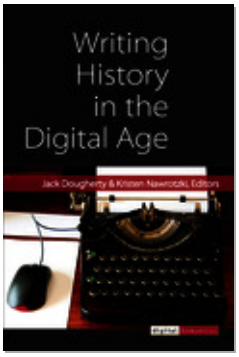
Writing History in the Digital Age

Jack Dougherty, Kristen Nawrotzki



Published by University of Michigan Press

Jack Dougherty. and Kristen Nawrotzki.
Writing History in the Digital Age.
Ann Arbor: University of Michigan Press, 2013.
Project MUSE. Web. 23 Sep. 2015. http://muse.jhu.edu/.

Historical Research and the Problem of Categories

Reflections on 10,000 Digital Note Cards

Ansley T. Erickson

Once while taking a break at an archive, I stood at the snack machine alongside a senior historian. She let out a tired sigh and then explained that she was at the beginning of a project, at the point "where you don't know anything yet." For historians, research often takes a nonlinear or even meandering form, through many phases of uncertainty and redefinition. As global historian William McNeill described it, we begin with a sense of a historical problem and explore it through reading, which cyclically "reshapes the problem, which further directs the reading." This back-and-forth can continue right up to publication. We might be more bold, like Stephen Ramsay, and celebrate the "serendipitous engagement" that happens when "screwing around" with sources, enjoying intellectually productive browsing and exploration. Whether we look forward to or struggle through these phases, much of our work happens while our research questions are still in formation.[1]

Uncertainty is, therefore, a core attribute of our research process, one that we might take as evidence that we are guided by our sources. Yet it can produce challenges as well. How do we proceed to do research—the real nuts and bolts of it—if we acknowledge such uncertainty? How can we organize information and keep it accessible in ways that will facilitate our ongoing thinking and writing, if we acknowledge changing focal points or areas of interest?

To research my dissertation, "Schooling the Metropolis: Educational

Inequality Made and Remade, Nashville, Tennessee, 1945–85," I started with various questions about desegregation in Nashville, Tennessee: Why did black students ride buses more and longer than white students? Was this due to power imbalances, ideologies, or explicit policies? Was the nature of Nashville's economy relevant? I gradually worked my way toward the question I came to address—how the pursuit of economic growth fed educational inequality.[2]

This essay considers a central challenge of historical research, one present in any long-term research endeavor but made more acute by shifting research questions: the challenge of information management. In the summer of 2006, I had a viable dissertation prospectus and was about to embark on the first of my trips to the archives. I was excited and I was scared that I would forget things. I knew what it took to manage the information involved in a seminar-length paper. Earlier, I had filled pages with handwritten notes or word-processed text, filtering through them as I built an argument. But how would I manage a project that would extend over years of research and writing? Where, in the most literal sense, would I put all of the information, so that I could find it when drafting chapters or, much later, revising for publication? I needed something that would backstop my own memory yet allow for shifts in my thinking. I also had to ensure that information stayed in the context of its originating source, while distinguishing between material from the sources and my interpretation of them.

Following the example of some more senior graduate students and one young faculty member in my department, I decided to use a relational database to keep my notes.[3] I was far from the cutting edge of digital history or information sciences. As I designed my database, I leaned on the very analogue metaphor of the note card. Rather than reconceptualizing my historical work in deep interaction with new tools, as many scholars in digital history (including several in this volume) have done, I used a new tool to do familiar aspects of research in a more accessible and efficient way.[4]

In the process, I came to see information management as a consequential aspect of historical research. How we organize and interact with information from our sources can affect what we discover in them. Scholars of the archive and of the social history of knowledge have long observed the consequences of how people keep information, and historians have considered the impact of archival practices on their own findings.[5] Their work raises useful questions about historians' own research processes—questions highlighted during work with databases. Particularly, where,

when, and how do we categorize information; how do we interact with these categories as we think and write; and what can we do so that we do not become bound up in the categories we create at the most uncertain stages of our research?

Although the quantity and functionality of digital tools for data management, as well as attention to these tools, has increased in the last few years, they are not yet fully woven into the fabric of the profession. Some of this may be generational; but it also results from our discipline's relative lack of formal conversation about methodology at the granular level. Graduate training programs paradoxically structure their training as internships in the consumption and production of history yet offer little explicit guidance on the mechanics involved.[6] When new tools emerge, their potential utility may not be appreciated fully. Database programs can have broad impact on how we interact with information, but much discussion of them emphasizes their use in the narrower work of bibliographic and citation management.[7]

While neither an early nor an innovating database user, I offer this account to illustrate some potential benefits and learnings from my modest use of this tool. I first lay out how I organized my research and how it related to my thinking and writing. (See images that document my process in the web version of this essay at http://WritingHistory.trincoll.edu.) Then I venture some connections between that process and questions in the social history of knowledge and the scholarship of the archive—questions about the making and impact of categories in thought.

Database Note Keeping

Having decided to keep notes in a database, I selected a program: FileMaker Pro. There are many alternatives: some designed for qualitative research (NVivo, Atlas.ti), some free and web-compatible (such as Zotero), and others emerge periodically.[8] Historians who write code can create their own. I began by creating two FileMaker layouts, one for sources and another for the "note cards" from those sources.[9] Guessing at how I might later sort and analyze my notes, I made a keyword field for themes I expected to recur. Zotero, which I use in current projects, provides a similar structure for sources, notes, and keyword "tags."

In trips to several archives over a year, I collected tens of thousands of pages of documents by taking digital photographs of these.[10] I read and took notes on a portion on site, in those collections that prohibited digital

copying or charged exorbitantly for physical copies. Because I had very limited time to work on-site at archives, most of my note taking happened once I returned home. I read digital copies on one screen. On the other, I entered notes in the database, putting direct quotes in one field, my observations and tentative analysis in another (see fig. 5). (Zotero uses a single note-taking field.) The vast majority of my note cards were descriptive, but when I had a thought that tied various sources together or hinted at an argument, I made a new note card, titled "memo to self," and then these entered the digital stack as well, tagged with keywords.

Once I had worked through most of my documents, I had nearly 10,000 note cards. I used the database as I began my analysis and sense making. I first ran large searches based on my keywords: searching hundreds of note cards on "vocational education," for example. I organized these cards chronologically—an action that takes only a few keystrokes—and spent a day or more reading them through. As themes or patterns began to emerge or as there were connections to other sections of my research that were not under the "vocational" heading, I ran separate searches on these, incorporating that material into the bin of quotes and comments I was building by cutting and pasting into a new text document. (Databases often have "report" functions that could help this process, but I did not explore that route.) Of course, sorting information can be done without a database. But I found it to happen quickly and more easily with one.

Having reviewed my research material, I began to draft a section of a chapter. I started to write before I was sure of the precise structure of the chapter or my detailed argument. I used writing as a way to find and refine my argument. Crafting a basic narrative often helped me identify what I was missing, what I needed to find out more about. Writing in this exploratory fashion was made easier by quick access to bits of information from the database as needed.[11]

Using a database did accomplish the most basic of my goals. It proved a reliable and convenient way to keep notes and contextual information in the same place, and it addressed my most basic fear of forgetting, by allowing searches for information in myriad ways—by title, content of notes, direct quotations, keywords, dates. As my writing advanced, I came to appreciate how the database's full-text searchability allowed me not only to follow my original questions but to explore ones that I had not anticipated at the start of my research. This mode of note keeping allowed me, as I thought and wrote, to access information that I would have missed otherwise—likely because of the difficulty of tracking down and reordering notes without such a database. Two examples illustrate this accessibility.

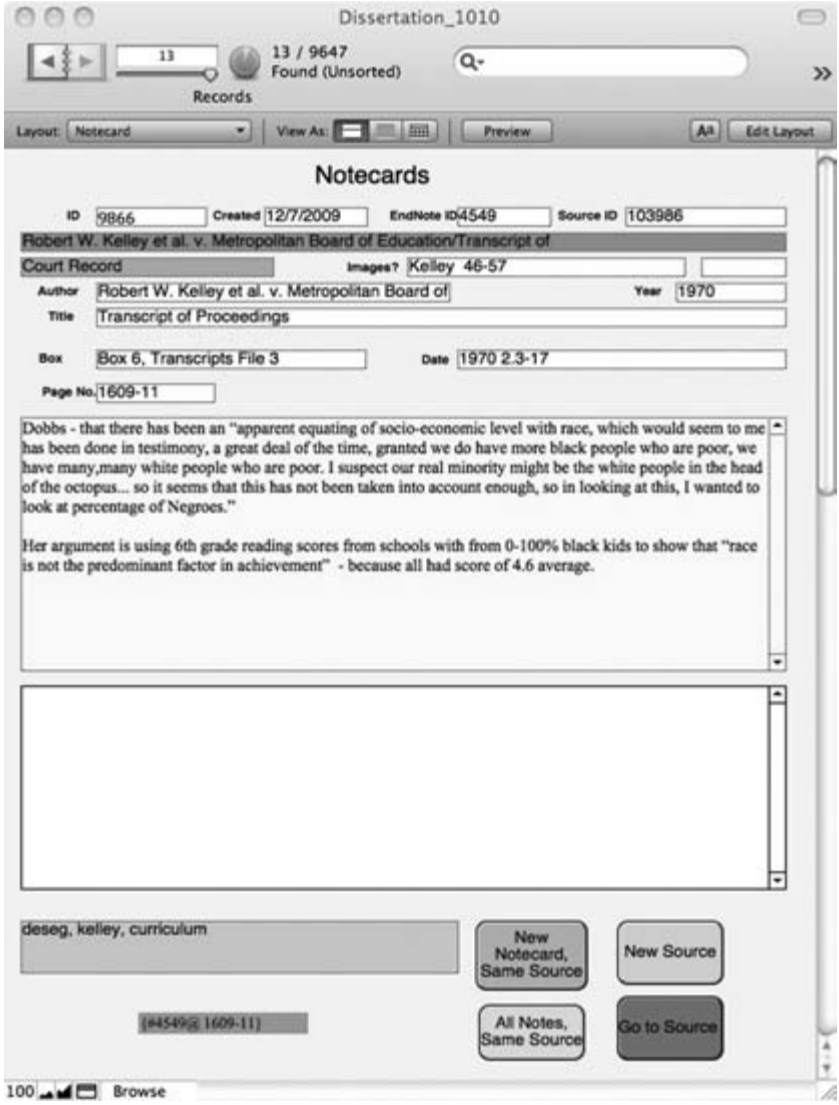


Fig. 5. FileMaker Pro screenshot of sample notes on a court transcript

One central problem in my work has been understanding the multiple layers of inequality at work in Nashville's desegregation story. There are, of course, salient and central differences by race and by class, but these divisions were often expressed in the language of geography. By the mid-1960s, residents, planners, and educators used the phrase *inner city* to indicate predominantly black neighborhoods or neighborhoods where planners predicted black population growth. I had noticed this pattern in my own reading and had captured examples of such language and other descriptions of geographic space with a keyword: *cognitive map.* To read about this phenomenon, I worked through all of my "cognitive map" notes, in chronological order. Through several conference papers and draft chapters, I developed an argument about how pro-suburban bias informed Nashville's busing plan. In early versions, I seemed to imply that in Nashville residents' cognitive maps, the correlations between suburban space and white residents and between urban space and black residents were absolute. But were there exceptions? What could I do to test this? I searched for instances where my sources used the phrase *inner city.* Of course, I may not have not written down each instance, as I did not plan for this textual analysis. Nonetheless, I had enough to begin.

When I read my sources in this way—some of which I had labeled "cognitive map," some not—I saw something new. Among the critics of schooling in the "inner city" and the smaller group of its defenders, there was a case that proved that the identification of urban space with black residents was not complete, at least for some city residents. I had earlier made notes and then forgotten about the story of a central-city school that was historically segregated white, remained largely working class, and had a local council representative fighting to retain the school in conjunction with what he labeled its surrounding inner-city neighborhood. William Higgins, the council representative, asked, "You're taking children from the inner city and busing them to suburbia. Why place the hardship on them? Why not bring children from suburbia to the inner city?" He later proposed, "All new schools . . . should be unified with the inner-city, otherwise the city finds itself a lonely remnant, disunited and eventually abandoned."[12] When I read these passages in the first years of my research, I had not thought to tag them with the keyword *cognitive map.* Thus they did not show up in that keyword search over two years later. I was able to discover them again because I could search for a phrase laden with meaning and insinuation. Doing so yielded access to notes that influenced my understanding of how categories of race, geography, and class overlapped in my story and where they diverged.

In another case, I found that the database allowed me to reframe an initial research question into a broader one. From the start, my dissertation was concerned with why schools were built where they were, how locations got chosen, to suit whose interests. I thought of schools as a good being struggled over in political and economic terms. After analyzing the local politics of school construction, I understood that my story was not about schools alone but about how the distribution of public goods reflected the political and economic structures that supported metropolitan inequality.

I had been tracing how urban renewal funds subsidized school construction and how, in the context of a metropolitan government, such subsidies could allow a municipality to shift more of its own tax revenues to its suburban precincts. I suspected that this use of urban renewal dollars to reduce the local commitment to supporting city areas in favor of suburban ones was visible in other areas of city services as well. How could I illustrate that broadened claim? I could see what my sources—planning reports, maps, records of community meetings—said about another kind of public good, to see if the dynamics were similar. I knew that I had made some notes about the building and repair of sewer lines for the city and surrounding suburbs, but I had not expected to write about them, so I had no related keyword. Text searchability of the database meant that I could very easily track down everything I had about sewers, organize it chronologically, and test if the pattern I saw for schools fit for sewers as well. Without fully searchable notes, I would have been looking through stacks of note cards, organized to fit another set of categories entirely. I may not have felt I had the time to expand my original question to a broader one.

In each of the examples just presented, the database helped relevant information jump out of the noise of years of research and thinking. It helped make that information available relationally, easily connected to other information.

Categories and the Making of Historical Knowledge

Reflecting on my use of this digital tool for note keeping has led me to questions about how we think about our research practice, how we understand the relationship between how we research and what we learn. Recent work in the social history of knowledge and the history of the archive share a core interest in categories—where they come from, what assumptions or values they represent, how they can be reified on paper or in practice.[13] These interests are relevant to our research methods. In researching and writing my dissertation, I was able to set out initial categories of analysis

(via keywords), but it was possible, at no great expense of time, to throw these out. Sometimes I used my initial keywords, and sometimes I skipped over these to evaluate new connections, questions, or lines of analysis. If I had used pens and notebooks or a set of word processing documents, regrouping information would have required a great expenditure of time. I would have been less likely, then, to consider these new avenues, and my earlier categories of analysis would have been more determinative of my final work. Those categories would have been highly influential even though I created them when, in the words of the historian at the snack machine, I really did not know anything yet. Since there was virtually no time expended in trying out new questions utilizing the database, I could explore them easily. Thinking about how my database facilitated my analysis got me thinking about how historians construct, use, and rely on categories in our work.

It makes sense that historians would think about categories, as we encounter them frequently in our work. As graduate students, we learn to identify ourselves by subfield: "I do history of gender" or "I'm an Americanist." We are trained implicitly and explicitly to organize information and causal explanations into categories of analysis—race, class, gender, sexuality, politics, space, and so forth—when, in fact, these categories are never so neat and separate, whether in an individual's life or in a historical moment. Then we research in archives that establish and justify their own categories—legal records divided by plaintiff or defendant, institutions that keep their records with an eye to confirming their power or reinforcing their independence. To make sense of a sometimes overwhelming volume of fact, all of which needs to be analyzed relationally, we rely on categories that we create as we work—like my database keywords.

This matter of categories connects to at least two fields of scholarship. Scholars of the history of knowledge, such as Peter Burke, have examined the organizational schemes embodied in curricula, in libraries, and in encyclopedias and have shown how these structures and taxonomies represent particular ways of seeing the world. For Burke, such schemes reify or naturalize certain ways of seeing, helping to reproduce the view of the world from which they came. They also make some kinds of information more accessible, and some less.[14]

Think, for example, of the encyclopedia. We are accustomed to its alphabetical organization of topics, but this structure, in fact, represented a break from previous reference formats that grouped subjects under the structure of classical disciplines. The alphabetized encyclopedia came about at a point when the previous disciplinary categories no longer could

contain growing knowledge. A new, more horizontal model took their place, a model that allowed readers access to information by topic, outside of the hierarchies of a discipline. Burke points us to the importance of how we categorize information, where these categories come from, and how categorizations affect our access to and experience of information.[15]

Anthropologist Ann Stoler comes to the problem of categories from a different perspective. She thinks of the archive as an active site for ethnography and seeks to understand how archives are live spaces in which the Dutch colonial state in Indonesia built, among other things, social categories. She traces how colonial administrators' use of archiving categorized and assigned particular rights and privileges to people with different national heritages. As they categorized, they made some peoples' experiences of the colonial state visible and obscured others. Stoler writes that categories are both the explicit subject of archives and their implicit project: "The career of categories is also lodged in archival habits and how those change; in the telling titles of commissions, in the requisite subject headings of administrative reports, in what sorts of stories get relegated to the miscellaneous and 'misplaced.'" She then frames the archive as a place to understand "how people think and why they seem obliged to think, or *suddenly find themselves having difficulty thinking*," in certain ways.[16]

The work of scholars like Burke and Stoler implies questions for historians' research processes. Burke's work suggests that we investigate how categories of thought, either between disciplines or within them, affect us. Think of academic subfields, for example, the boundaries of which still shape the literatures we read (even as many try to transcend them) and still guide which archives we pursue or whether we think of particular questions as part of our domain. Stoler raises a different kind of question. At what points in our research, out of pragmatic necessity, out of a desire for intellectual order, or for yet other reasons, do we set out categories of evidence or thought that influence what we see and what we do not see? What kinds of tools could help us be more aware of these categories or could give us the flexibility to move beyond them when necessary or desirable?

I hypothesize here that databases offer a kind of flexibility that can allow us to create and re-create categories as we work with notes, to adjust as we know more about our sources, about how they relate to one another and how they relate to the silences we are finding. That flexibility means that we can evaluate particular ways of categorizing what we know and then adapt if we realize that these categories are not satisfactory. In doing so, we are made more aware of the work of categorization and are reminded to take stock of how our ways of organizing help and what they leave out.

The matter of flexible categorization touches on another strand of scholarship: archivists debating what postmodernism means for their work. How does the growing understanding of archives as spaces in which certain kinds of power are codified and justified and where information has to be understood relationally matter for the practice of archiving? Archival theorist Terry Cook argues that finding aids and item descriptions should be constantly evolving, adapting to new relevant knowledge about the item's sources and its relationship to other archived and unarchived materials.[17] Working with databases provokes historians to think about how our note-keeping practices could seek such flexibility and relationality.

Yet there are at least two cautions as well. One comes from the flatness of databases like the one I used. In Burke's terms, my database was not a reference text organized along disciplinary lines. It was more like an alphabetized encyclopedia. Without hierarchies that keep each fact locked in relationship to others—through the structure of earlier historiography, for example, or through the categories of an archive's collections—the historian has to be more intentional about seeing information in its context. If we can look across all of our notes at a very granular level and make connections across categories that we or others created, it becomes too easy to look at these bits of information devoid of context—a danger visible even in my own way of cutting and pasting out of my database. I linked bits of notes only to a source code, meaning that they could be read in less-than-direct connection to their origins. Digital bits seem very easily severed from their context. Zotero's structure links sources and notes visually, which may help safeguard against this.

More important, despite its usefulness in helping us see things we might otherwise have forgotten or missed, no database does the work of analysis. The two are, of course, interdependent—as they are in any digital or non-digital form of note keeping. The analytical work, the crucial sense making that pushes history writing from chronology to critical interpretation, still happens in our own heads. There other implicit categories or habits of thought might shape our analysis. There we decide whose stories to tell first, or we prioritize one set of historical drivers over another. Some of these habits reflect our deepest-held assumptions and beliefs. It is less easy to talk of these, and certainly less easy for an author to identify his or her own, than it is to speak of note keeping. Maybe bringing critical consciousness to the mechanical can prompt more reflection about the conceptual as well.

It is also worth considering what kinds of concerns may arise for histo-

rians who have not yet made use of digital tools like databases in their own research. Historians surely value, maybe even romanticize, the encounter with sources in the archives. Does converting that textual, even textural, experience into digital note cards somehow deaden it? Does it render our research uncomfortably close to a social scientist's coding and writing up of findings? Charlotte Rochez, responding to an earlier version of this essay, explained that she worried about sacrificing "some of deeper insights, interpretations and understanding induced from being more involved in sorting and interpreting the sources."[18] Digital note taking may add to but does not of necessity replace varied encounters between researcher and sources—even "serendipitous engagement." It remains possible to meander through your notes from a given collection or source, to look back at the original page (even in PDF or photocopied form). But it becomes newly feasible to look broadly across those collections and sources.

One prompt for this volume came from the *Journal of American History*'s 1997 special issue that made public the process of academic peer review. David Thelen's introduction to that issue raised questions about the work of history writing that seem important to revisit in light of digital innovations. The centerpiece of the issue was a submission by Joel Williamson, in which Williamson recounted his failure to perceive lynching's centrality to and origins in American and Southern history. Two reviewers received Williamson's piece with shock and dismay that he could have missed what they had appreciated as central for years. Despite this disagreement, or perhaps because of it, Thelen saw Williamson's piece as issuing a challenge to historians to "think about what we see and do not see, to reflect on what in our experience we avoid, erase, or deny, as well as what we focus on."[19] I see my attention to categories, to the possibilities and implications of how we choose to organize the information on which our interpretations rest, as a kindred effort.

Acknowledgments: The author thanks Jack Dougherty and Kristen Nawrotzki for the invitation to reflect on research practice and for good feedback on this essay, Courtney Fullilove for reading suggestions, Seth Erickson for ongoing conversations about archives and information architecture, and all those who commented on earlier versions for their helpful remarks. The dissertation research described here was supported by a Spencer Dissertation Fellowship, a Clifford Roberts/Eisenhower Institute Fellowship, and a Mellon Interdisciplinary Graduate Fellowship at the Paul Lazersfeld Center, Institute for Social and Economic Research and Policy, Columbia University.

Notes

1. McNeill quoted in Tracy L. Steffes, "Lessons From the Past: A Challenge and a Caution for Policy-Relevant History," in *Clio at the Table: Using History to Inform and Improve Education Policy*, ed. Kenneth Wong and Robert Rothman (New York: Peter Lang, 2008), 267–68; Stephen Ramsay, "The Hermeneutics of Screwing Around; or What Do You Do with a Million Books" (paper delivered at the conference "Playing with Technology in History," History Education Network/Histoire et Éducation en Réseau, Niagara-on-the-Lake, Canada, April 2010), http://www.playingwithhistory.com/wp-content/uploads/2010/04/hermeneutics.pdf.

2. Ansley T. Erickson, "Schooling the Metropolis: Educational Inequality Made and Remade, Nashville, Tennessee, 1945–1985" (PhD diss., Columbia University, 2010), http://faculty.tc.columbia.edu/upload/ate11/Erickson_TitlePage&Abs_DEP.pdf.

3. A relational database keeps bits of information in relationship to one another without establishing hierarchies. In my database, as discussed in this essay, each source was related to multiple notes, yet I could also search across notes from multiple sources. More elaborate uses of relational databases can include specifying types of relationships between sources or analyzing the density of relationships between attributes of the database so as to discern patterns that may not be otherwise apparent.

4. Consider other examples of database technology in historical writing, such as the archival collection *as* database (full-text searchable to facilitate text mining) or the database as scholarly product. See Jean Bauer, The Early American Foreign Service Database, 2010, http://www.eafsd.org/.

5. See, for example, Peter Burke, *A Social History of Knowledge: From Gutenberg to Diderot* (Cambridge: Polity, 2000); Ann Laura Stoler, *Along the Archival Grain: Epistemic Anxieties and Colonial Common Sense* (Princeton: Princeton University Press, 2010); Antoinette Burton, ed., *Archive Stories: Facts, Fictions, and the Writing of History* (Durham: Duke University Press, 2005); Joan M. Schwartz and Terry Cook. "Archives, Records, and Power: The Making of Modern Memory," *Archival Science* 2 (2002): 1–29.

6. For views of this problem, see Steffes, "Lessons," 267–68; Amanda Seligman, comment on "Historical Research and the Problem of Categories," in *Writing History in the Digital Age*, web-book ed., Fall 2011 version.

7. For one recent example, see Amanda Morton, "Review of Digital Tools: Zotero and Omeka," *Journal of American History* 98, 3 (2011): 952, http://jah.oxfordjournals.org/content/98/3/952.

8. FileMaker Pro, http://www.filemaker.com; NVivo, http://www.qsrinternational.com/products_nvivo.aspx; ATLAS.ti, http://www.atlasti.com/; Zotero, http://www.zotero.org. New options include DevonThink and Adobe Lightroom, discussed in the "History and the Digital Image Forum," *Perspectives* (October 2012), http://www.historians.org/perspectives/issues/2012/1210/index.cfm.

9. You can download my FileMaker template at http://faculty.tc.columbia.edu/upload/ate11/Erickson_FMTemplate.fp7.

10. Kirklin Bateman et al., "Taking a Byte out of the Archives: Making Technology Work for You," *Perspectives*, January 2005, http://www.historians.org/perspec-

tives/issues/2005/0501/0501arc1.cfm; Konrad Lawson, "The Articulated Arm of an Archive Raider," *Chronicle of Higher Education*, *ProfHacker*, December 7, 2010, http://chronicle.com/blogs/profhacker/the-articulated-arm-of-an-archive-raider/29243.

11. Lynn Hunt argues for a similar approach, encouraging scholars not to delay writing by organizing notes or other getting-ready activities, as well as to use writing to further thinking, in "How Writing Leads to Thinking (and Not the Other Way Around)," *Perspectives*, February 2010, http://www.historians.org/perspectives/issues/2010/1002/1002art1.cfm. Digital note keeping helps earlier and more fluid writing, as the accessibility of information reduces barriers between getting ready and actually writing. James B. McSwain shared a similar story about his experience with Nota Bene in *Writing History: How Historians Research*, *Write*, *and Publish in the Digital Age*, ed. Jack Dougherty and Kristen Nawrotzki, October 6, 2010, http://writinghistory.wp.trincoll.edu/2010/10/06/organize/#comment-148. I wrote more about my approach on the same site, in "Keeping the Writing (and Thinking) Going," http://writinghistory.wp.trincoll.edu/2010/10/06/erickson-thinking/.

12. William Higgins, "Suggestions for Development of Guidelines for an Unitary Plan for the Metropolitan Board of Education," 1979, box 69, file 8, Kelly Miller Smith Papers, Vanderbilt University Special Collections and University Archives; Saundra Ivey, "School Closing Plan Draws Fire," *Tennessean*, November 23, 1977.

13. See the works cited in note 5, and see Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, MA: MIT Press, 1999).

14. Burke, *Social History*.

15. Burke, *Social History*, especially 184–87.

16. Stoler, *Along the Archival Grain*, 36.

17. Terry Cook, "Fashionable Nonsense or Professional Rebirth: Postmodernism and the Practice of Archives," *Archivaria* 51 (Spring 2001), http://journals.sfu.ca/archivar/index.php/archivaria/issue/view/428/showToc.

18. Charlotte Rochez, comment in *Writing History: How Historians Research*, *Write*, *and Publish*, http://writinghistory.wp.trincoll.edu/2010/10/06/organize/#comment-803.

19. David Thelen, "What We See and Can't See in the Past: An Introduction," *Journal of American History* 83, no. 4 (1997), http://www.jstor.org/stable/295898. This idea also supports Fred Gibbs and Trevor Owen's call for "a new kind of methodological transparency," in "The Hermeneutics of Data and Historical Writing" in the present volume, pp. 159–70.