Data feminism in the wild: case studies of indexing projects and their futurities

Rachel Pierce (University of Gothenborg), Hana Marčetić (University of Borås, Sweden)

Introduction

The past decade has seen increasing calls for more feminist data studies and doing data feminism, culminating in the book *Data Feminism* by Catherine D'Ignazio and Lauren F. Klein (2020). The compilation of bibliographic information and work with indexing and cataloging is central to library work but understudied as a practice, though data feminists have highlighted how central subject description is for finding information in complex physical and digital environments (Adler 2017), and queer library and information science scholars have urged practitioners to "queer metadata" (Ruberg, Boyd & Howe 2018). This emphasis on description and networking of individuals and materials has historical, analog roots that scholars have uncovered (McKinney 2020; Eichhorn 2013). These roots affect the development of current indexing and digital tools, as well as how feminist indexers think about future opportunities.

Previous work has attempted to identify data feminism in contemporary scholarship (Marcetic and Nolin 2022) and the futurities implicit in these studies. This paper will continue that work while serving as a starting point for investigating whether and how data feminist principles are implemented in cataloging projects that are constrained by funding, scope, and other resources, as well as by histories of subject description. The paper will examine the Homosaurus LGBTQ+ thesaurus, the Swedish women's history-oriented KVINNSAM database, and the QueerLit project, which is working to develop a Swedish-language database that bridges the gap between Homosaurus and KVINNSAM to make prose, poetry, and drama texts by or about LGBTQ+ individuals and themes more visible and findable. Using these examples, this paper expands the scope of analysis beyond the individual project, providing a multinational scope that places American feminist work and English-language indexing at the periphery. As intersectional data feminists have argued, the implementation of feminist work demands a focus on the local (Risam and Josephs 2021), and these projects will provide an opportunity for examining how local identities provide opportunities, constraints, and new questions for data feminism.

Theoretical Framework

Currently, the Science and Technology Studies (STS) field focuses on networks and materiality, largely overlooking questions of power and normativity. But datafication or translation of social action into data is a significant driver of contemporary surveillance (Zuboff 2019) and social regulation (Eubanks 2018, O'Neil 2016), and deepening inequalities (Noble 2018). The theoretical concept of sociotechnical imaginaries centers the role of science and technology in investigations of how contemporary society understands its various potential futures. Scholars employing this concept are concerned with "collectively held and performed visions of desirable futures" and normative beliefs about the development and future effects of technology (Jasanoff & Kim 2015, p. 11).

Feminist, LGBTQ+ and critical race studies and projects explicitly position themselves as alternatives to dominant discourses that privilege capitalistic, market-driven, and top-down approaches to technology and technological change. Often positioned as activism, feminist interventions in information and data creation and dissemination are framed as simultaneously utopian and obtainable (Mandell 2019). Less well-understood is what data feminist scholars and cultural heritage practitioners take for granted and what they see as malleable or changeable. This paper will use the theoretical concept of sociotechnical imaginaries to investigate how the contours of the possible are constructed within practical, on-the-ground data feminism within contemporary library indexing and thesaurus work.

Research Questions

What are the overlaps and gaps in focus between data feminist scholarship and current feminist indexing and thesaurus work?

How do current indexing and thesaurus projects distinguish between work to be done now and work that must wait for a more equitable future?

What is currently prioritized? For what future technological worlds have indexers planned?

Methods and Data Sources

This study focuses on feminist indexing projects that span the analogue and the digital, as well as having international and national scopes. Literature on data feminism has tended to focus on North American empirical material, most often in English. The projects we have chosen exist across these divides. We will examine project funding texts, the structure of various indexes and thesauri, library and researcher team goals, and information on future planning. This material array will allow for comparisons of intended scope, audience, and ideological principles while also providing a sense of how interconnected (or not) these projects/infrastructures are.

Homosaurus[1] is a controlled vocabulary project that is the cooperative effort of American and Dutch librarians. This project is currently working on expanding its subject list while translating information into linked data to make searching across institutional and linguistic boundaries easier. The project has generated some attention from scholars but has not been placed within a broader context of digital controlled vocabulary and linked data projects beyond the United States (Cifor and Rawson 2022).

Meanwhile, KVINNSAM[2] is a subject indexing project that began in the late 1950s, when three women's librarians and researchers realized that it was difficult to find material on women and women's history using the Swedish subject list (SAO). They built a catalog and indexing list that KvinnSam, the National Resource Library for Gender Studies, has updated and translated into digital form, so that it can be integrated with Libris, the national library catalog for Sweden. Existing analyses are almost entirely devoted to whether the KVINNSAM database is representative of and provides an adequate entryway into contemporary feminist scholarship.

Lastly, QueerLit[3] is a current project funded by Riksbankens jubileumsfond, a national Swedish research funding body. Researchers at the Centre for Digital Humanities at the University of Gothenburg are working with librarians at KvinnSam, as well as the National Library and the Archives and Library of the Queer Movement in Gothenburg. The end product will be an indexing of literature (novels, poetry, and drama) by and about LGBTQ+ individuals and themes, with a subject list in Swedish that is compatible with Homosaurus and integrated into Libris. A stand-alone database will also eventually be available.

Preliminary Results

An initial survey of these three projects demonstrates both overlap and distinctions between scholarly discourses and the "in the wild" reality of implementing data feminist principles. Two trends are immediately visible and worth more concentrated work: (1) the way in which the pre-existing array of technologies, resources, and systems constrains or alters the prioritization and implementation of "utopian" suggestions prominent in scholarly literature, and (2) how "archivism" or get-it-done approaches that prioritize visibility mean that some of these more "utopian" goals run the risk of repeatedly being set aside in practice. In other words, looking at the concrete projects of data feminism can expose potential for conflicts between the goals articulated by this area of scholarship.

In further analysis, we will pay special attention to the underlying notions of desirable and undesirable data practices as well as where these projects draw the line between what is currently

---

[1] https://homosaurus.org/ (accessed 29 Oct 2022).

[2] http://libris.kb.se/form_extended.jsp?f=kvin (accessed 29 Oct 2022).

[3] Information on the project can be found here: https://www.gu.se/en/research/queerlit-data-base (accessed 29 Oct 2022).

implementable and what can only be implemented in a more equitable future. KVINNSAM's status as the most local of the projects is interesting, considering that this database project is the longest-lived and, because of this history and pre-existing infrastructure, the most politically traditional. Understanding how locality and longer-term institutionalization affects feminist indexing practices and understandings of the future is central. The positioning of actors and digital initiatives in relation to their target audiences/communities is one of many structures shaping the way in which these communities define themselves and parse current from future goals.

Discussion

Via this analysis, we will be able to identify what is in the process of becoming normative within the increasingly structured world of data feminist scholarship and practice. The sociotechnical imaginaries framework will allow us to reflect on the nature and prioritizations of feminist data praxis. This praxis will structure the future of feminist resistance to seemingly monolithic marketized discourses about technological "advance" as neutral, democratic, and inevitable. Importantly, this discussion will center the role of libraries and librarians in structuring and advancing data feminist goals through indexing and cataloguing work while decentering the United States. The examples presented also provide more information about the role of locality, which is central in both data feminist scholarship and concrete proposals such as the CARE principles for linked (open) data (Carroll et al. 2020). Localities in these projects operate on multiple axes that include linguistic (Swedish; multilinguistic), institutional (library, research project) and critical perspectives (gender and queer), which interact with geographical boundaries. These localities affect how data feminists structure and imagine the future (impact) of their projects.

Conclusion

Data feminist scholarship has always been focused on applicability – its theoretical underpinnings are situated in relation to and are supposed to be attuned to real-world use (D'Ignazio & Klein 2020). Metadata construction and subject classification have been central to data feminist discussions of openness, democratic access, and representativeness within digital milieus. Concrete proposals for greater inclusivity are abundant, but there are still relatively few studies of project work and implementation that span several projects and look at project work as a human, material set of communities and processes. This study is a start.

REFERENCES

Adler, Melissa. (2017). Cruising the Library: Perversities in the Organization of Knowledge. New York: Fordham University Press.

Carroll, Stephanie Russon, Garba, Ibrahim, Figueroa-Rodríguez, Oscar L., Holbrook, Jarita, Lovett, Raymond, Materechera, Simeon, Parsons, Mark, Raseroka, Kay, Rodriguez-Lonebear, Desi, Rowe, Robyn, Sara, Rodrigo, Walker, Jennifer D., Anderson, Jane, and Hudson, Maui.

(2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19(1), 43. DOI: http://doi.org/10.5334/dsj-2020-043

Caswell, Michelle. (2021). Dusting for Fingerprints: Introducing Feminist Standpoint Appraisal. In: Radical Empathy in Archival Practice, eds. Elvira Arroyo-Ramirez, Jasmine Jones, Shannon O'Neill, and Holly Smith. Special issue, Journal of Critical Library and Information Studies, 3(2). DOI: 10.24242/jclis.v3i2.113.

Cifor, Marika and Rawson, K.J. (2022). Mediating Queer and Trans Pasts: The Homosaurus as Queer Information Activism. Information, Communication & Society. DOI: 10.1080/1369118X.2022.2072753.

D'Ignazio, Catherine and Klein, Lauren F. (2020). Data Feminism. Cambridge, MA: The MIT Press.

Eichhorn, Kate. (2013). The Archival Turn in Feminism: Outrage in Order. Philadelphia, PA: Temple University Press.

Eubanks, Virginia. (2019). Automating Inequality: How High-tech Tools Profile, Police, and Punish the Poor. New York, NY: Picador.

Jasanoff, Sheila and Kim, Sang-Hyun. (2015). Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power. Chicago: London: The University of Chicago Press.

Mandell, Laura. (2019). Gender and Cultural Analytics: Finding or Making Stereotypes? In: Debates in the Digital Humanities 2019, eds. Matthew K. Gold and Lauren F. Klein. Minneapolis, MN: University of Minnesota Press, p. 3-26.

Marcetic, Hana and Nolin, Jan. (2022). Feminist Data Studies and the emergence of a new Data Feminist knowledge domain. First Monday, 27(7). DOI: 10.5210/fm.v27i7.12295

McKinney, Cathy. (2020). Information Activism: A Queer History of Lesbian Media Technologies. Durham, NC: Duke University Press.

Noble, Safiya Umoya. (2018). Algorithms of oppression: How search engines reinforce racism. New York: New York University Press.

O'Neil, Catherine. (2017). Weapons of math destruction: How Big Data increases inequality and threatens democracy. London: Penguin Books.

Risam, Roopika and Josephs, Kelly Baker, eds. (2021). The Digital Black Atlantic. Minneapolis, MN: University of Minnesota Press.

Ruberg, Bonnie, Boyd, Jason, and Howe, James. (2018). Toward a Queer Digital Humanities, Bodies of Information: Intersectional Feminism and Digital Humanities, ed. Elizabeth Losh and Jacqueline Wernimont, Minneapolis, MN: University of Minnesota Press, pp. 108-127.

Zuboff, Shoshana. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. New York: Profile Books.