Datafication Dilemmas: Data Governance in the Public Interest

Linda Huber
ludens@umich.edu
University of Michigan
School of Information
Ann Arbor, Michigan, USA

Anubha Singh
anubha@umich.edu
University of Michigan
School of Information
Ann Arbor, Michigan, USA

Lynn Dombrowski
lynndombrowski@gmail.com
Georgia Institute of Technology
School of Interactive Computing
Atlanta, Georgia, USA

Shion Guha
shion.guha@utoronto.ca
University of Toronto
Faculty of Information
Department of Computer Science
Toronto, Ontario, Canada

Jean Hardy
jhardy@msu.edu
Michigan State University
Department of Media &
Communication
East Lansing, Michigan, USA

Naja Holten Møller
naja@di.ku.dk
University of Copenhagen
Department of Computer Science
Copenhagen, Denmark

Abstract

Within CSCW, there is a robust body of literature examining the infrastructures, institutions, and labor processes enabling pervasive processes of datafication. Key public sectors and industries are being increasingly datafied, often in the service of implementing AI tools to improve the efficiency, efficacy, and equity of public goods and services. This panel specifically examines the role of CSCW scholars in advancing public-interest technologies by engaging with datafication through the register of *data governance.* Each panelist brings perspectives from diverse domains of datafication - including healthcare, education, and agriculture - and state, corporate, and community forms of data governance. We invite panelists and attendees to explore a series of provocations focused on reconceptualizing our role and responsibility as researchers in intervening in processes of datafication and data governance.

CCS Concepts

• **Human-centered computing**; • **Social and professional topics** → **Computing / technology policy**;

Keywords

datafication, data governance, public interest technology

1 Motivation and Background

We are in a period of ever-intensified datafication: new domains of life are being digitized, made into fodder for new AI and data-driven products, services, and tools of algorithmic control. Business leaders and policy makers position datafication as key to creating more efficient, modern modes of governance and value-production. To feed the "data imperative" [3], work across all roles and sectors is transformed into data work [2, 5, 9].

As a field, CSCW has attended closely to the politics and practices of datafication [1, 6, 11]. Yet, this engagement has primarily focused on datafication through the lenses of work practices or design, and relatively little on topics of data governance. As Jackson, Gillespie and Payette [4] have argued, governance and policy are not simply arcane legal constructs applied after the fact to already-stabilized technological artifacts and practices: formal and informal policies help to create the conditions or terrain upon which certain technical designs and practices become possible.

This panel specifically seeks to engage with data governance both as a site of study and a domain of potential intervention. We use the term "data governance" quite broadly, referring to policies, procedures and protocols developed by corporate platforms, state regulatory bodies, and in communities. We also consider more diffuse modes and practices of corporate "governmentality" [10] - i.e., how corporations enact control and power through data.

As datafication extends beyond the realm of social media or gig work platforms into key public sectors such as health, education, housing and agriculture, critical examination of the potential harms and benefits of emerging infrastructures of datafication are increasingly crucial. Data governance provides a possible inroad for critically examining both goals and values driving datafication, as well as for possible avenues for intervention at the design/policy/practice nexus. This panel opens up a conversation about the role and responsibilities of CSCW researchers in relation to data governance, and seeks to expand the "mosaic" [14] of potential contributions that HCI scholars can make with regards to datafication specifically.

This panel focuses on datafication processes in key public domains such as healthcare and social work, urban infrastructures, agriculture, and education. As we study those who are crafting technical, policy, and social infrastructures of datafication in the public interest - what role do we, as CSCW researchers, have in

moving beyond knowledge production or design recommendations towards directly intervening in processes of datafication and data governance? Where do our responsibilities and loyalties lie? What kinds of unexpected collaborations, interventions, or techniques can we bring to the table?

Through this conversation, we specifically explore research methods and approaches to examining datafication and data governance that move beyond abstract critique, and which strengthen our capacity not only to mitigate the harms of extractive systems, but also to propose and advocate for systems, policies, and data practices which are foundationally different than those offered by corporate actors [12]. We explore approaches from our own work, including critical engagement with the problematizations and which motivate state-led processes of datafication, or historicizing processes of datafication or situating them within larger projects of economization [7]. We also explore how our own positionalities and loyalties as researchers shape the kinds of questions we can ask and the kinds of stories we can tell [8]. For instance, how do we navigate between critique of nascent data infrastructures, while maintaining our relationship with well-intentioned interlocutors (e.g.,policy makers, data workers, and public servants)? We present some strategies available beyond research or design recommendations, including curriculum design, counter-narration, and scholar-activist approaches , to intervene in processes of datafication and data governance. We draw inspiration from calls to scholar activism, to unexpected collaborations and contributions between HCI researchers and the "publics" that are impacted [13].

Within the current financialized hype cycle of the "AI summer," this is a crucial moment for opening up broader questions about exactly how public interest is being construed and enacted within datafied infrastructures, and for developing a multiplicity of tactics for intervening within those systems and processes. This includes reclaiming data governance as a kind of everyday practice, rather than relegating it to the work of legal experts and corporations.

2 Attendees

This panel is intended for CSCW researchers examining processes of datafication (e.g., data collection, analysis), particularly within public sectors or for the development of public-interest technologies. Through this open dialogue we hope to provide fodder for attendees to think differently about the scope of possible intervention with regards to datafication and data governance.

3 Panel Structure and Format

Each panelist will briefly introduce themselves and speak for 3-4 minutes about the domain(s) in which they are studying datafication and governance. We will then move into round-robin discussion of each of the three provocations below. After each panelist responds, we will move to open-forum discussion, inviting audience members to also answer the provocation – or to build upon or extend the panelists' responses. For each question, we will allocate a maximum of 20 minutes of discussion across panelist responses and attendee engagement.

The draft provocations are as follows: 1. What techniques or strategies are available to you as a researcher to intervene in regimes of datafication, or to advance justice and the "public interest" through

datafication? 2. Drawing from your experiences examining datafication in public sectors, where would you situate yourself in between "AI abolitionist", "critical pragmatist", or "reluctant AI booster" - and why? 3. Within the datafication domain(s) that you study, have there been specific moments or practices or dreams articulated by your interlocutors that help to point towards more liberatory or just data regimes? What goals or visions do you share for the datafied future?

To help facilitate this open conversation, we will have both a discussion facilitator and a moderator. The moderator will keep time for each set of provocations, ensure that the conversation keeps moving and that participants are "sharing the air." The moderator will also help to ensure accessibility by repeating questions into the mic as necessary, etc. The discussion facilitator will be responsible for helping to develop a coherent, constructive conversation, identifying shared themes, areas of omission or which warrant further discussion, and synthesizing the take-aways from the conversation at the end of the panel.

4 Panelists

Each panelist engages in datafication and data governance from different domains/perspectives. This includes different forms of governance ranging from formal lawmaking, corporate governance and policy-making, and community governance. The panelists also draw from several different diverse domains, including healthcare, education, agriculture. Additionally, although unified by an interest in advancing justice and equity in 'public interest technology', the panelists also situate themselves differently on the "fix it to burn it down" spectrum [11], and have diverse tactics for intervening and engaging with their fields.

Lynn Dombrowski is an Associate Professor at the School of Interactive Computing at the Georgia Institute of Technology. She uses empiric and design methods to investigate the possibilities and limits for social change/justice using sociotechnical, participatory, and community-based approaches. Her study topics have ranged from various types of work/workplaces, policing, addiction, food insecurity, domestic violence, and so on.

Shion Guha is an Assistant Professor at the Faculty of Information and Department of Computer Science at the University of Toronto. He is the director of the Human-Centered Data Science Lab and the coordinator of the Human-Centered Data Science graduate degree program. His research focuses on integrating technical, computational methods with critical, interpretive inquiry to study issues of data and algorithmic biases in public sector agencies such as child welfare, public health, homelessness and higher education.

Jean Hardy is an Assistant Professor of Media & Information at Michigan State University where he directs the Rural Computing Research Consortium. At the intersection of CSCW and rural sociology, his research uses ethnographic and design methods to understand the growing role of digital technology in rural economic development.

Linda Huber is a PhD candidate at the University of Michigan School of Information. Her research examines how business models and organizational practices are shifting in relationship to emerging digital technologies. Specifically, she uses ethnographic methods to examine processes of datafication and platformization within

the healthcare industry. This research contributes to the domains of feminist science and technology studies (STS), organization studies, and critical data studies.

Naja Holten Møller is an Associate Professor in the Department of Computer Science, University of Copenhagen, Denmark. She is also a member of ACM's Future of Computing Academy. Her research unfolds through a deep engagement with issues of responsibility, and the enactment of ethics through participation of stakeholders in technology use- and design. In particular, Naja Holten Møller is noted for her research on "data work" and the challenges and opportunities opened up by the use of large datasets and algorithms to optimize work in the future workplace.

Anubha Singh is a PhD candidate at the University of Michigan School of Information. Through long term ethnography of the onion supply chain in Western Maharashtra, she studies how data-driven technologies are restructuring farming and redefining the future of agriculture in India. Her work is informed by and contributes to the fields of Postcolonial and Feminist STS, Ethnography of Computing and Agriculture, and HCI.

References

[1] Stevie Chancellor, Shion Guha, Jofish Kaye, Jen King, Niloufar Salehi, Sarita Schoenebeck, and Elizabeth Stowell. 2019. The Relationships between Data, Power, and Justice in CSCW Research. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing* (Austin, TX, USA) *(CSCW '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 102–105. https://doi.org/10.1145/3311957.3358609

[2] Joel E Fischer, Andy Crabtree, James A Colley, Tom Rodden, and Enrico Costanza. 2017. Data Work: How Energy Advisors and Clients Make IoT Data Accountable. *Comput. Support. Coop. Work* 26, 4 (June 2017), 597–626. http://dx.doi.org/10.1007/s10606-017-9293-x

[3] Marion Fourcade and Kieran Healy. 2016. Seeing like a market. *Socioecon Rev* 15, 1 (Dec. 2016), 9–29. https://academic.oup.com/ser/article/15/1/9/2656165

[4] Steven J Jackson, Tarleton Gillespie, and Sandy Payette. 2014. The policy knot: re-integrating policy, practice and design in cscw studies of social computing. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 588–602. https://doi.org/10.1145/2531602.2531674

[5] Hanlin Li, Nicholas Vincent, Stevie Chancellor, and Brent Hecht. 2023. The Dimensions of Data Labor: A Road Map for Researchers, Activists, and Policymakers to Empower Data Producers. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency* (Chicago, IL, USA) *(FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1151–1161. https://doi.org/10.1145/3593013.3594070

[6] Milagros Miceli, Martin Schuessler, and Tianling Yang. 2020. Between Subjectivity and Imposition: Power Dynamics in Data Annotation for Computer Vision. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2 (Oct. 2020), 1–25. https://doi.org/10.1145/3415186

[7] Michelle Murphy. 2017. *The Economization of Life.* Duke University Press. https://www.degruyter.com/document/doi/10.1515/9780822373216/html

[8] Ihudiya Finda Ogbonnaya-Ogburu, Angela D R Smith, Alexandra To, and Kentaro Toyama. 2020. Critical Race Theory for HCI. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3313831.3376392

[9] Kathleen Pine, Claus Bossen, Naja Holten Møller, Milagros Miceli, Alex Jiahong Lu, Yunan Chen, Leah Horgan, Zhaoyuan Su, Gina Neff, and Melissa Mazmanian. 2022. Investigating Data Work Across Domains: New Perspectives on the Work of Creating Data. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22, Article 87)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3491101.3503724

[10] Nikolas Rose, Pat O'Malley, and Mariana Valverde. 2006. Governmentality. *Annual Review of Law and Social Science* 2, Volume 2, 2006 (Dec. 2006), 83–104. https://doi.org/10.1146/annurev.lawsocsci.2.081805.105900

[11] Divy Thakkar, Azra Ismail, Pratyush Kumar, Alex Hanna, Nithya Sambasivan, and Neha Kumar. 2022. When is Machine Learning Data Good?: Valuing in Public Health Datafication. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22, Article 322)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3491102.3501868

[12] Jacob Thebault-Spieker, Stevie Chancellor, Michael Ann DeVito, Niloufar Salehi, Alex Leavitt, David Karger, and Katta Spiel. 2021. Do We Fix it or Burn it Down? Towards Practicable Critique at CSCW. In *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, USA) *(CSCW '21 Companion)*. Association for Computing Machinery, New York, NY, USA, 234–237. https://doi.org/10.1145/3462204.3483281

[13] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R Rick, and Lilly Irani. 2021. HCI Tactics for Politics from Below: Meeting the Challenges of Smart Cities. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21, Article 297)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3411764.3445314

[14] Qian Yang, Richmond Y Wong, Steven Jackson, Sabine Junginger, Margaret D Hagan, Thomas Gilbert, and John Zimmerman. 2024. The Future of HCI-Policy Collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24, Article 820)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3613904.3642771