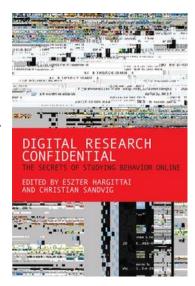Eszter Hargittai and Christian Sandvig (Eds.), **Digital Research Confidential: The Secrets of Studying Behavior Online**, Cambridge, MA: MIT Press, 2015, 271 pp., $26.00 (paperback).

Reviewed by
Lisa Farman
Ithaca College, NY, USA

What *really* happens during the digital research process? In *Digital Research Confidential: The Secrets of Studying Behavior Online*, editors Eszter Hargittai and Christian Sandvig have collected real-life accounts from established and rising scholars, providing the reader with a behind-the-scenes look at research projects involving the Internet and digital media. This volume contains an introductory chapter by the editors, followed by 10 chapters written in first-person, each dedicated to a different digital research project and its challenges.

The book is advertised as a follow-up to Hargittai's previous edited volume, *Research Confidential*, and its mission is largely the same. In the introductory chapter, Hargittai and Sandvig suggest that method sections of scholarly articles are often too defensive. They implore the scholarly community to reconceptualize "method," arguing we should think of it as a creative act and be more open to discussion about the process. In this spirit, the editors invoke the term "bench social science" to describe the accounts contained in this volume, all of which are transparent about the authors' struggles along the way.

Not surprisingly, its sole focus on digital research is where *Digital Research Confidential* departs from its predecessor. The editors argue that now more than ever we need to be more transparent about the research process, as much of digital media research is uncharted territory and the research area continues to develop. They offer two frames for how to think about digital media research—digital media as instrumentation, or an "unprecedented tool we can use to see ourselves" (p. 6), and digital media as a new object of study, where "pointing an existing method at a new object [the Internet] often raises difficult intellectual questions related to method" (p. 7). Unlike the previous volume which was organized around themes, this time it's up to the reader. The chapters in this volume are not explicitly tied back to these frames, and many chapters arguably address both.

The chapter topics run the gamut in terms of the type of digital media used or studied. The first two chapters, Megan Sapnar Ankerson's "Read/Write the Digital Archive: Strategies for Historical Web Research" and Virág Molnár and Aron Hsiao's "Flash Mobs and the Social Life of Public Spaces: Analyzing Online Visual Data to Study New Forms of Sociability," focus on the many challenges created by the changing nature of online content, particularly when researchers are trying to do longitudinal or historical research. The next two chapters tackle the challenges of doing fairly traditional ethnographies that also incorporate digital environments. In "Making Sense of Teen Life: Strategies for Capturing Ethnographic Data in a Networked Era," danah boyd writes about exploring teens' engagement with social media, and in

Paul M. Leonardi's "The Ethnographic Study of Visual Culture in the Age of Digitization," he tackles how ethnographers might record and write about people's interactions with digital artifacts. In "Social Software as Social Science," Eric Gilbert and Karrie Karahalios talk about building "social software" (in this case, a Twitter application) to study more traditional concepts like tie strength. Aaron Shaw's chapter, "Hired Hands and Dubious Guesses: Adventures in Crowdsourced Data Collection," wrestles with the problems created by using strangers as research assistants—in this case, what to watch out for if you outsource coding tasks via Amazon Mechanical Turk.

The next three chapters get a bit more technical, focusing on big-data research techniques. In "How Local is User-Generated Content? A 9000+ Word Essay on Answering a Five-Word Research Question," Brent Hecht and Darrent Gergle discuss mapping geographic points of origin in user-generated content repositories (in particular, Wikipedia and Flickr). In "The Art of Web Crawling for Social Science Research," Michelle Shumate and Matthew S. Weber discuss using a Web crawler to collect data. They touch on topics including the relative merits of building your own web crawler versus using existing software, as well as the importance of knowing how your own technology works. Brooke Foucault Welles wrote "Big Data, Big Problems, Big Opportunities: Using Internet Log Data to Conduct Social Network Analysis Research," which is the most accessible of the three chapters. She talks about problems created by using data provided by a company like Second Life, including how to position your project to an Institutional Review Board, the importance of having a detailed codebook, and philosophical struggles regarding the relationship between users' online and offline behavior. The final chapter, "When Should We Use Real Names in Published Accounts of Internet Research" by Amy Bruckman, Kurt Luther, and Casey Fiesler, deals with old ethical challenges made new again, such as when anonymity might actually hurt participants and whether to gather participant feedback before you publish your research.

The book's greatest strength might also be its weakness. The spirit of the book as outlined by the editors charged the authors to write about the nitty-gritty details of their challenges on specific projects. Inarguably, the purpose of the book is well achieved. It's comforting, especially for junior scholars, to read about other researchers' struggles and know they are not alone. Reading about the design process also sparks the reader's imagination. The authors not only offer inspiring accounts of finding and overcoming design issues, but they also discuss new online tools and processes that could be used by readers for future projects.

However, the projects and their associated challenges in some cases seem overly specific, and the amount of detail provided makes some chapters read like the authors' cathartic journal entries. It's easy for a reader to get lost amid, in one example, highly specific details about the inner workings of a company that serves as an ethnographic field site, making it harder to see the overall lessons contained there for ethnography as a method. The focus on digital media also means many of the researchers' challenges were technical, so some of the chapters go into technical detail about topics like creating databases, tracking IP addresses, and building web crawlers. This is another place where it can be hard to see the forest for the trees for researchers who aren't working on these kinds of projects. In addition, despite the breadth of topics covered in the book, one audience that might feel left out is the experimentalists. Because the title of the volume mentions studying behavior online, it was surprising to find there were no experiment-focused chapters. Experimentalists face many challenges in the digital

media landscape, such as trying to replicate the social media environment in a controlled experiment. It would have been nice to see this method represented in the volume.

However, there are recurring themes throughout the book that are useful for all scholars, regardless of their research approaches or interests. In many cases, these themes take the reader back to methods basics, even as the authors point out unique challenges created by the Internet. For example, several authors discuss the difficulties they had drawing a proper sample, especially when sampling online content rather than users. Some authors provide recommendations to the reader for how to draw better samples online when probability sampling is not an option, and others remind readers that in many cases researchers must limit their claims to the population studied. Similar recurring themes include reliability and validity of data, the importance of conceptualization, ethical decision-making, and navigating the IRB.

Another significant theme is the importance of theory. Although some methods books completely ignore theory, the editors make it clear from the introduction that this is not their approach, saying: "Methods can be usefully discussed by themselves and their theoretical frames deferred, as long as theory is not put off forever" (p. 14). Considering the debate among social scientists regarding the role of theory in the new world of big data (Anderson, 2008), it was refreshing to find that the chapters focused on big data in fact provided the strongest arguments in favor of theory. These authors contributed not only philosophical but also practical arguments in favor of developing research questions and hypotheses before data collection—otherwise, the amount of data is too overwhelming. This sentiment is summed up nicely by Shumate and Weber, who say: "While Web crawlers can collect vast quantities of textual and intertextual information, as a researcher you must determine what is significant before you begin" (p. 205).

It would have been nice if the editors had also provided a final, wrap-up chapter. Their introduction provided a nice setup for the chapters to follow and laid out the argument clearly, but a conclusion chapter would have provided additional continuity and allowed them to tie together the themes mentioned throughout the 10 chapters by contributors. It would also have given them the opportunity to reiterate their call to the scholarly community to be more honest and transparent when writing about method.

In summary, *Digital Research Confidential* is a well-written volume useful for scholars who are engaged in digital media research projects or who would like to read about classic method problems in a new context. The book could be used in a graduate-level research design course, but the projects are likely too complicated and specific for an undergraduate methods class. The volume would probably be most helpful for scholars undertaking the same kinds of research projects outlined in the book, to help them avoid the same pitfalls the authors endured. However, as the editors say, people appreciate hearing about others' research struggles, even when the projects aren't like their own. More than that, although the Internet gives new reason for scholars to be more transparent about their processes and view method as a creative act, this rallying cry is something all scholars, regardless of discipline, can get behind.

Reference

Anderson, C. (2008, June 30). The end of theory: Will the data deluge make the scientific method obsolete? *Edge: The third culture*. Retrieved from http://www.edge.org/3rd_culture/anderson08/anderson08_index.html