

Owning Your Digital Shadow: Students' Right to Their Own Data

Computers and Writing, George Mason University: May 26, 2018

Mike Edwards

Washington State University

1. WHAT DATA

I'll talk today about how data and metadata are being used to algorithmically disenfranchise groups in multiple ways, including separating people from the products of their labor. However, my presentation proposal did not focus on social media, so I'm not sure how I wound up here: algorithmically selected? I feel an obligation though to suggest applicability, given that in the past few days many of us received a flurry of General Data Protection Regulation enforcement notifications. So I recently asked undergraduates to create a data visualization that embedded themselves in relation to a dataset, and some of them downloaded their JSON-formatted Facebook data and I worked with them to use Python and Tableau to manipulate and visualize those datasets. The arguments I make here can apply to those datasets, to Blackboard datasets if you turn on all the surveillance settings, and to others. The challenge, as End-User License Agreements or EULAs make clear, comes with arguments over who owns that data.

The discourse of intellectual labor ownership rights operates among four frameworks: utility, identity, labor, and civic. The framework of identity is a familiar framework with which to talk about intellectual property and data: in the information economy, we are our data, or at least considerably constituted by the trails of data we leave in our interactions with the world. I'm talking here about two forms of data: first and more familiar is metadata that summarizes data in order to facilitate manipulating and tracking it. Here's a typical view of one form of metadata

called Dublin Core. Metadata can include information about how and why data was created, when and where it was created and by whom, who is entitled to use it and how, and so on. Jeffrey Pomerantz defines data as “the means by which complexity is represented in a simpler form” (in other words, a move fostering efficiency) and argues that today metadata is infrastructural. The other form of metadata I’m interested in here is educational “paradata,” automatically generated data about how students use digital systems in the course of their educations. Paradata is sometimes called “data exhaust” because of the way it can be easily generated as an algorithmic by-product of students’ engagement with digital tools.

2. WHY DATA

In her 4Cs address, Linda Adler Kassner pointed out that such data can be misused, critiquing the impulse of the “education-industrial complex” “to be guided by big data and [predictive] analytics systems” (326). All digital writing scholars do use binary data, and those who focus on its forms understand that studies involving word counts and the quadratic kappa of inter-rater reliability and error rates and other rhetorics of quantification have advanced the field. Metadata can be a form of that quantification and digitization, and while an increasing amount of recent writing scholarship has indicted the rhetorics of big data, I have yet to see any that actually engage significantly large datasets. So my question would be: what if we did? What could we do investigating a large volume of paradata? In order to do so, we would need the use rights for such a corpus of metadata. And the problem with owning metadata is “in the US today. . . information about you is not your property; it’s owned by the collector” (Schneier 195).

Jessica Reyman points out that in contexts like these, user data is understood “as a by-product of technological algorithms . . . 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... the terms-of-use policies governing social and

participatory Web services make problematic distinctions . . . between authored texts and technology by-products... [D]ata is presented as technology-generated artifact, or as a neutral by-product of technology usage. . . 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” (517–529). These are two competing views of data ownership, and the capitalists in this scenario will insist that contract law and the EULA govern the collection of data.

Many paradata-generating LMS and SNS applications rely on EULAs, or do not offer users ways to export their own paradata, and Bruce Schneier cites a Supreme Court decision that “a person has no legitimate expectation of privacy in information he voluntarily turns over to third parties” (68): does the idea of information privacy extend to information ownership? My interest here is economic, and the algorithmic capital of data harvesting makes freshly visible economic inequality: because they, like all technologies, replace labor with capital, which we might therefore see as the economic activity that is at the center of all computers and writing research. We understand, with Adler-Kassner, that “When predictive analytics are done crudely—when the data are bad, when the algorithms are incorrect, or when they fail to take into account consequences—results can be enormously problematic” (327), and this extends to the ways capitalist organizations ignore the consequences of their paradata use and data analytics for the individual student.

3. WHO DATA

Adler-Kassner’s caution about the misuses of data is well-taken, but follows a pattern observed by Victor Villanueva: we “tend to think of ‘economics’ as a numbers game. And we humanities types tend to fear numbers” (58). Those accustomed to market-based understandings

of economics may caution that such quantitative monitoring of composing paradata lends itself to surveillance and capitalist exploitation. So in the case of student composing labor, as Reyman argues, “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. . . Data becomes, at its inception, free to be appropriated and controlled by those responsible for the technology and not by users” (527). Such circumstance would seem to be only slightly different from the form of exploitation critiqued by Marx, but equally pernicious when the value of students’ writerly labor gets appropriated without their consent.

However, in an expanded Marxian view of economy, we don’t need to presume appropriation without consent. First, we should understand that EULAs are contracts, and contracts rely on consent. We can create environments that do not demand students consent to the appropriation of their paradata. Second, we can imagine diverse academic economies wherein laborers appropriate the value of their own labor rather than being exploited by the capitalist, based on different understandings of property rights: in addition to utilitarian and identitarian notions of ownership, we can also consider property rights in the Lockean sense of labor transforming resources into property and the civic rights associated with the Constitution’s copyright provisions for the advancement of knowledge. As Danielle DeVoss and Jim Porter point out, “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” (194), or, in the definition I extend from Marx, the production, distribution, use, and re-production via labor and capital of artifacts, processes, and systems of value. The value of written work can be appropriated by the writer, and the value of learning work by the learner. This appropriation can happen at any stage: at the stage of production, distribution, use, or re-production.

Imagine in such a post-capitalist context what Bill Hart-Davidson observes: one “by-product of the phenomenon of all facets of the writing process—from composing to reading—taking place in digital environments is that writing researchers now have access to very rich, time-indexed sequences of events” (168). Time-use datasets harvested by digital technologies—paradata—allow us to more fully account for the value of writing. In fact, Hart-Davidson continues, “If we consider ‘moments of contact’ with a text to be the subjects of a time-indexed study of the writing process, for example, we could construct fascinating accounts of how [a text] came to be, how it circulated during the drafting, review, and revision stages. . . , and how it will travel to the screens... of others, perhaps becoming part of other texts” (169). Reflection on writing processes promotes knowledge transfer, and we can aid that reflection by incorporating paradata. Lisa Dush has observed the beginnings of a similar tendency for writing on digital networks, which she notes “are too vast, too dispersed, and too diverse to presume to know, especially in advance of a composing task: they favor adaptation over prediction. Content creators iteratively assess audience, using analytics tools” (177), offering the potential for students to become more deeply aware of how their idiosyncratic learning processes operate and evolve. That’s what students’ rights to their own data can do.

4. HOW DATA

This is a call for data sovereignty for students. I would welcome seeing the developers of digital writing tools incorporating the ability for students to easily download, own, and manipulate the user data and analytics generated by their labor. Some of the creators of digital composing applications already provide such paradata: Eli Review offers a robust set of student feedback data downloadable by teachers as comma-separated value .CSV files, which I applaud, and would ask the creators to consider doing that for students as well. If you create it and own it in fixed form, it’s yours and copyrighted, and no one can appropriate it from you. 750words.com

makes granular time-indexed sequences of composing data downloadable as .CSVs as well. Teachers in Blackboard can access an alarmingly deep suite of student engagement surveillance tools, which I would urge to be cracked open for student access. Beyond those forms, though, consider: what if we were to promote academics adding Creative Commons style permissions metadata to our documents? What if we expanded that to Unix-style file permissions metadata stating which parties (users, groups, others, world) could anonymously or identifiably read, write, cite, delete, copy, modify, or duplicate? What if we acknowledged and revived the practice Collin Brooke and Derek Mueller instituted in their version of CCCO in incorporating not only lists of Works Cited but also trackback hooks for lists of Works Citing? What if we incorporated feedback and revision and circulation histories?

Part of the answer there is that soon we're talking about a robust set of paradata descriptors too cumbersome to easily fit into an XML document head. And I'm well aware of the dangers of the algorithm: in our age of automated inequality, algorithmic surveillance, and computational falsification, there's a danger that paradata's "data exhaust" will become obfuscatory corporate data smog — and that's why I want to release that paradata to those who generate it and who should, by virtue of law and human agency, own copyright to it. I'm aware that metadata is not equivalent to activity just as algorithms are not analysis and the map is not the territory. Those considerations should not stop us from thinking about how to develop a metadata schema for circulating academic writing; to ask, what terms would we as a profession want to see added to the Dublin Core? That's I think an achievable goal.

But I'll close with a more ambitiously pie-in-the-sky-and-a-pony-too suggestion. Imagine for composition what Don Tapscott characterized as a continuously growing "vast, global distributed ledger or database running on millions of devices and open to anyone, where not just information but anything. . . can be moved and stored securely and privately," a list of timestamped records of transactions linked and secured using cryptography (Iansiti and

Lakhani), with each block containing a cryptographic hash of the previous block (Narayanan et al.). I'm talking, of course, about blockchain, the technology behind Bitcoin. Blockchain can be used both as ledger and as transport layer for documents that incorporate forms of paradata that students and academics can use to track and improve their writing and help us develop, in the words of Lisa Dush, "the never-done knowledge of how writing develops, within a person or a populace" (355).

References

- Adler-Kassner, Linda. December 2017. "CCCC Chair's Address: Because Writing Is Never Just Writing." *CCC* 69:2. 317–340.
- Bloome, David. 2008. Foreword to *Affirming Students' Right to Their Own Language: Bridging Language Policies and Pedagogical Practices*. Jerrie Cobb, Dolores Y. Straker, and Laurie Katz, eds. Florence, KY: Routledge, Taylor & Francis.
- DeVoss, Daniele Nicole, and James E. Porter. 2006. "Why Napster Matters to Writing: Filesharing as a New Ethic of Digital Delivery." *Computers and Composition* 23: 178–210.
- Dush, Lisa. 2015. "When Writing Becomes Content." *CCC* 67.2: 173–196.
- Hardt-Davidson, William. 2007. "Studying the Mediated Action of Composing with Time-Use Diaries." In *Digital Writing Research: Technologies, Methodologies and Ethical Issues*, edited by Heidi A. McKee and Danielle Nicole DeVoss. Cresskill, NJ: Hampton.
- Iansiti, Marco, and Karim R. Lakhani. January 2017. "The Truth About Blockchain." *Harvard Business Review*. Cambridge, MA: Harvard University. <https://hbr.org/2017/01/the-truth-about-blockchain>
- Marx, Karl. 1976. *Capital: A Critique of Political Economy. Volume 1*. Translated by Ben Fowkes. New York: Vintage.

- Marx, Karl. 1993. *Capital: A Critique of Political Economy. Volume 2*. Translated by David Fernbach. New York: Vintage.
- Marx, Karl. 1993. *Capital: A Critique of Political Economy. Volume 3*. Translated by David Fernbach. New York: Vintage.
- Narayanan, Arvind, Joseph Bonneau, Edward Felten, Andrew Miller, and Steven Goldfeder. 2016. *Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction*. Princeton, NJ: Princeton University Press.
- O’Neil, Cathy. 2016. *Weapons of Math Destruction*. New York: Crown.
- Parks, Stephen. 1999. *Class Politics: The Movement for the Students’ Right to Their Own Language*. Urbana, IL: NCTE.
- Pomerantz, Jeffery. 2015. *Metadata*. Cambridge, MA: MIT Press.
- Reyman, Jessica. 2013. “User Data on the Social Web: Authorship, Agency, and Appropriation.” *CCC* 75.5: 513–533.
- Schneier, Bruce. 2015. *Data and Goliath*. New York: Norton.
- Tapscott, Don. May 2016. “The Impact of the Blockchain Goes Beyond Financial Services.” *Harvard Business Review*. <https://hbr.org/2016/05/the-impact-of-the-blockchain-goes-beyond-financial-services>
- Villanueva, Victor. 2005. “Toward a Political Economy of Rhetoric.” Laura Gray-Rosendale and Steven Rosendale, eds. *Radical Relevance: Toward a Scholarship of the Whole Left*. Albany, NY: State University of New York Press. 57–68.
- Wilkof, Neil. April 2014. “Theories of Intellectual Property: Is It Worth the Effort?” *Journal of Intellectual Property Law and Practice*. Oxford, UK: Oxford University Press. <https://doi.org/10.1093/jiplp/jpu018>