The Politics of Mass Digitization. Nanna Bonde Thylstrup. Cambridge, MA: The MIT Press, 2018. 216 pp. $35.00 (hardcover). (ISBN 9780262036870)

Paul Conway
Associate Professor
University of Michigan School of Information
E-mail: pconway@umich.edu

The Politics of Mass Digitization is a watershed work of scholarship that establishes a new theoretical discourse on the wholesale digital transformation of cultural heritage resources. The particular focus of this book, worked out through three case studies and framed by innovative theoretical synthesis and insight, is on the socio-political complexities of digitized content aggregated and networked at scale. The analysis functions simultaneously under the surface of technical processes, at the boundaries of state-controlled intellectual property regimes, and at the borderless domain of global information infrastructures. As such, the book provide a needed intellectual and conceptual reorientation of the cultural meaning of mass digitization.

Nanna Bonde Thylstrup is an assistant professor in the School of Communication and Culture at Aarhus University, Denmark. Her book is a thorough reconfiguration of her dissertation from the University of Copenhagen, properly updated with facts and new developments practically up to the point of publication. The author structures her argument in three parts. Part 1 is a new framing of the intellectual challenges of mass digitization through the theoretical constructs of "assemblage" and "infrapolitics." In Part 2, Thylstrup interrogates Google Books and the Europeana Collections in this new light and contrasts the corporate and collaborative nature of these two well-known projects with a selection of community-driven shadow digital libraries. Part 3 applies the conceptual frameworks introduced in Part 1 by cutting across and extending the interpretation of the three case studies.

Thylstrup establishes her argument in Chapter 1 with a brief history of origins and distributed efforts of early large-scale digitization projects that draw on the holdings of libraries, archives, and museums. Her review is not comprehensive in scope, but rather is a stage-setting exercise designed to reorient the reader from the bounded and parochial writing on mass digitization, which tends to focus on policy development or technical implementation. For Thylstrup, mass digitization is primarily about the aggregation of digital content or metadata across institutional and national boundaries at scale. The technologies and procedures of analog to digital conversion are far less interesting and important than the implications of the huge networked piles of digitized books, archives and manuscripts, photographs, and works of art.

The two theoretical concepts of "networked assemblage" and "infrapolitics" are fundamental to Thylstrup's analysis. The construct of "assemblage" has its roots in the efforts of scholars across a spectrum of humanistic inquiry to complicate the notion of the fixed and interpretable cultural archive in favor of mutable, absent, or imaginary evidence. "Thinking about mass digitization as an 'assemblage'

allows us to abandon the image of a circumscribed entity in favor of approaching it as an aggregate of many highly varied components and their contingent connections" (p. 23). In this book, assemblage serves simultaneously as a heuristic device for describing the database and metadata models, technical standards, networking platforms that enable the accumulation of digitized content across institutional and political boundaries, as well as an ontological device for articulating the transformation of power and control in new corporate or collaborative configurations.

Thylstrup is not unique among scholars in pointing out that the digital transformation of cultural heritage resources is no doubt political. The question for this author is how to characterize the politics at play in mass digitization assemblages. Getting to an answer in this book involves the re-coining of the concept "infrapolitics" as sort of mashup (assemblage?) of globalizing infrastructure and trans-border politics. Thylstrup adopts the term from the work of James C. Scott (2009) and others who focus on the politics of the "infra" or the micro processes hidden under the largely more visible processes represented by the engagements of geographically bounded entities such as countries or territorial alliances. "This volume suggests shifting the lens to focus on a different kind of infrapolitics, however, one that not only ties the space of resistance but also of maintenance and conformity, since the story of mass digitization is both the story of contestation and the politics of mundane and standard-seeking practices" (p. 25). Thylstrup thus mobilizes "infrapolitics" as a theoretical mechanism to bring to the surface the underlying structures of the "new normal" of networked digital content assemblages and expose the challenges to sovereign power presented by corporate, collaborative, and community driven assemblages.

The centerpiece of *The Politics of Mass Digitization* is a close read of the infrapolitics of three large digital assemblages, each one distinctive for the power relations of their creation and maintenance. Chapter 2 concerns Google Books, the seemingly grand scheme to digitize millions of books as a way to "organize the world's information." Google is deceptively transparent about its intentions while also notoriously protective of its proprietary technical processes, requiring non-disclosure agreements for all but the most mundane technical inquiries. Thylstrup traces the origins and infrapolitics of the Google Books project using the limited insider knowledge in published form, along with a few patent applications oddly chosen from among the many hundreds that Google has filed to protect its proprietary processes. Hers is a compelling story of a complex assemblage of technological affordances and processes that function at the level of the digital scanner and the scan operator, through interactions and contracts with dozens of libraries across the globe, and with interventions in areas of law and public policy. In Thylstrup's telling, Google Books is an excellent example of infrapolitics at work – the unseen and unknown interacting with the visible in a variety of political contexts.

If Google Books is an assemblage that exists in the context of the global corporation, the Europeana Collections is an equally borderless networked resource, but one that is structured to function well within the negotiated intragovernmental context of the European Union. Thylstrup's reading of Europeana in Chapter 3 constructs a story of the products that result from the marriage of standardized technical infrastructure and a political regime that seeks to protect the rights and prerogatives of the member states that choose to contribute to the shared platform. Europeana is low-risk aggregation,

centered on embracing a shared metadata model while deferring to contributors on the choice of content, all of which must meet the very high bar of intellectual property permissions. "Europeana produces a new form of cultural memory politics that converge national and supranational imaginaries with global information infrastructures" (p. 58). The infrapolitics of Europeana, as portrayed by Thylstrup, creates a deceptively seamless integration with Google, Facebook, Twitter, and other large-scale globalized network infrastructures and an assemblage of over 58 million digital objects that is deeply lacking in diversity, equity, and inclusion. "In reality it represents a highly fragmented image of Europe" (p. 76).

Nanna Bonde Thylstrup's inclusion of shadow libraries as a distinctive case study (Chapter 4) is a brilliant stroke of original scholarship. In her case study, Thylstrup considers three diverse shadow libraries. The first, Lib.ru, is a Russia-based file sharing platform initially developed by a single individual for the purpose of sharing books outside the boundaries of existing intellectual property regimes. The second, Monoskop, built collaboratively among Bratislava, Slovakia's thriving digital scene, is an attempt to establish an intellectual platform for the study of avant-garde cultures in a globalized environment that does not universally recognize the principle of fair use. The third shadow library, UbuWeb, is a non-commercial file-sharing site based in the United States and focused on sharing a wealth of English-language avant-garde sound art, video, and textual works.

On the surface, exposing the terms and conditions driving the creation of community-based and seemingly rogue assemblages of DIY digitized books, zines, and avant-garde art and performance is a jarring contradiction. Shadow libraries exist beyond the legal challenges of corporate digitization and beyond the complexities of cross-boundary collaboration on copyright and the canon, in an environment of community activism. Their existence within the structures of the global information infrastructure is a confrontation, a manifesto, a statement about the futility of bounding the flow of cultural property by country or organization of origin or by the shifting perspectives on intellectual property as manifested in a networked environment. "Shadow libraries instigate new creative relations, the dynamics of which are infrastructurally premised upon the medium they use" (p. 100). Thylstrup is working counterintuitively to raise the infrapolitics of large-scale assemblages to the level of a general principle, one that becomes transferrable to other forms of digital assemblage not revealed in her book.

In Part 3, Thylstrup broadens her analysis of mass digitization by deepening her critique of mass digitization as a global phenomenon that is driven not solely by corporate self-interest, border spanning collaboration, and stateless community collaboration, but rather by a rich combination of the these three forces. Her analysis is ambitious and challenging, but she never loses sight of the implications of infrapolitics and assemblages for simultaneously fostering and obscuring meaning making. Thylstrup directs the reader's attention to the scale (mass) of digital assemblages as a source of power and intimidation. She calls for a sociocultural analysis of mass digitization that goes well beyond the technical and the procedural to "acknowledge the pathologies" of digital assemblages in all of their rich psychological dimensions.

Toward this end, Thylstrup's creates in Chapter 5 a nuanced analysis of the tension that those who enter a product of mass digitization face between the comfort and pleasure of convenient networked access, on the one hand, and the confusion of the unknown landscape of the assemblage, on the other. "Caught up in the commodified labyrinth of the modern digitized archive, the digital flaneur of mass digitization might just as easily get stuck in a repetitive, monotonous routine of scrolling and downloading new things, forever suspended in a state of unfulfilled desire, than move about in meaningful and pleasurable ways" (p. 113).

Thylstrup is equally critical of the efforts to overcome this tension through the algorithmic engineering of serendipity. Programmed recommender systems combined with randomized presentation of "interesting" digital objects to the casual visitor seek to create an impression of comprehension and control, where neither of these exist. Corporate and collaborative assemblages thrive on the false belief that the whole is always greater than the sum of the parts. In her concluding Chapter 6, Thylstrup anchors her work neither in the reactionary language of loss and nostalgia nor in practical implications but rather in a scholarship that seeks to understand the transformative realities of globalization. "Mass digitization assemblages, and their globalization of knowledge infrastructures, thus crystalize the more general tendencies of globalization as a process in which people participate by choice, but not necessarily voluntarily; one in which we are increasingly pushed into a game of social coordination, where common standards allow more effective coordination yet also entrap us in their pull for convergence" (p. 30).

In a work of original scholarship that turns on the relevance of the chosen cases, there is always room for additional cases and alternative perspectives. Thylstrup presents a cohesive rhetorical argument that opens at least three avenues for further research that exists between the lines of her analysis. First, digital humanists would benefit from understanding the extent to which the infrapolitics of assemblages apply to aggregations of metadata and/or digital content constructed in academic contexts. Research should apply Thylstrup's analytical framework to the wide range of what Carole Palmer (2004) terms "thematic collections" and Katrina Fenlon et al. (2014) dub "scholar-built collections." Second, in focusing so intently on the assemblage, Thylstrup's analysis skips over the equally important infrapolitical issues associated with creating digital surrogates worthy of aggregation. Mats Dahlstrom (2010) for example calls for a "critical digitization" that seeks to interrogate the layers of decision making that influence the shape and meaning of a digital product, prior to even considering the implications for aggregation. Similarly, Paul Conway (2015) points to the importance of surfacing the archival properties of large collections of digital surrogates that are neither fixed nor transparent.

The Politics of Mass Digitization is likely to make an outsized contribution to ongoing debates about benefit and loss in the ongoing transformation of cultural resources from analog artifacts to digital assemblages. This is not the book to read for technical or procedural details about mass digitization projects. It is not the book to read for understanding the impact of mass digitization projects on the libraries, archives, and museums across the globe that are providing the content-substance of large scale assemblages. It *is* a book, however, that should reconfigure the discourse about the meaning of digital aggregations that live and thrive beyond the boundaries of cultural heritage organizations and

beyond the borders of nation-states. Digital assemblages of the sort theorized and interpreted in this volume are the ways in which we are increasingly constructing the reality of our past lives. Thylstrup relocates the discourse where it belongs now.

References

Conway, P. (2015). Digital Transformations and the Archival Nature of Surrogates. Archival Science 15 (1): 51-69. DOI: 10.1007/s10502-014-9219-z

Dahlstrom, M. (2010). Critical Editing and Critical Digitization. In Text Comparison and Digital Creativity: The Production of Presence and Meaning in Digital Text Scholarship, edited by W. Th. Van Peursen, Ernst Thoutenhoofd, and Adriaan van der Weel. Brill, pp. 80-97.

Fenlon, K. et al. (2014). Scholar-built collections: A study of user requirements for research in large-scale digital libraries. Proceedings of ASIST 2014, November 1-4, 2014, Seattle, WA, USA.

Palmer, C. (2004). Thematic Research Collections. In A Companion to Digital Humanities. Blackwell Publishing.

Scott, J. C. (2009). Domination and the Arts of Resistance: Hidden Transcripts. New Haven, CT: Yale University Press.