UCLA
Publications

Title

Private Platforms, Metadata, and the Enclosure of Data Access: Urgent Issues for Knowledge Infrastructure Research

Permalink

https://escholarship.org/uc/item/48c7b5p2

Author

Acker, Amelia

Publication Date

2020-02-26

Copyright Information

Private Platforms, Metadata, and the Enclosure of Data Access:
Urgent Issues for Knowledge Infrastructure Research

Amelia Acker, University of Texas at Austin

Two of the most urgent issues for knowledge infrastructure researchers concern the rise of private platforms and their models for data access that rely on profit-driven standards and ontologies. Capitalist in their design, platforms aggregate user data *en masse* so that it can be transformed, bought, sold, and accessed (sometimes for free) as an integral part of a profit model of data expropriation away from users to data brokers (Gehl, 2014). The standards and ontologies used to classify and exchange data in private platforms are different from web standardization efforts and international standards engineering of the early Internet and open web (Russell, 2014). Platforms are frequently described as 'walled gardens' for user experiences. These walls are extended to inhibit and enclose methods for data access as well, because platforms rely upon the 'accumulation by dispossession' of data from individuals (Thatcher, O'Sullivan, & Mahmoudi, 2016).

The generation and reuse of metadata through the asymmetric creation and collection of user data in platforms is now an epistemic mark of our present data culture. Increasingly we see that the semantic relationships between users' data and data derived from their environments (such as work, gym, church, or commute) are classified with corporate taxonomies and then become central to the user experience and product design of personalized platform products from Amazon, Yelp, or YouTube. For example, Spotify recently announced its personas tool, which builds on years of user-centered design techniques to cluster and examine listening behaviors of user groups in the U.S. (Torres de Souza, Hörding, & Karol, 2019). Personas are now part of the platform's internal vocabulary to support the identification of users, categorize their listening habits, and drive algorithmic recommendations. Yet the classified personas label of Spotify users is not available to users who generated this tool. While metadata standards like the personas product taxonomy that describe people and their behaviors have always been essential to social platforms, the intensification and datafication of ICTs we now see proves that some user categories can and have been used in support of profiling, social sorting, and redlining minority and vulnerable populations with platforms (Eubanks, 2018; Noble, 2018).

With profit-driven platforms premised on targeted advertising based on user metadata, the power of legacy standards and the categorization of users has far-reaching consequences when deployed in big data applications, where large-scale data-mining and data analytics are dependent upon initial taxonomies and classification systems to sort users. More and more, we see engagement metadata being gamed for misinformation campaigns, media manipulation and disinformation efforts, even large-scale discrimination tactics across platforms (Acker & Donovan, 2019). For example, ethnic affinity categories are among the options that data brokers and advertisers can use to promote content and direct targeted content to users on platforms (Angwin, Mattu, & Parris Jr., 2016). Platforms typically classify users by a range of affinity categories that may be unknown to users themselves. These metadata structures can be used to control the information that is accessed by some users and not others, as well as categorizing content for specific users based on personalized predictions. These metadata are then used to display new content, push news alerts, and personalize search results to users through affinity categories on platforms such as Google, YouTube, Instagram, even navigation apps.

Grouping users into audiences for targeted advertising is not new. Data markets such as newspaper, television media, food service, and insurance markets existed for decades before mobile networks and social platforms. But with knowledge infrastructures that enable near-constant data creation and collection, platforms can leverage large-scale social networks, environmental sensors, and rapid data-

processing to create new gateways of control and access to collections of data. So, in our data culture of constant creation and collection of data in platforms, metadata standards and ontologies that underwrite these networked infrastructures remain hidden and are often not accessible by most of the creators who produce them or researchers who wish to investigate them (Acker, 2018).

Another characteristic of this moment is that data creators are ceding control and access to their data to profit-driven platforms, whilst platform intermediaries' profit from consuming and providing access to these collections to data brokers who curate them for long-term value. Despite the fact that knowledge infrastructures can now extract data at scale, intermediaries and brokers assert control over collections of user data by enclosing access and inhibiting oversight, criticism, and research (Acker & Kreisberg, 2019; Bruns, 2019). This distance between creators who produce data and collection contexts where data and metadata are accumulated and stored in private platforms is where KI research can have swift and lasting impact. These uncertain archives are made of data and metadata accrued from big data apparatuses such as transportation and mapping apps, social media, mobile devices, learning management systems, and internet infrastructures assembled by data intermediaries to resell and repurpose user data to data brokers. Few of these data archives are accessible to the creators who produced these digital traces and are impacted by these categories the most. Indeed, platform users cannot opt out of affinities, audience profiles, schemas, corporate taxonomies or personas once they have been classified as such, because these metadata standards and ontologies belong to the intermediaries and brokers who control them. While these metadata are not always understood or preserved for the people who created them, they are actively leveraged by platform intermediaries, data brokers, and third-party data consumers to create personalization profiles, predictive analytics, algorithmic recommendations, and user data collections.

All KI reinforce and redistribute authority, power, and control with gateways like metadata standards and ontologies. Today, creating data is a form of belonging, but there are few avenues in place for users to control or access metadata generated about them, or to withdraw from categories once they have been sorted and enrolled into them. If the space that separates users from collections of their data is an epistemic mark of our current moment, then the access and enactment of metadata in private platforms will continue to be how this space grows and more walls are built. It is imperative for KI researchers to confront not just the commodification of data in private platforms, but the impact of enclosure of data access regimes we are witnessing with their rise.

Works cited

Acker, A. (2018). A Death in the Timeline: Memory and Metadata in Social Platforms. *Journal of Critical Library and Information Studies*, *2*(1), 27.

Acker, A., & Donovan, J. (2019). Data craft: A theory/methods package for critical internet studies. *Information, Communication & Society*, *0*(0), 1–20. https://doi.org/10.1080/1369118X.2019.1645194

Acker, A., & Kreisberg, A. (2019). Social media data archives in an API-driven world. *Archival Science*. https://doi.org/10.1007/s10502-019-09325-9

Angwin, J., Mattu, S., & Parris Jr., T. (2016, December 27). Facebook Doesn't Tell Users Everything It Really Knows… [Text/html]. Retrieved August 31, 2018, from ProPublica website: https://www.propublica.org/article/facebook-doesnt-tell-users-everything-it-really-knows-about-them

Bruns, A. (2019). After the 'APIcalypse': Social media platforms and their fight against critical scholarly research. *Information, Communication & Society*, *0*(0), 1–23. https://doi.org/10.1080/1369118X.2019.1637447

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, NY: St. Martin's Press.

Gehl, R. W. (2014). *Reverse Engineering Social Media: Software, Culture, and Political Economy in New Media Capitalism*. Philadelphia, Pennsylvania: Temple University Press.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism* (1 edition). New York: NYU Press.

Pomerantz, J. (2015). *Metadata*. Cambridge, Massachusetts ; London, England: MIT Press.

Russell, A. L. (2014). *Open Standards and the Digital Age: History, Ideology, and Networks*. New York, NY: Cambridge University Press.

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016). Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environment and Planning D: Society and Space*, *34*(6), 990–1006. https://doi.org/10.1177/0263775816633195

Torres de Souza, M., Hörding, O., & Karol, S. (2019, March 26). The Story of Spotify Personas. Retrieved May 2, 2019, from Spotify Design website: https://spotify.design/articles/2019-03-26/the-story-of-spotify-personas/