spending can be made based on seemingly innocuous purchasing details, such as buying cotton balls and unscented lotion, drawn from massive data
sets resulting from tracking many users across many purchasing activities. Additionally, user data can be mined from many users across services and Websites to be sold to third parties through data aggregation and analysis. Acxiom Corporation, a database marketing company, collects and sells what a *New York Times* reporter calls the “consumer genome,” consumer data analyses that reveal useful information about consumers’ past, current, and potential purchasing activities as well as about their ethnic backgrounds, health issues, and religious preferences (Singer). In this way, data mining provides a benefit to private companies by revealing who users are, how they use services, and what motivates those uses—information that can then be interpreted to improve the design, development, delivery, and sales of services and products.

This vast array of information is collected with such ease and finesse that users are often unaware of what they are contributing. The generation, collection, and use of data occur with a surprising lack of transparency. Terms-of-use policies that describe data collection and use are required by law, but these are lengthy and difficult to understand when read at all. Even more problematic is the fact that everyday users are often led to believe that the data they contribute is advantageous to them. Technology companies argue that Internet users see a benefit from the generation and release of data in that they receive increased access to the tools for participation, personalized services, enriched online experiences, useful advertising and promotions, contextualized searching, and nuanced interactions among humans and technologies. Google, for instance, attempts to convince users of the benefits of data mining by saying that collecting personal information from users enables Google to make “services even better—to show you more relevant search results and ads, to help you connect with people or to make sharing with others quicker and easier” (“Privacy Policy”). Social Web services claim to tame the chaos and filter massive amounts of information into something manageable, creating “a unique universe of information for each of us” (Pariser 9). User tracking and data aggregation are often posited as adding value to Web content and services for users themselves, despite the fact that users have limited understanding of the underlying technological workings and little control over terms of use. It often comes as a surprise to users to find out that their information is being appropriated and applied toward commercial ends, breaching users’ expectations concerning privacy and information flow (boyd). Users can feel unsettled when they believe that “filters” might be manipulating their online activities and making decisions for them—when advertisements appear to be too specialized and too targeted, when search results seem too narrow, when technologies appear to be too smart at determining their interests and behavior.

Facebook is a notable example of a social networking company with a long (relatively speaking) history of managing tensions between technology development and user expectations about control and ownership of user information. Since its inception
in 2004, Facebook has had its fair share of debacles surrounding its management of user data. Facebook’s history is well known. Created by undergraduates at Harvard, Facebook spread quickly to all universities and colleges in the United States and many globally. The service enabled users to create profiles of themselves that included photographs and personal information such as academic major, club membership, interests and hobbies, and other content to be shared with new acquaintances and friends. It also offered various modes of socializing and connecting with others, allowing users to friend one another, browse others’ profiles, and “poke” each other.

Facebook expanded to allow high school students in 2005, to include those with a working email address in 2006, and to allow anyone with access to the Internet in 2007. The site’s features change quickly, but several existing modes of participation have stirred debates about user control, including (1) the news feed, which automatically displays users’ activities to everyone in their network on the initial sign-in page; (2) tagging, which allows users who post links, events, photographs, and other content to tag them with the names of people and link them to users’ profiles; (3) connections with external applications, which share users’ personal information with third-party sites and services when accessed through a Facebook account; and (4) social plug-ins and instant personalization, which allow services that have partnered with Facebook to access some of a user’s information and friends’ information in order to make their sites “social.” Facebook offers a vast array of modes of participation and interaction that have revolutionized social networking, but the development of the service has met storms of protests at various points in its history.

Facebook’s data use policy states, “You own all of the content and information you post on Facebook, and you can control how it is shared through your privacy and application settings” (“Information We Receive and How It Is Used”). However, this policy also claims a right to collect and use what it identifies as “data,” which the policy treats differently from “content” and “information.” According to Facebook’s terms of use, “data” includes a plethora of information types: name, email address, birthday, gender, status updates, photographs, comments, friends, groups, likes, recommended links, tags identifying people or locations, searches, browsing activities, messages, metadata from uploaded content, information about the devices used to access the site including IP address, information from advertising partners about shopping activities, and information from partner services. Facebook also notes that it aggregates users’ data with that of their friends and others they connect with through the service. Additionally, Facebook reserves the right to access, save, and share users’ information in response to legal requests, to detect and prevent potential criminal behavior, or for other security reasons (“Some Other Things You Need to Know”).

Although Facebook’s data use policy is one of the more comprehensive policies I’ve encountered, it presents a dizzying array of types and uses of information it collects (and potentially shares), and it likely does little to clarify for most users
the flow and ownership of information. And it’s important to note that I’ve chosen Facebook’s policy to explore as an example not because it is unique among other social networking services, but because Facebook has attempted to present a comprehensive explanation of what have come to be common practices in the management of user data by many social Web services.

Facebook explicitly states that users own their content, but it assumes ownership over what it calls user “data” when it claims the right to collect, control, and use the vast array of information types just noted. Many believe that the appropriation of user data is unproblematic because there is a general lack of concern for control over user data among users themselves, who are choosing to share their activities, their likes and dislikes, their photographs, their beliefs, and their behaviors with their large networks. A common justification for corporate ownership of user data is that “privacy is dead” in an age of social networking. Facebook founder Mark Zuckerberg makes this claim himself in a 2010 interview:

People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people. That social norm is just something that has evolved over time. [...] We view it as our role in the system to constantly be innovating and be updating what our system is to reflect what the current social norms are. (Kirkpatrick, “Facebook’s Zuckerberg”)

Zuckerberg claims that he is merely responding to social norms in the development of new technologies and the policies that govern them. According to Zuckerberg, there is little harm in this type of data appropriation by social networking services; in fact, users understand this kind of data management as acceptable behavior in the age of social media. However, recent debates surrounding Facebook’s changing practices and policies reveal that Zuckerberg’s implicit claim that “privacy is dead” may not accurately reflect everyday users’ perceptions.

Despite the fact that user sharing on social media sites is increasing, users are nevertheless resistant to giving up control of their contributions. Among the most well-known examples of public outcry occurred in response to Facebook’s 2006 introduction of its news feed feature, which surprisingly displayed to all users’ Facebook friends their activities within the system (boyd). Although protests over the news feed blew over quickly and the feature is now fundamental to the workings of the social networking site, following it came a more intense resistance to another Facebook feature, Beacon, in 2007. Beacon was an advertising system developed by Facebook that tracked—and reported to friends via the news feed—the consumer activities of users on third-party sites, including Blockbuster, Zappos, and forty-two others (Perez). The release of Beacon sparked a public debate and a class action lawsuit in 2009, and, in a blog post from Zuckerberg in which he reflects on the design and release of Beacon, he admits that “we’ve made a lot of mistakes” in the introduction
of the new feature (Zuckerberg). Then came the Facebook privacy debate of 2010, the largest public protest over Facebook to date, when the company changed many of its default settings to make personal information more openly available and to limit control by users (Kirkpatrick, “Facebook Privacy Debate”). Although many users complained and some cancelled their accounts, by 2010 Facebook had become too valuable a social networking tool for most users to abandon it. In these examples, Facebook appropriated user data in ways unintended by users themselves, in ways that failed to meet their expectations about control over their information. And users’ collective outrage reveals their resistance to giving up control of their data contributions within the social media site.

In public debate, the data ownership issue is often framed as one about user consent or user choice: policymakers note the importance of users giving permission to companies that collect and use their information, and of users being able to opt out of data collection activities if they so choose. However, the issues of consent and choice are not universally understood and ethically resolved. Facebook’s Data Use Policy, for instance, does not discuss at length the issue of users giving consent; in fact, the only mention of user consent for the appropriation of data is in this statement: “Granting us this permission not only allows us to provide Facebook as it exists today, but it also allows us to provide you with innovative features and services we develop in the future that use the information we receive about you in new ways” (“Information We Receive and How It Is Used”). It is not clear how permission is granted or even what “this permission” specifically refers to; indeed, further reading of Facebook’s policies reveals that the company appears to assume ownership over users’ data as it is created, and that users appear merely to be granting pro forma permission to use it. Facebook Terms and Policies states simply that these (and others) are “[t]erms you agree to when you use Facebook.” User consent, therefore, does not require users to take an active role in understanding their rights and responsibilities or take a specific action to provide consent, but rather requires only that users sign up for and use the service. Further, even if users do seek out and read the terms of use, often they do so without full understanding of the complex Web of interrelated terms that affect the ownership and use of their information.

A 2008 study conducted by Carnegie Mellon University researchers regarding the difficulty of reading and comprehending lengthy and complicated terms-of-use agreements found that, based on an analysis of a large sample of policies, it would take an average user seventy-six workdays per year to read all privacy policies on every Website visited (McDonald and Cranor). And what is potentially more problematic for users is that terms-of-use policies may change at a moment’s notice. Technology service providers reserve the right, as Facebook does in its terms of use, to make changes to agreements—and they often do, which can leave users even more confused. Whether users have been fully informed of terms of use is debat-
able, and although service providers claim to collect and use information only with users’ consent, interpretation of what constitutes consent is questionable. McKee questions the ethicality of this structure for consent, calling it “manufactured and illusionary” (284).

The concept of choice is similarly misleading when describing data use practices. Facebook’s Data Use Policy is introduced with the directive to users, “Get to know the privacy settings that help you to control your information on facebook.com.” Users are told that they have choices surrounding their information, that they can customize settings to control the flow of their information. Users may be comforted by the notion of options and settings: customization options give the illusion of user ownership and control over their own data, suggesting that users have the right to retain some ownership over some information types. However, default settings offered by technology companies largely determine the use of social media services by the vast majority of users. In order to have the capability to choose different settings, users must be aware of the default settings, what their implications are, where to locate their controls, and how to change them. User choice requires that users understand how the different settings affect their use of the service and which settings to implement to get specific results. In noting the problematic default settings in Google, and the nature of “choice architecture,” Siva Vaidhyanathan argues that “meaningful freedom implies real control over the conditions of one’s life. Merely setting up a menu with switches does not serve the interests of any but the most adept, engaged, and well-informed” (Googlization 88–90). Because many users lack the understanding or the technical know-how to make different choices, default settings often result in most users allowing their data to be appropriated and used by technology companies, except in rare circumstances for the most informed and technologically adept users. The issue of choice is further complicated among college student users who are asked (or even required) by their professors or college administrators to complete assignments or perform other academic activities using social Web technologies that are not of their own choosing.

This overview of Facebook’s practices and policies regarding user data gives a glimpse of a larger trend toward diminished user control over information produced within and collected by social and participatory Web services. The terms of use and default settings express a particular ideology concerning data flow that privileges technology companies’ ownership and control rather than favoring truly flexible platforms that offer user customization and control. Terms of use are based on agreements made between unequal partners, between uninformed users and powerful companies that hide much of their activities from users. The problem surrounding user data is not caused by how much people are sharing (“users are revealing too much”), or even by the type and nature of information they are willing to share (“if users are concerned about privacy, it’s because they have something to hide”). Nor
should we resign ourselves to this situation (“we’re all being tracked online all the time, so who cares?”). Based on an ambitious research study about human-technology interactions, Sherry Turkle reports on the general lack of privacy and resulting anxiety that adolescents, teenagers, and young adults feel when online. Arguing for the value of a more user-controlled space for communication, interaction, participation, and reflection, she warns social media users that they should not turn a blind eye to data-mining practices and “live the fiction,” pretending “as if the shadow were not there rather than simply invisible” (260–61). Turkle’s study supports the idea that a solution to the data use problem on the social Web requires more than users’ self-discipline or a change in user behavior to simply “be discreet” when online.

The issue of data collection and appropriation on the social and participatory Web is, fundamentally, about the terms of control over user-generated information. It is a problem of information authorship and ownership: from what source does information originate, where does it circulate, what is its reach, and how is it appropriated and used? The policies governing the control and ownership of user information play a determining role in the nature of online participatory and productive activity. Dànielle DeVoss and James Porter note that participating, communicating, and creating online all take place within particular “economies of writing”:

Economics has to do with money, but not only money. It has to do more broadly with value, exchange, and capital; with production and consumption of goods; with giving, receiving, and sharing; with purpose, desire, and motivation; with the distribution of resources, products, and services; and with the systems of understanding that people rely on when they engage in such activities. Writing—*all* writing, we would say—resides in economic systems of value, exchange, and capital. (194)

The appropriation of user data creates a particular economy for writing, and establishes a particular value and exchange system for user contributions on the social and participatory Web. Current data-mining practices create an unequal exchange between unequal partners. Although the public debate regarding data mining has focused on the concepts of consent and choice as touchstones for fair and ethical practice, more readable terms-of-use policies and finer-grained settings and options for users will produce limited results. Instead, by considering data as coauthored with technologies, texts, and other users, we can reconceptualize user data as dynamic, living texts rather than as technology by-products to be bought and sold.

**Authorship, Agency, and Appropriation of User Data**

User data contributions in social and participatory Web environments are often handled, as is the case with Facebook’s data use policy, as the property of the technology company that offers the service. Opponents of strengthened user privacy rights
think of data as a product of companies’ investments in developing technologies that generate the data, compile it, aggregate it, and mine it for specific ends. Technology providers, from this perspective, have discovered and made good use of a resource that previously had little value, perhaps not unlike discovering and mining untouched natural resources. The data itself isn’t viewed as the result of human creativity or effort. Technology companies and data brokers, in this sense, are taking something that has little or no value as separate, individual data points and creating something of commercial value through aggregation and interpretation. Such appropriation is based on the assumptions that data as property is separable and unique from individual users’ creative activities on the social and participatory Web and, further, that data is a technology-generated by-product. As authorless objects, the argument goes, data has little value until it is aggregated and transformed into massive data sets and applied toward target marketing and customization of services.

When justifying practices of data mining, terms-of-use policies make a clear distinction between two types of property to be owned: user content and user data. Legal ownership of user content is governed by copyright law, and specific rights regarding ownership and distribution are assigned through terms-of-use agreements established by technology providers. Facebook, for instance, assigns copyright ownership of “content that is covered by intellectual property rights, like photos and videos,” to users, but stipulates that by posting content to Facebook, users are granting Facebook “a non-exclusive, transferable, sub-licensable, royalty-free, world-wide license to use any IP content that you post on or in connection with Facebook” (Statement of Rights and Responsibilities). This statement gives Facebook the legal right to use content such as photos and videos on Facebook in certain ways, for example by displaying it on a user’s Timeline and in others’ news feeds, and in connection with Facebook services. The conditions of use grant Facebook the right to use and display a user’s content to others (as is typically the purpose of posting photos and other items on Facebook), but stipulate that the user retains the right to share and profit from distributing the content elsewhere.

However, user data is governed under different terms. Although the site owners do not make an explicit claim of ownership of user data, Facebook freely appropriates and uses it for self-serving purposes. Facebook confusingly states that “you always own all of your information,” but also that “you are allowing us to use the information we receive about you” (“Information We Receive and How It Is Used”) for a range of unstated purposes and without compensation. Facebook reserves the right to use the information for the operation and development of its services, a right granted when a user signs up for and uses Facebook. A claim to appropriate user data, then, is not based on intellectual property law but, rather, on the assumption of a meaningful distinction between the content and data contributed by users (and on manufactured user consent).
This distinction between original works of authorship (or “content”) and data is not unfounded. One hesitation I have when discussing the management of user data as an IP issue results from the fear that establishing property rights to user data could erroneously lend more support for exclusive ownership rights to uncopyrightable facts. US copyright law grants authorship to particular expressions—“original works fixed in a tangible medium of expression”—and not to objective facts and data. Though copyright law does protect some compilations of facts, it applies only to works that “entail a certain degree of creativity” and “are sufficiently original” (Feist Publications v. Rural Telephone). The distinction between data and content is central to the assertions of ownership made by technology companies. Understanding data as objective facts, and, further, as unowned property, works to remove users from a creative role in its production. What such a distinction neglects to address, however, is that user data is a form of user contribution that is inextricably bound to an author-user’s content creations: the generation of data is tied—by user, time, activity, technology—to the production of content, and vice versa. When posting a photo, for instance, a date, time, and location are automatically attached to it, and when generating data about your networks and activities, status updates and comments are composed. Based on the asserted distinction between content and data, users maintain the rights to content shared on participatory Web platforms, but site owners unquestioningly assume control over the data attached to it. Such a distinction fails to recognize the interactive nature of production in social media spaces, in which user contributions of content and data are inextricable.

An underlying assumption of the agreements governing data ownership on many social networks is that data is authorless, a collection of technology-generated artifacts rather than an integral part of users’ productive activities. Within terms-of-use policies, user data is posited as a by-product of users’ interactions with technology rather than as products that users have authored within the technological space. A by-product of use, by definition, is a secondary and unexpected result from the making of a principal product. It is a derivative of the intended product but is not the intended outcome of an activity or interaction; in fact, it can even be undesirable in its outcome (think side effects that result from prescription drug interactions). Terms-of-use policies do not grant users a role as authorial agents in the creation of data; rather, the generation of user data is viewed as a technological product. User agreements emphasize that technology companies “collect” and “receive” information when users interact with the site, rather than positing that data is “produced” by the user. For instance, Google states, “When you use our services or view content provided by Google, we may automatically collect and store certain information” (“Privacy Policy”), and Facebook states that “we receive data about you when you interact with Facebook” (“Information We Receive and How It Is Used”). User data appears to be a neutral result of user-technology interaction, naturally occurring
and available to be appropriated and mined by the technology provider. The user’s role a passive one, while the technology itself plays an active, generative role in what appears to be a neutral, even positive process of making good use of an otherwise wasted by-product. Such terms-of-use statements position user data and data records as information without a human author or productive force, without meaning until it is mined and put to use by the site owner.

The understanding of data as objective facts that preexist rather than products of authorial agency precludes an understanding of data as authored texts to be owned in part by users themselves. Krista Kennedy challenges the notion of technologically automated tasks as unaugmented or agent-less in her analysis of Wikipedia bots as authors of encyclopedic entries. Kennedy urges scholars to “recognize more and more that writing happens as an interactive process that involves exchanges between multiple agents, texts, and influences,” and posits that “textual curation” and the products of other mundane tasks performed by automated bots might be considered the intellectual property of authors, or at least “a newly identifiable creature” (308). The nonhuman agents at work on the social and participatory Web compose through solicitation, aggregation, and categorization of data according to technological code. However, this technological agency is intertwined and articulated with human agency at key stages. Technology developers make decisions and implement choices as they design systems for data collection and mining, as well as when they develop interfaces that direct user interaction. Additionally, it is the interaction between humans and technology that results in the productive act of composing data: as users view, contribute, and share with social and participatory Web technologies, with other texts, and with each other, they become collaborative agents in generating data. Despite the fact that data is produced by interactions between humans and technology, the terms-of-use policies noted earlier fail to recognize human agency in the generation of data. The treatment of data as authorless by-product precludes the consideration of data as having creative and intellectual value, as well as the granting of ownership status to users.

The question is not whether social and participatory Web technologies demonstrate some agency at strategic moments in the generation, collection, and use of data, but whether a technological agent is acting alone in the production of user data. According to the terms-of-use statements for social and participatory Web technologies, user data is not a product of a human agent; the only agency demonstrated in its generation is a nonhuman, technological system. If we grant sole agency to technological systems, we should understand clearly the potential implications of excluding human agency from this aspect of digital composition. A consequence of separating an author-user from the productive act of generating user data is that it privileges a structure in which the technology provider automatically assumes ownership and control over user information. If user data is posited as a neutral by-
product of a technological system, it becomes impossible for users to claim ownership or control of it, or even to understand data contributions as a result of acts of authorship. Data becomes, at its inception, free to be appropriated and controlled by those responsible for the technology and not by users. Further, the type of property that is produced—user data—is not understood as holding intellectual and creative value, but only as a commodity to be bought and sold.

IP law is based on the idea of balancing rights, “[t]o promote the Progress of Science and useful Arts” by granting some exclusive rights to authors, but at the same time by placing limitations on those rights to serve the public interest, as do fair use provisions (“Constitutional Provision Respecting Copyright”). The balance between individual ownership rights (which, arguably, provide incentives for creation) and public interests (in accessing, learning from, and building on the works of others) establishes a system of information flow that is characterized by gray areas, fine-grained distinctions, and context-specific rights and responsibilities. In recent years, many have argued that applications of copyright law have been out of balance (for example, Gillespie; Lessig; Logie; Reyman; Vaidhyanathan, Copyrights and Copywrongs), yet control over user data remains an unexamined issue from an IP perspective. Control over user data is largely unregulated from a legal standpoint and greatly imbalanced in its lack of consideration of public interest. Data-mining practices and technology policies unquestioningly place ownership in the hands of technology companies and compel users to surrender control of their own contributions on the social Web. In seeking a more balanced distribution of rights over user data, we might question the assumptions about authorship and agency on which data management is based. One justification for corporate ownership of user data is that technology companies and data brokers invest resources in its collection and use, while users do not, and this lends support to appropriation of user data as the property of corporate author-users. Another justification is that user data is a technological by-product, a result of technological agency rather than the result of an individual user’s productive act. But we should remember that the contributions of users make data possible: user data generation depends on users, on their interactions, participation, and production. It does not exist without them.

The current system for control over user data promotes the social Web as a space for commercial activity, but it does not harness the power of user data for intellectual, creative, or civic purposes. Although technology companies posit that the appropriation of user data improves their services and allows for more integrated, personalized experiences online, these benefits are primarily accomplished through targeted marketing and personalized advertising of goods and services. Instead, users might retain some rights over their data—to control how it is generated, how it can be viewed and by whom, and toward what ends—in ways that they can also apply toward the creative and intellectual acts of exploration, discovery, research,
and learning. The practices of generating, aggregating, and interpreting user data could be understood as collaborative, authorial acts of technological and human agency. With a balancing of users’ and technology companies’ rights over user data, the social and participatory Web could be nourished as a space that provides access to tools for participation and production, and also recognizes the value of human agency required for rich, meaningful social networks.

**Balancing User Data Rights: Conclusions and Implications for the Social Web**

Many user contributions on the social and participatory Web do not easily fit within traditional definitions of authored works. User contributions are complex, situated within a network of users, content, and texts online. They are made even more complex by the Web of interrelated data that is generated behind the scenes. Although the generation of data is inextricable from composing processes within participatory Web spaces, the activity is not typically perceived as producing valuable contributions from human author-users. Rather, data is treated as unclaimed property free for the taking, or as by-products of technology use to be collected and used by technology providers toward largely self-serving ends. Rather than accept the imbalanced distribution of rights over users’ data on the social Web, we should seek more fair and ethical practices that make data collection transparent and that openly recognize the value of users’ data contributions to the cocreation of digital culture.

A new distribution of rights is currently being discussed in legal and public spheres. The issue of data mining has not gone unnoticed by the federal government, which released a White House white paper in March 2012, signed by President Obama, titled “Consumer Data Privacy in a Networked World.” The report contains a code of conduct for technology developers and a call for Congress to pass legislation that applies a privacy “bill of rights” to commercial sectors. One month later the Federal Trade Commission (FTC) released a report on the topic, which encourages technological fixes for the privacy problem, such as the Do Not Track mechanism. Such a mechanism would allow Internet users to participate online without leaving a trail of data behind, though as with the opt-out approaches adopted by some technology services, users must be well informed and tech-savvy enough to employ it. Additionally, using the Do Not Track mechanism may prevent users from being able to use certain social media features and services. This model—in which users must employ technological solutions to address their individual privacy “problems”—fails to offer users an opportunity to shape the development of technologies and policies governing use, and does not allow users to harness the power of collective user data themselves. These ideas for reform offer a strong starting point, but many believe that the proposed solutions will not become law and that voluntary industry compliance is
unlikely. Users themselves need to advocate for a future Internet based on increased user participation in the development of data-mining practices and policies, and for a more balanced ownership structure to manage user data.

The 2012 White House white paper framed the issue of data mining as one about preserving “individual dignity” (53) and the “right to be let alone” (3). For those in English and textual studies, in particular, the issue of controlling user data is about more than “being let alone.” Management of user data is an issue affecting the concept of authorial agency on the participatory Web, which requires maintaining some user control over data, and also openly recognizing the value of data contributions within economies of production online. On a more practical level, management of user data will also determine developing privacy rights and codes of ethics that govern our students’ online academic activities. Yes, users are able to connect with others through social networking services in their personal, professional, and academic activities, and they have the technical ability to participate through free social Web services. But at what cost does this access come? In exchange for technological access, users are giving up their personal information to be used to deliver “social advertising” on Facebook, and to “create one beautifully simple and intuitive experience” as Google claims (Policies & Principles). Although Google tells users that its data use settings will make their lives better by creating a unified online experience, the fact remains that data mining is, at the same time, pushing technology companies’ stock values higher and challenging the boundaries of what users should accept as the cost for participation.

Taken together, the terms-of-use policies governing social and participatory Web services make problematic distinctions between content and data, between agent-directed and agent-less activity, and between authored texts and technology by-products. Photos, postings, and other “member content” are considered to be copyrightable, owned by the user and used for particular purposes by the technology service provider, but data associated with such content and other interactions is not treated as a user contribution. Rather, despite its role as a central good on the social Web, data is presented as technology-generated artifact, or as a neutral by-product of technology usage, to be appropriated and used for a wide variety of purposes by technology companies. Contrarily, user data might be considered authored texts or valuable compositions produced by human agents acting collaboratively with texts and other individuals within a technological environment. Data comes from a productive activity, a result of human interaction within a dynamic technological space. Further, it is inseparable from composing processes of the social Web. Even though the generation of data might be regarded as subsidiary to traditional acts of composing online, the massive amount of information produced by users is central to the value of the participatory Web, enabling collective intelligence, personalized experiences, and responsive technological spaces. Data, then, might be understood
as providing the basis for a sustainable networked, participatory Web.

In this quickly evolving rhetorical environment, a new level of literacy is required for digital composition, teaching, and learning. At the time of this writing, users can employ powerful tools for online composition and communication, but only under terms that are set by outside parties with largely privatized, corporate interests. Universities adopt social Web technologies that aggregate student data, without clearly understanding the workings and potential consequences of these behind-the-scenes practices. After examining the information flow and control systems set up by social and participatory Web platforms in their terms-of-use policies, I see an economic dynamic in which users contribute content and other information of great value to technological systems over which they have very limited control. These platforms, while promising the freedom and ability to participate, through their terms-of-use policies and hidden data-mining functions, disempower users. The policies and terms of use set forth by technology providers often compel writers to surrender control of their own contributions on the social Web.

We should work to educate ourselves, our colleagues, and our students about this unfair exchange system within many social Web environments. Our English classrooms are appropriate sites for discussions about information flow and control and how it relates to literate practices in digital environments. Attention to data-mining practices in our scholarship and teaching raises technology users’ (including our students’, our colleagues’ and our university administrators’) awareness of the types of tacit arrangements on which the social Web is formed. By contributing to our own and our students’ network literacy, we encourage more safe, ethical, and productive participation online. We should advocate for terms-of-use policies—which are couched in both ownership and privacy policies—that limit privatized control of productive resources and that do not compel writers to surrender control of their own contributions.

Participation and composition online often occur on commercial sites that collect and store users’ data: the details of their social networks, their shared conversations, text, pictures, videos, and music. However, the conditions of access to popular platforms for social networking and communication include submission to forms of surveillance, data mining, and target marketing that support a particular view of the Internet: an Internet based on the logic of online commerce, data-driven mass customization, and target marketing. The danger presented is that the contributions by everyday users will potentially be transformed into increasingly exclusive forms of proprietary data, available to the few for use on the many. Instead, if we understand user data as collaboratively authored texts, we can better recognize their worth to the social Web as valuable resources that produce digital culture. We might envision an alternative system that increases transparency in data-mining practices and offers more user-driven environments where users themselves are involved in data
management and the policymaking activities surrounding social technologies. Doing so could contribute to a more fair and ethical balance between incentives to offering tools for participation and protection of the public interest in accessing, using, and creating with them.

Notes

1. Carolyn Miller addresses the question of whether automated computerized systems can demonstrate agency in her discussion of automated tools used for writing assessment in composition programs. She defines agency as “the kinetic energy of rhetorical performance” and argues that “if agency is a potential energy, it will be thought of as a possession or property of an agent […] but if agency is kinetic energy, it must be a property of the rhetorical event or performance itself” (147). Miller views agency as that which requires “attribution,” and recognizes it as a gift of empowerment (153). Marilyn Cooper challenges Miller’s separation of agency from individual human embodiment. Instead of a definition centered on empowerment, she argues, we need an understanding of agency that relies on the concept of responsibility. Cooper recognizes individuals’ conscious and nonconscious rhetorical acts (as, I would add, seem to occur in the production of social data) as acts of agency that have real consequences: “Rhetorical agency is a big responsibility. It means being responsible for oneself, for others, and for the common world we construct together” (444).

2. One exception is law scholar Julie Cohen’s article “Examined Lives,” in which she applies theoretical arguments about IP law to the privacy debate. Ultimately, Cohen argues that the concepts underlying IP law may not serve well the development of fair privacy laws for online media.

Works Cited


