

Marina Micheli

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

(doi: 10.2383/85296)

Sociologica (ISSN 1971-8853)

Fascicolo 2, maggio-agosto 2016

Copyright © by Società editrice il Mulino, Bologna. Tutti i diritti sono riservati.

Per altre informazioni si veda <https://www.rivisteweb.it>

Licenza d'uso

L'articolo è messo a disposizione dell'utente in licenza per uso esclusivamente privato e personale, senza scopo di lucro e senza fini direttamente o indirettamente commerciali. Salvo quanto espressamente previsto dalla licenza d'uso Rivisteweb, è fatto divieto di riprodurre, trasmettere, distribuire o altrimenti utilizzare l'articolo, per qualsiasi scopo o fine. Tutti i diritti sono riservati.

Book Review

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

doi: 10.2383/85296

This collection of essays edited by Eszter Hargittai and Christian Sandvig remains true to its title’s promise to provide “confidential” insights into digital research. Browsing the pages of this volume feels like being allowed to look into a researcher’s “secret (methodological) diary.” As the authors of one chapter eloquently state, the essays are a “behind-the-scenes look at both the content and the practice of a challenging and chaotic interdisciplinary research adventure, with an emphasis on how [the researcher] solved or dealt with the various problems that came [his or her] way” [p. 175].

The book comprises eleven chapters: the first is a helpful introduction by the editors that does a great job of putting the various contributions in context, while the remaining ten chapters are accounts of research projects that include detailed first-person methodological analyses.

As the title suggests, this edited collection examines cutting-edge research both *with* and *about* digital media. Sandvig and Hargittai explain that the contributors conceptualize the Internet (and digital media in general) as both an “instrument” and an “object of study.” The great enthusiasm surrounding the potential of “big data” for social research should be tempered by the concrete case studies presented in this book. There is indeed an unprecedented amount of online data available to researchers. Digital media can also make exciting research tools. However, scholars should be aware of both the different possible approaches to conducting research with online data, and the associated limitations. This volume informs the reader about these methods and their challenges, while presenting creative ways of using big data for social research. Individual chapters deal with different types of big data such as: web site interfaces [Chapter 2]; *YouTube* videos [Chapter 3]; *Wikipedia* entries and photos uploaded on *Flickr* [Chapter 8]; hyperlinks [Chapter 9]; and friendship networks in virtual worlds [Chapter 10]. Two chapters even propose new methods for collecting and analyzing online data: Gilbert and Karahalios describe a web application developed ad hoc to collect large amounts of data for their sociological research on tie strengths [Chapter 6], while Shaw tells us about using crowdsourced data collection and classification to perform extensive content analysis on hundreds of websites [Chapter 7].

The book, however, is not about big data *per se*, but about research *with* and *about* digital media. Therefore, the chapters illustrate the adoption of different digital research methods reflecting different research aims. These methods may be pioneering, but are not necessarily so. On the one hand, the book includes chapters describing established ethnographic research – such as boyd’s study on teenagers’ engagement with social media and Leonardi’s organizational ethnography with engineers using visual digital data in a workplace setting. On the other hand, it outlines research projects that inevitably required innovative methods – such as Hect and Gergle’s study on the local dimension of user-generated content (that is to say, the extent to which users edit and upload content about

the area in which they live) or Shumate and Weber's analysis of hyperlinking amongst websites "as a way to map and analyze trends in groups and organizations" [p. 204].

The book also offers detailed descriptions of web tools for digital research such as the *Internet Archive Wayback Machine* [Chapter 2] and web crawlers [Chapter 9], which may be useful to those who are new to digital research.

Aside from the innovative approaches and topics presented in this volume, it is particularly valuable how the authors set their work in relation to traditional research methods. All of them, at some point, stress that their work stands in continuity with traditional scholarship on social research methodology. Throughout the book, the reader will frequently encounter key issues faced by every social researcher such as: sampling, framing a research question, data cleaning [Chapter 10], representativeness and generalizability [Chapter 3], reliability of data [Chapter 7], triangulation, confidentiality, access to an organization for conducting ethnographic research [Chapter 5], ethical research on human subjects [Chapter 11] and so on.

Interestingly, this edited collection helps the reader to grasp the multiple limitations both of the information available in online big datasets and digital research methods. Ironically, traditional socio-demographic variables are the most difficult to obtain. Therefore online analysis is insufficient to the pursuit of certain research aims, such as in boyd's ethnographic study of teenagers' engagement with social media. As the author herself states: "Teen life is increasingly intertwined with technology, but the traces that teens leave through technology are not rich enough to convey their practices. [...] [G]etting at the nuances of teen life in a networked era requires going back to foundational research methods" [p. 99-100]. However, as the editors emphasize in their introductory chapter: "a claimed continuity in method is often quite the opposite; it is a way to enlarge and refine the definition of traditional research methods while still receiving full credit for rigor and tradition" [p. 24].

A significant theme emerging from these accounts is that, even in the case of the most advanced techniques, technology is not powerful enough to function on its own, nor, of course, to substitute human work. On the contrary, many of the authors describe doing a lot of "manual" work (e.g., to devise and set up automatic tasks conducted by algorithms). In certain cases, these methods seem to be even more time-consuming than traditional ones. For example, to "crawl" a "seed sample of 76 websites" that "ended up with a network of 25,628 domains" [p. 216], took Matthew Weber five months. Furthermore, how big data are organized and archived is generally obscure: for instance, Molnár and Hsiao inform us that it is impossible to find out exactly how the *YouTube* search algorithm works, additionally several videos claim to be on a certain topic (through their tags) but instead are not. With regard to the second of these issues, for example, the researchers were obliged to intervene "manually" to filter out irrelevant videos from the dataset and code the remainders.

It is particularly noteworthy that the authors of these essays openly share their mistakes, also described as "failures" [p. 189], "embarrassing" and "costly" [p. 161]. This makes these chapters rather unique, given that it is almost impossible to find such frank and detailed descriptions of empirical research in journal articles. As Shumate states "the standard methods section does not permit explanation of many of the tacit choices that have to be made when cleaning Web crawling data" [p. 219-220].

Of course, the candid admission of mistakes is justified by the fact that these researchers embarked on projects without an established methodological tradition and have been among the first to do so. Another issue that can make this kind of research particularly challenging is the constant evolution of the platforms that scholars are studying or using to extract data. Even the most well-known platforms can be “unstable.” A few examples: *Wikipedia* removed the option to download the entire history of its encyclopedia for each language edition precisely when a research team was relying on this information to conduct its analysis [Chapter 8]; *YouTube* changed the configuration of its search engine so that it was no longer possible to search for videos uploaded within a custom time range just as a researcher was using this technique to obtain her sample [Chapter 3]; and *Twitter* implemented a new feature for creating lists of contacts while a scholar was in the process of observing tie strengths among its users [chapter 6].

Overall, the book is particularly suited to social science graduate students at the early stages of their research careers. Graduate students will probably be the most engaged by the personal tone of these chapters, in conjunction with the book’s in-depth examination of research methods and the authors’ self-reflexivity. Indeed, such is the level of self-reflection offered in these contributions, that it occasionally becomes a sort of confession: “[it] has forced me to recognize some of my shortcomings as a scholar” [p. 158], “If this PhD does not work out, I will leave and open a used children’s bookstore” [p. 223].

While this might confuse some readers, I feel certain that graduate students going through the difficulties of setting up their own research will appreciate it and find themselves relating to many of the reflections put forward in this book. Although, conceivably, no one student (nor scholar) will use all the methods described in these chapters, read together they may inspire the pursuit of unconventional and innovative empirical research.

Naturally this is a good read for anyone embarking in a research project *with* or *about* digital media and needing a detailed overview of innovative and emerging methods. However, it is particularly useful for those at the early stages of their work, such as the research design phase, because it may encourage them to try out new and unprecedented approaches. A shortcoming of this book is possibly that, while grounded in social research, it appears to be addressed to students and scholars working in interdisciplinary teams. Multiple areas of expertise are required to pursue some of the research methods described in this volume: many of the projects involved computer scientists, geographers, archivists, software engineers, or HCI researchers in addition to social scientists. An open mind and, in certain cases, a rudimentary knowledge or understanding of technical languages (programming, database or web markup languages) are prerequisites for adopting the methods illustrated in this edited collection. Some of the contributions, especially Chapters 5 and 9, delve very deeply into technical issues and the “nitty gritty” details of research work. This might discourage some readers. Overall, while the rationale for providing this level of detail is to save other scholars time, a broader overview of the kind of findings these studies obtained could have made this book even more meaningful and extended its potential readership. Nonetheless, for all those embarking on a “digital research project,” it may be very useful to have this “confidential” on their desk, partic-

Micheli

ularly for those inevitable moments in which they will find themselves struggling with the challenges of their innovative methods.

Marina Micheli
University of Zurich