Asking Caro to address those questions is asking him to write the book archivists would like and not necessarily the book he has in mind. He is not writing for archivists per se. But the current book might teach archivists a little—and maybe more than a little—about how seasoned researchers experience what archivists have to offer. And the writing is so enjoyable that it is worth a look.

© Dan Harper
University of Illinois at Chicago

Responsible Operations: Data Science, Machine Learning, and AI in Libraries

By Thomas Padilla. Dublin, OH: OCLC Research, 2019. 35 pp. Open Access PDF.
ISBN 978-1-55653-151-4. DOI: https://doi.org/10.25333/xk7z-9g97.

Published in December 2019, *Responsible Operations: Data Science, Machine Learning, and AI in Libraries* is a new Research Position Paper from OCLC that is particularly prescient in the light of recent prominence in the mainstream media of the activism by the Black Lives Matter movement to end the kind of structural inequalities perpetuated by data-driven policies. Indeed, decisions based on flawed data science further uphold structural inequalities. This paper addresses just one area of library and archival practice where the need for change to tackle structural inequalities is urgent.

We live in a data-driven world, and, whether we like it or not, we are affected by hidden algorithms that drive our internet search engines and social media, as well as feed the decision-making processes of policy makers, insurance providers, law enforcement agencies, and so on. In her 2016 book, *Weapons of Math Destruction*,[1] Cathy O'Neil discussed a wide variety of misuses and abuses of data to drive decision-making and policy that affect the lives of millions of people across the world. She discussed the ways insurance companies, probation services, and police authorities use algorithms to make decisions about people and communities that reinforce and uphold structural inequalities linked to race, gender, and residential location, among other factors. The algorithms that drive data analysis are opaque and closed to scrutiny. No opportunities exist to provide feedback on results to create a force for good rather than a machine that reinforces inequalities and the status quo. Safiya Umoja Noble's 2018 *Algorithms of Oppression*[2] took this further to highlight the damaging effects of the racist and sexist substructures of search engines (and their results). She further drew a direct link between the historical roots of library cataloging practice and the development of data science and search engines: "Information organization is a

matter of sociopolitical and historical processes that serve particular interests."[3] We cannot live in a time like this and ignore the consequences of both our actions and our inactions as library and information professionals, which have a direct bearing on the lives of others.

This research paper boldly sets an agenda for tackling the positive and negative impacts of data science, machine learning, and artificial intelligence (AI) in libraries. OCLC has long been an excellent source of publications that distill practical guidance from what can be an intimidating landscape of theory and complex research questions, and this paper is no exception. The author, Thomas Padilla, has a strong background in digital scholarship, digital literacy, inclusion, and data management and is exceptionally well placed to comment on the intersection between data science and libraries. In this paper, he brings erudition, wisdom, and a political dimension to a complex area, which is, as he points out, often siloed to particular sections of an organization.

The paper is a self-described call-to-action, which is welcome at a time of heightened awareness that "something needs to be done." If you find yourself asking the question, "but what can we do to be an ally?," then addressing the recommendations in this paper might be one step along this road. The tone is polemical as befits a profoundly political subject and requires a systemic shift in thinking and resource allocation to center activities that are not yet a holistic part of good practice. As Padilla writes, "Diversity is not an option—it is an imperative" (p. 9).

Responsible Operations is a product of interviews and meetings with experts in the fields of libraries, archives, and digital scholarship coming together to discuss the impacts that data science, machine learning, and AI have on libraries. Padilla provides the caveat that, as a result of time and resource constraints, the group of experts is drawn largely from the United States and the English-speaking world. He acknowledges that the debate needs to be widened. The ethics of machine learning and AI in libraries and information has global reach and impact in a digital world, where all libraries and data sets can potentially be connected, and all collections harvested and analyzed. This world offers huge opportunities but holds equally huge responsibilities.

The paper was commissioned to chart the library community's engagement with data science, machine learning, and AI. As the author makes explicit, this is an extremely complicated, multifaceted agenda. To the credit of its author and contributors, this paper brings together a cogent set of areas of investigation which, although separate, are interdependent; no single part of the agenda can be considered or tackled independently. These elements are successfully synthesized to describe a holistic agenda with which the library community can begin to take the steps necessary to collect, manage, and make accessible their collections in a responsible fashion.

The audience for the paper is explicitly "Library administrators, faculty, and staff, University administrators and disciplinary faculty" (p. 7), then the rather more loosely defined professionals operating in commercial and nonprofit contexts interested in collaborating with the aforementioned. The reviewer takes this to mean those professionals working in operations that create, use, or otherwise manage data and algorithms. This is interesting but not further explored in the paper. Finally, Padilla addresses funders as part of his key audience since the recommendations in one way or another all necessitate the injection of resources. However, the scope of the recommendations is thrown wider and the onus put on each and every one of us engaged in library and information work: "No single country, association, or organization can meet the challenges that lie ahead. Progress will benefit from diverse collaborations forged among librarians, archivists, museum professionals, computer scientists, data scientists, sociologists, historians, human computer experts and more. All have a role to play" (pp. 6–7). Although focused on the United States, the paper offers much of value and use to an international audience. The agenda is explicitly multifaceted and diverse—therefore, it needs to be international and multilingual in its operation.

The paper attempts to bring together key areas of investigation (p. 9), the first of which is an overall commitment to the concept of the ethical application of machine learning and AI to collections care and the centering of data science in library and information practice. In the "responsible operations" of the title, Padilla refers to Rumman Chowdhury's concept of *responsible operations*, which brings ethical considerations to the use of AI. Under the headings of "Description and Discovery," "Shared Methods and Data," and "Machine-Actionable Collection," Padilla explicitly discusses how this practice can directly apply to collections care. The final three areas of "Workforce Development," "Data Science Services," and "Sustaining Interprofessional and Interdisciplinary Collaboration" comprise the framework for making all of this operational with a focus on development, training, and collaborative working.

While some of the issues are relatively well understood, bridging the gap between theory and practical application has thus far proven difficult. Padilla's paper does highlight areas of good practice and case studies worthy of attention, but he identifies these as "archipelagic"—disparate and disjointed—when we should instead aim for an interconnected community of practice. Good practice is out there but not widespread, and Padilla calls for the creation of venues to facilitate joint work, publications outlets, and platforms for exploring methods of practice and funding sources to enable and encourage its wider adoption.

Padilla calls for the creation of more machine-actionable collections with a broad audience in mind. He argues that too often the intended audience for a collection is conceived of in very narrow terms, which leads to the work as

being either lauded as cutting-edge or dismissed as "boutique." In neither case does this help center and operationalize the kind of work called for here.

This consideration of audiences should be central to any library or archival service delivery—we collect and maintain collections to make them available and accessible to our users, both current and future. What we collect and how we make the materials available are political acts. This ties into the key concern of bias within collections that further drives structural inequalities in machine learning: "Historic and contemporary biases in collection development activity manifest as corpora that overrepresent dominant communities and underrepresent marginalized communities" (p. 15). Here, the recommended actions are rightly to prioritize the creation of machine-actionable collections that speak to underrepresented communities. Underneath this lies a whole raft of essential work around building relationships that require time, effort, and resources. Unquestionably, these tasks are necessary but should not be underestimated.

It can seem a bit overwhelming. Every recommendation includes forming a working group, and this is off-putting for the individual practitioner. As a framework document, however, this paper is comprehensive, and it really remains for institutions to take up the challenge going forward of how to implement these actions and see real change that will benefit not just library and information professionals, but all users and stakeholders.

To center and diversify the work required to move forward on these recommendations, advocacy strategies are clearly needed, and something about how these could be developed would have been welcome. How does the data-science-engaged librarian or archivist speak to the senior manager juggling a thousand other competing tasks to prioritize the imperative to act in this area? Although both the language and discourse of this paper are aimed very squarely at academic libraries, those working in other institutions may wonder where they fit in this. The target audience is likely managers and policy makers, and the wider the recommendations can be spread the better. Everyone can listen to and participate in the call-to-action, but it will be most effective if those in a position to allocate resources to the kind of research required pay the most attention.

Overall, *Responsible Operations* is a superb synthesis of the issues and obligations that fall on all of us who manage digital collections and data. The recommendations come as a call for all of us to enact responsible operations and start to implement the kind of changes we want to see in the world.

© Rachel MacGregor
University of Warwick

Notes

[1] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (New York: Crown, 2016).

[2] Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: New York University Press, 2018).

[3] Noble, *Algorithms of Oppression*, 138.

Partners for Preservation: Advancing Digital Preservation through Cross-Community Collaboration

Edited by Jeanne Kramer-Smyth. London: Facet Publishing, 2019. 240 pp. Softcover and EPUB. $93.99US, £69.95UK. Softcover ISBN 978-1-78330-347-2; EPUB ISBN 978-1-78330-349-6.

While cultural heritage and memory institutions traditionally responsible for saving materials are leading digital preservation efforts, the challenges of digital preservation require the involvement of new stakeholders. These new stakeholders include many organizations and individuals with diverse needs and priorities, including rights holders, lawyers and lawmakers, data scientists, architects, and hardware and software developers. Collaboration with these stakeholders is required to make digital preservation activities effective and to allow programs to mature. The complexities of digital preservation activities and the higher stakes in digital preservation are creating a stimulus for an interdisciplinary approach to digital preservation. *Partners for Preservation: Advancing Digital Preservation through Cross-Community Collaboration* seeks to build bridges between archivists and stakeholders in other professions outside the GLAM (galleries, libraries, archives, and museums) fields to address common issues and struggles inherent in digital preservation. From digital inheritance rights to data visualization, this book builds a compelling case that "Archivists cannot navigate the flood of technology and changes alone" (p. xxii) and should seek mutually beneficial partnerships with other professions to traverse the digital landscape. Archivists must look beyond their profession and identify those in other domains who may also have a stake in building mature and sustainable digital preservation practices and programs.

Partners for Preservation is edited by Jeanne Kramer-Smyth, who, after a twenty-year career as a software developer, graduated with an MLS from the University of Maryland College of Information Studies and is currently an electronic records archivist with the World Bank Group Archives. Kramer-Smyth expertly weaves together a multifaceted discussion of digital preservation challenges through the lens of ten subject matter experts whose backgrounds span legal studies, journalism, architecture and design, information security, statistics, and data visualization.