

The Past Web: Exploring Web Archives

Edited by Daniel Gomes, Elena Demidova, Jane Winters, and Thomas Risse.
Switzerland: Springer Nature, 2021. 297 pp. eBook. \$119.00.
ISBN 978-3-030-63291-5.

In 2006, Julien Masanès wrote what some consider to be the first book on web preservation called *Web Archiving*.¹ During the subsequent sixteen years, the number of publications on web archiving increased to reflect the overwhelming volume, societal relevance, and technological complexity of the Web. These publications, all variations on acknowledging the transience of the Web, focused on a diverse range of subjects, such as digital archives,² the basics of web archiving,³ digital cultural memory,⁴ and web history.⁵ In the past five years, the combination of the COVID-19 pandemic and antiracist protests created web archivists out of many information professionals; thus, collecting and preserving web-based materials in both traditional and nontraditional institutions became an urgent priority, even though efforts to archive the Web have been underway for twenty-five years (p. vii). From these global events arose innovative and creative ways to preserve online content, and new technology and tools followed.

Published in 2021, *The Past Web: Exploring Web Archives* is a welcome addition to the collective *oeuvre* of web archives scholarship. Because the first sentence of the book acknowledges that “[m]any people do not know that the Web is being archived” (p. vii), one may wonder if the book is not relevant for those outside the GLAM fields. However, readers will quickly discover as they continue reading that the book is intended for Web users of any skill range. Thus, one of the book’s strengths is that it can be read as an introduction to practical information about web archives for academics in a broad range of research areas who are new to the topic. For professionals with advanced technical skills, the book presents unique projects involving web archives, raises new challenges and concerns, and explores recent research results about web archives initiatives focusing on copyright and privacy concerns, technology and tools, and the curatorial and creative uses of web archives.

The book is edited by digital preservation and web archiving scholars Daniel Gomes (Fundação para a Ciência e a Tecnologia Lisbon Portugal), Elena Demidova (University of Bonn), Jane Winters (University of London), and Thomas Risse (Goethe University Frankfurt) who recruit an impressive list of thirty-nine contributors from all over the world. Structured in six parts, each major section offers an assortment of chapters that call attention to various web archiving initiatives, as well as discuss tools and services currently available to assist users in exploring the archived Web. In addition, each part includes a short introduction that acts as an overview of the section’s theme, provides readers with a general summary of the

chapters within the section, and encourages enthusiasm about the forthcoming content.

Part 1 discusses web ephemera, the problems of transient web-data, and web archiving as representation of digital collective memory. In “The Problem of Web Ephemera,” Daniela Major delves into the problem of link rot (p. 7), which she carries over into “Web Archives Preserve Our Digital Collective Memory,” coauthored with Daniel Gomes. This chapter offers a brief history of web archiving that is helpful for readers who are new to the concept.

Part 2 focuses on the problem of deciding what to preserve and investigates the different strategies used to preserve web collections. Paul Koerbin, Ivy Huey Shin Lee, and Shereen Tay describe the web archiving efforts of the national archives of Australia and Singapore and approach these topics from the context of larger, comprehensive collections. The chapters focusing on archiving Twitter and event-centric web content call attention to collections that may require specified technology or tools. Elena Demidova and Thomas Risse’s chapter is fascinating because it delves into manual curation technology and crawl-based methods used to access and create event-centric web collections.

Part 3 introduces challenges to web archives access and offers constructive and efficacious methods to solve these problems through overcoming technological barriers. The technical lexicon in this chapter may be intimidating for beginning web archivists, but the value of the content is worth parsing through the jargon. Zeynep Pehlivan’s essay on “Linking Twitter Archives with Television Archives” stands out for exploring resourceful ways to use an amalgamation of televisual and Twitter archives (p. 138) to provide access to social media data.

Part 4 addresses the theories, methods, and practices of web archives as applied in case studies. Researchers in both the humanities and the social sciences have used the past Web for groundbreaking research; thus, these essays explore various qualitative and quantitative investigations that demonstrate the complexities and achievements of case studies. Saskia Huc-Hepher and Naomi Wells’s essay, “Exploring Online Diasporas: London’s French and Latin American Communities in the UK Web Archive,” is particularly interesting. The authors’ multilingual approach to web archiving guarantees the diversity of voices and communities “across digital and physical environments” (p. 189) and highlights how web archives can be applied linguistically and culturally.

Part 5 argues that web archives should move from a niche service to a critical infrastructure for modern societies. Miguel Won’s chapter, “Political Opinions on the Past Web,” offers an illuminating look into the Portuguese Arquivo de Opinião (Archive of Opinion), an online archives that collects opinion articles with the goal of studying political commentary via web archived articles (p. 243). Another interesting method of turning web archives into critical infrastructures is introduced in Dragan Espenschied and Ilya Kreymer’s chapter discussing the Oldweb.today

legacy browser. This service allows users to search the Web using obsolete browsers, “instantly and without any previous configuration necessary” (p. 256). In addition, to encourage more research of the past Web, Niels Brügger argues in “The Need for Research Infrastructures for the Study of Web Archives” for web archives to be viewed as research infrastructures.

Last, Part 6 reflects on the development of the Web, web archiving practices, changes to preservation efforts, and the long-term sustainability of web archives. Written by Julien Masanès, Daniela Major, and Daniel Gomes, this singular chapter offers some final thoughts on the “ethical duty to archive public content” (p. 291) and on ways to make this content accessible. The fragility and quick-paced evolution of the Web presents myriad difficulties in collecting and preserving everything, but the threat of irrelevance and obsolescence is too great of a jeopardy in this technological age.

Overall, *The Past Web* is eloquent and demonstrates innovative concepts in web archiving theory and practice. The essays cover a lot of ground in around ten pages each, and their short length may leave the reader unfulfilled and wanting more—perhaps a good excuse for the editors to compile more essays for another volume. The acknowledgments section is quite meta; because this work is available in e-book form, the editors worked to preserve all the links in the electronic version of the book. Consequently, any links that happen to be broken can be recovered from the Arquivo.pt Web Archive. This is a nice salutary nod to the problem of link rot discussed in some chapters.

One critique is the amount of technical lexicon found throughout the book. The specific terminology, complex web crawling concepts, and programming-specific language may be overwhelming for readers who do not have computer programming or coding experience. In the preface, the editors state that the book aims not only to promote web archiving as accessible to all but also to provide background information for newcomers (p. xii). The editors admit that readers’ levels of technical knowledge will vary and that part of their motivation in writing the book was to help those interested in web archiving to get started (p. xii). Another motivation was to offer educators a source to use in teaching web archiving to students. Thus, with these purposes in mind, the editors compiled a book not only for web archivists but also for all web users interested in the past (p. xii). While these ideas make sense separately, they conflict when juxtaposed. An excess of complicated or obscure language creates potential barriers to comprehension. Accessibility for all means using language for all, and this book does not always live up to that ideal.

Thankfully, criticisms of the book are few, and the diversity and ingenuity of topics found in the chapters make the work a welcome addition to the literature of the field; the reference sections alone offer a comprehensive bibliography of web preservation. The book’s significance within the scholarship of the web archiving field is evident, and I cannot think of another work that offers as distinct an array

of creative web archiving initiatives. This work not only describes the theoretical purpose and methods of web archiving; rather, this book shows readers what can be done with web archiving and how various projects and case studies can work toward preserving digital content in a holistic sense.

© Amanda Greenwood
Schaffer Library, Union College

NOTES

- ¹ In 2005, Ross Harvey published *Preserving Digital Materials*, which references web archives and archiving web content, but the book by Masanès is thought to be the first comprehensive one published about web archiving. Ross Harvey, *Preserving Digital Materials* (München: K.G.Saur, 2005); Julien Masanès, *Web Archiving* (Switzerland: Springer Nature, 2006).
- ² William E. Landis and Robin L. Chandler, eds., *Archives and the Digital Library* (London: Routledge, 2008).
- ³ Niels Brügger, *Archiving Websites: General Considerations and Strategies* (Århus: Centre for Internet Research, 2005); Adrian Brown, *Archiving Websites: A Practical Guide for Information Management Professionals* (London: Facet, 2006).
- ⁴ Abigail De Kosnik, *Rogue Archives: Digital Cultural Memory and Media Fandom* (Cambridge, MA: MIT Press, 2016).
- ⁵ Niels Brügger and Ralph Schroeder, eds., *The Web as History: Using Web Archives to Understand the Past and the Present* (London: UCL Press, 2017); Niels Brügger, *The Archived Web: Doing History in the Digital Age* (Cambridge, MA: MIT Press, 2018); Ian Milligan and Niels Brügger, eds., *The SAGE Handbook of Web History* (London: SAGE, 2018); Niels Brügger and Ditte Larson, eds., *The Historical Web and Digital Humanities: The Case of National Web Domains* (New York: Routledge, 2019).