Interest from academics in the humanities and social sciences in studying the cultural dimensions of computing can be traced back at least as far as the early 1980s. As personal computers became increasingly common in homes, offices, schools, and universities, scholars like Jay David Bolter, Lucy Suchman, and Sherry Turkle began the work of adapting the research methodologies of their respective disciplines to study computing in these new contexts. Personal computing devices and the myriad of cultural activities we juggle through them have since become more and more complicated; however, at the same time user-friendly approaches to design encourage us to take this complexity for granted. Indeed, as Eszter Hargittai and Christian Sandvig note in their introduction to *Digital Research Confidential: The Secrets of Studying Behavior Online*, there is now often little institutional reward for writing about how the broad integration of computing affects academic research about culture and behavior. In this respect, their collection serves as an effective argument for the value of sustaining in-depth conversations on the effects that constantly changing technological conditions have on research methods. It is also an engaging introduction to the wide range of research in these fields being conducted digitally.

Not to be confused with the “digital humanities” — a field largely devoted to studying and preserving pre-twentieth century texts due to the copyright’s influence over digitization efforts — the “digital media studies” represented in Hargittai and Sandvig’s collection focuses on contemporary and often digitally created cultural activities. This interdisciplinary research area involves scholars from a variety of disciplines including, but not limited to, literature and language, writing studies, history, anthropology, sociology, communications, information science, and computer science. Representing every methodological approach in this incredibly diverse field in a single volume would be an impossible task. Hargittai and Sandvig have wisely chosen to limit their collection by focusing primarily on scholars whose work involves Web 2.0 technologies. This decision gives each of the essays in the book some common ground while still allowing the collection to highlight the breadth of subjects covered by digital media studies. The collection discusses projects involving the Internet Archive’s Wayback Machine, YouTube, Twitter, computer-aided drafting software, Amazon’s Mechanical Turk system, Wikipedia, Flickr, homemade webcrawlers, and Second Life.

Hargittai and Sandvig’s excellent introduction is able to synthesize the wide-ranging research trends in this field around a tension between digital media as instrumentation and digital media as object of study. Digital research in the humanities and social sciences is often framed in one of two ways: either online platforms that offer “a new kind of microscope” allowing us to understand an area of offline behavior we are already familiar with in a new way, or online social activity sharing enough similarities to offline behavior that we can comfortably transfer our existing assumptions about how humans relate to one another into these new contexts. Both of these approaches assume that theory and method can be discussed separately. Most research leveraging these approaches, in other words, assumes either that digital tools add to — but don’t disrupt our — core understandings of cultural and behavior or that online activities function as relatively seamless extensions of offline behavior. Yet as Megan Sapnar Ankerson concludes in her essay, each project in the collection involves a moment when a researcher “could not help but notice the ways (their) entire engagement with (their) project was thoroughly organized through software” (p. 47). While the widespread acceptance of particular methods in the humanities and social sciences allows for basic assumptions about culture and society shaping them to be taken for granted, digital tools are often designed with other needs in mind. Conducting social research...
in digital contexts, which is to say through software, requires us to revisit and re-evaluate them against the affordances and constraints of the tools we use. Digital Research Confidential thus makes a strong argument for both the practical and analytical value of software and data carpentry in the humanities and social sciences.

The ten essays that follow the introductory chapter can be roughly divided into four categories. The first two essays following the introduction explore how database construction shapes archival research practices. Megan Sapnar Ankerson shares her experience working with the Internet Archive's Wayback Machine. Initially taking the project's slogan to “surf the web as it was” at face value, Ankerson eventually realized that some sites were reproduced in a form that had never existed in the first place due to the way that the project’s webcrawler tried to fill in gaps in content by selecting temporarily adjacent versions of files with themselves a part of these debates in the sense that what they capture, recreate, or wholly ignore reflect beliefs held by its creators about what the web would become. In the chapter that follows, Virág Molnár and Aron Hsiao discuss a project on tracing the evolution of flash mobs using recordings published on YouTube. In addition to sharing some of their results, Molnár and Hsiao discuss how they documented their retrieval process in the interest of reproducibility, both so they could update their work later but also so that others might have a model to follow in similar projects. Yet when revisiting their results after the initial capture period, they discovered that YouTube’s search interface had been altered, leaving them unable to reproduce their procedures. Like Ankerson, Molnár and Hsiao’s essay demonstrates how digital research methodologies are influenced even before a project begins by technical decisions that researchers ultimately have no control over. While neither essay attempts to dissuade researchers from using public databases, they both point to a need to address the fact that decisions made in the name of efficiency or usability often have consequences for researchers that are not anticipated by software designers.

Three essays in the collection discuss ethnographic methods, each demonstrating that close observation of the social use of technology requires creative adaptation to the various ways that technology is incorporated into our personal and working lives. The first, by danah boyd, discusses her work in documenting the roles that social media play in the lives of teenagers; however, boyd’s essay rarely mentions technology. As she explains, her interviewing techniques reflect the ways that technology flitters in and out of their lives. While the Internet remains a significant conduit for teenage social behavior, she notes that it is rarely a focus of the narratives they tell about their lives. For anyone already familiar with boyd’s work, this observation is not new, but the chapter offers important insights as to how she learned to adapt her ethnographic interest in computing to study a group that does not spend much time thinking about it. Technology also appears, at least initially, to exist at the margins of Paul M. Leonardi’s essay on observing automobile engineers. Leonardi recounts his time conducting field research within a private company, offering a wealth of practical advice for every step of the process from explaining the idea of software as data in an Institutional Review Board proposal, to gaining the trust of management while on site. More importantly for the aims of Harigitäi and Sandvig, Leonardi’s work demonstrates the importance of incorporating a technological awareness into his observational method, as Leonardi is able to discover many of the disagreements he documents between engineers result from the way their software resolves conflicts when trying to integrate components designed by different groups before an impact test. Finally, Amy Bruckman, Kurt Luther, and Casey Fiesler’s essay examines the role of Institutional Review Boards in universities, placing many of the same ideas raised in these two earlier essays into a historical context, showing how the social norms of online environments often diverge from the privacy and protection needs of offline social contexts.

There is now often little institutional reward for writing about how the broad integration of computing affects academic research about culture and behavior.

Crowdsourcing has already proved effective in a number of research contexts, but as a pair of essays here show, the patterns of engagement it affords may conflict with some
best practices of survey design and academic division of labor. Eric Gilbert and Karrie Karahalios’ essay on developing a Twitter interface based on theories of “social ties” explores the difficult task of balancing the expectations of social science research against the feedback from user testing. Despite succeeding from a technological perspective, a question of representativeness lingers over the project, as start-up style beta testing is not necessarily compatible with norms of sample selection in the social sciences. Following their essay, Aaron Shaw’s narrative of incorporating crowdsourcing into a large-scale content analysis project discusses the problem of user error in research software and expands into a broader reflection on how academic labor is valued. But this decision comes with its own time-consuming problems. Whereas professionally trained research assistants often disagree over interpretation, Shaw notes that disagreements in crowd-sourcing are often also the result of usability issues in the project’s software interfaces and documentation. Shaw’s account of addressing miscommunications between himself and software engineers on the importance of collecting data in specific ways will ring true to anyone who has participated in complex collaborations and is a must read for anyone considering one, regardless of their role.

The collection also includes a trio of essays on doing social research using “big data” retrieved from online sources. In their conversationally styled essay, Michelle Shumate and Matthew S. Weber offer two complementary, detailed defenses of software carpentry through a discussion of programming web-crawling tools. Their essay offers a look at the same issues raised by those before them in the collection but from the other side of the interface. While the labor of programming may not yet be institutionally rewarded in the humanities and social sciences, Shumate and Weber offer a strong argument for the ways that it can enrich those products of research that are. Brent Hecht and Darren Gergle extend discussion of this topic by examining how disparate data sources often make very different assumptions about metadata and its presentation within a database. Their essay raises important questions about the bias of certainty present in research involving big data. Similarly, Brooke Foucault Welles’ discusses how larger data sets can actually create a greater degree of uncertainty by examining how seemingly familiar concepts like “friendship” are understood via comparatively narrow observational contexts. In addition to offering detailed looks at the labor behind large-scale data collection and preparation, these three essays show that close involvement in the algorithmic processes performing it will lead to increased confidence in the theories derived from it.

*Digital Research Confidential’s* breadth, in short, is ambitious and representative of the way that digital media research in the humanities and social sciences is often peppered across academic departments on most campuses rather than centralized under a single umbrella. Readers will likely find that not all subjects or methods offer insights applicable to their own research. For instance, someone interested in combing through digital archives may not find the lengthy discussions on gaining trust during interviews as considerations of how databases prejudice our sense of historical memory.

Each project involves a moment when a researcher could not help but notice the ways their engagement with their project was thoroughly organized through software.

If Hargittai and Sandvig have aimed the book at fostering conversations among seasoned practitioners in the fields represented about the long-term effects of recent developments in Internet-based technologies on their home disciplines, then in that regard it succeeds wonderfully even if every reader will take away something different from it. Nonetheless, each chapter is generally written in a manner beneficial to specialists and non-specialists alike, making the collection also an excellent choice for an upper-level classroom, scholars already wrestling with similar problems themselves, or technological professionals looking to better understand the computationally-driven research of their humanist and/or social scientist collaborators.

**Reviewer Information**

Michael L. Black is an Assistant Professor of English as UMass Lowell, Lowell, MA. His current book project examines the influence of market pressures and technical discourses on popular conceptions of software usability and personal computing.