

Appendix B: Ebook Bibliography

Kindle files are compiled at point of download on Amazon's servers and cannot be treated as consistent beyond individual files. Ebook bibliography requires a new vocabulary to distinguish between unique editions and account for the variability between dedicated e-readers and their simulation via Kindle for PC. Academic style guides have been slow to react to the changing demands of ebooks. The eighth edition of the *MLA Handbook*, published in 2016, largely ignores ebook platforms within a broad conceptual framework of media in the early twenty-first century. The revised *MLA Handbook* introduced the media metaphor of "containers" and "content" to its citation system, where a citation can simply note that the version consulted was an "e-book" and specify the copy as a "Kindle ed." As I demonstrated in chapter 4, this approach cannot provide sufficient detail of the item in a cloud-generated format specification. The in-text reference section mentions only citing a "part number" if it is "explicit (visible in the document) and fixed (the same for all users of the document)" but does not offer an example of this in principle.¹ The seventeenth edition of *The Chicago Manual of Style*, published in 2017, offers a more nuanced breakdown of ebook citations, although the information is contradictory, wavering between adding the broadest possible title of a platform (such as "Kindle edition") to the end of an otherwise regular reference and including location information.² As variations exist within platforms, let alone between them, scholars need a more robust framework for identifying and citing ebooks as unique digital objects.

In the appendix to *I AM ERROR*, Nathan Altice demonstrates the need for closer attention to medium-specific bibliography for digital objects, noting that “platform studies owes much to the discipline of bibliography.”³ Altice further argues that even though the fidelity of these objects and surrogates, such as cartridges and their ROMs, may appear to be identical, “we know that they often carry along unseen textual artifacts of their history, circulation, and distribution. None of these aspects should be ignored when we take bibliographic account of our digital objects.”⁴ The work of platform studies helps to illuminate the material conditions of creation and circulation, but this needs to be supplemented by traditional bibliographical rigor to encourage appropriate documentation in nonspecialist works.

Digital bibliography often focuses on the exceptional—video games,⁵ electronic literature,⁶ digital-physical hybrids⁷—but mundane digital objects such as web pages, facsimiles, Instagram feeds, and Word documents have not received the same level of attention. Ebooks appear familiar on the surface, drawing on the rich bibliographical and logistical traditions of print, but despite this supposed dull exterior, the objects present new challenges for bibliography, as evidenced throughout *Four Shades*. It is possible to see the EPUB source files created by a publisher or self-publishing author and compare them to a local copy of the final publication, but Amazon’s black-box cloud processing and delivery of ebook content ensure that the Kindle format remains a distant object. The descriptive bibliography of Kindle ebooks requires an understanding of cloud computing bibliography and the book-as-continuous-object. I offer some suggestions for referencing ebooks in this appendix, but conventions and standards evolve on the basis of further knowledge and developments.

Mechanisms remains a foundational text in both platform studies and digital bibliography due to Kirschenbaum’s introduction of digital forensic analysis into both media and literary studies. Kirschenbaum provided case studies for a range of static digital objects including Michael Joyce’s *afternoon* and Roberta and Ken Williams’s *Mystery House*, which provide the foundations for understanding less-fixed artifacts.⁸ The dynamic generation of ebooks ensures that every copy is unique. The typical Kindle ebook does not have sufficiently complete and verified metadata on Amazon product pages, prompting the need for further forensic analysis. Amazon maintains control over publicly available metadata on product pages, but this information can be unreliable. For example, the product page for Christopher Rice’s *Bone Music* features a file size in the “product details” of 1,960 KB, but when the book is downloaded, the size ranged from 786 KB to 812 KB depending on the file format.⁹ An opened copy of the title on

Kindle for Mac OS 1.20.2 clocked in at 1,772 KB with ancillary documents. Similarly, the product page does not list the “location” size but references the “Print Length” accompanied by “Page Numbers Source ISBN,” a useful measure for comparison, but not accurate enough to ensure that page numbers transfer between media.

As with forensic evidence, it is also important to note the difference between tampered and untampered files. An unopened Kindle file is substantially different from a loaded ebook, since the ancillary files are only downloaded once a user begins to read. This creates a significant difference for the user experience beyond any bibliographical variation. Amazon facilitates the collection of “new” ebooks through the “Download and transfer via USB” option when managing Kindle content associated with a device. It is possible to collate a full bibliographical record from this file alone other than the location length, which requires an open app. Since metadata around ancillary files and the location length are necessary to create accurate bibliography records, one should assume that files have been opened in generating bibliographies.

From Collation to Location

Ebook bibliography has more in common with the pre-nineteenth-century print book than with mass-produced twenty-first-century books. Early modern books featured unreliable pagination with a mixture of unnumbered pages and signatures that might mix letters, roman numerals, and other elements primarily intended to aid the book binder in ensuring correct order. Pagination for Kindle titles is equally unreliable, and page numbers are algorithmically generated according to nonstandard rules. The book trade typically identifies the title page as the first page with Arabic numerals, ensuring the first page of the main body rarely begins with page 1. Conversely, the Kindle’s algorithmic generation assumes the main body marks page 1, ensuring any additional accuracy is at least a couple of pages out of synchronization.

The early modern bibliographical practice suggested that books should be handled according to each unique copy. The formula includes the binding of the book (folio, quarto, octavo) followed by a method of identifying an edition. In *Principles of Descriptive Bibliography*, Fredson Bowers provides an example issue formula of “4^o: A²B-C⁴D²” to represent “a quarto [consisting] of the following gatherings: 2 leaves signed A, 4 leaves signed B, 4 leaves signed C, and 2 leaves signed D.”¹⁰ The alphabetical signatures were used to allow binders to collate the book effectively rather than to aid navigation, but this was a useful tool for checking if two versions of the

same book were printed in the same run. Just as early modern bibliographers adapted to the material conditions of letterpress printing, ebook bibliography must develop conventions suitable for formats rather than algorithmic page numbers. The solution is platform dependent, as iBooks, for example, removes all traces of static pages, rendering in-text citations impossible. For the Kindle, we have two prebuilt solutions. First we can use the underlying byte register location used in the internal files for marking the position of highlights, index entries, and reading locations. The byte-level location is difficult to access and requires use of ancillary files such as an ebook's PHL. Second, the location information is available for all titles, even if it is obfuscated by an attempt to render "real page numbers," and is the most appropriate in-text citation mechanism.

Kindle Bibliography

The following principles assume that it is sufficient to identify a publication at the higher level rather than individual copies of digital objects. All extra details should be available to compile from various sources without recourse to subverting DRM.

Format

Creator(s) (Release date). *Title* [Hardware and firmware or Software and version].
Format. Catalog ID. Publisher. Version. Location.

The first three elements—creators, date, and title—follow standard bibliographical conventions and can be adapted according to the referencing template of a specific editing style.

The middle section, comprising the software through the catalog ID, represents the most valuable medium-specific information for identifying differences in the same ebook. The platform is usually noted after the title, but this should be supplemented by further information about the method of access. The required information depends on whether the text has been accessed via a dedicated ebook reader or through software. E-readers should feature information about the device generation and firmware; software should indicate the software and version. For the Kindle Cloud Reader, version and build date appear in the website's source code.

The format should be the default when the file is downloaded. The same file may be rendered differently according to its format on the same hardware-software combination. This can be obfuscated through reuse of the same file extension despite rendering different content. The main cutoff occurs with the launch of KF8 and KFX, where newer devices download content in this format. Older AZW files can also be opened

with a PRC extension. TPZ most often refers to Apple desktop and mobile implementations.

The catalog ID for Kindle titles is the ASIN, which can be found through the file name of the ebook or through the URL of the Kindle title's product page. This is best recorded at time of download, as new editions of the same text can occasionally generate new ASINs for the same ebook. An ASIN is a persistent identifier even after an ebook has been removed from the Kindle Store, and therefore is a useful tool for distinguishing between editions of a text when not clearly marked, as has happened with Amazon-issued public domain texts.

Information about the publisher should be restricted to name, as the location of publication is difficult to ascertain and irrelevant in the case of Kindle Direct Publishing titles. Where a book has been published via Kindle Direct Publishing without an imprint, the "sold by" field should be included. For example, in the British market, this is typically "Amazon Media EU S.à r.l." Occasionally the publisher will include version information, which should be included, although this might not be updated with successive updates. The copyright page of Ben Goldacre's *Bad Science* not only distinguishes between print and ebook ("Ebook Edition © December 2008 ISBN: 97800007283194") but continues to note that the current ebook is "Version: 2013-09-03."¹¹ The versioning depends on good publisher workflows and will not apply to updates from Amazon, but once again, this offers a rough guide of the book's provenance.

The location should be the total location count for the entire book in lieu of a page count. This information can be important for identifying minute differences between editions, as the unit of a location is lesser than the equivalent page.

Examples

- Albanese, Andrew Richard (2013). *The Battle of \$9.99: How Apple, Amazon, and the "Big Six" Publishers Changed the E-book Business Overnight* [Kindle Cloud Reader 011201999]. B00DH8JCOC. PWxyz. Loc. 979.
- Goldacre, Ben (2008). *Bad Science* [Kindle for Android 8.0.0.68]. AZW3. B002RI-9ORI. HarperCollins. Version 2013-09-03. Loc. 5474.
- Isaacson, Walter (2011). *Steve Jobs: The Exclusive Biography* [Kindle 8, 5.9.4]. KFX. B005J3IEZQ. Little, Brown. Loc. 12321.
- Pitzer, Andrea (2017). *One Long Night: A Global History of Concentration Camps* [Kindle 2, 2.5.8]. AZW. B01N5XNWEF. Loc. 7442.
- Sterne, Laurence (2012). *The Life and Opinions of Tristram Shandy* [Kindle for Mac 1.20.2]. AZW. B008NZGMV0. HarperCollins. Loc. 9863.
- Sterne, Laurence (2012). *The Life and Opinions of Tristram Shandy* [Kindle for PC 1.21]. KFX. B0082RZPVG. Amazon Media EU S.à r.l. Loc. 7773.

Copy Specific

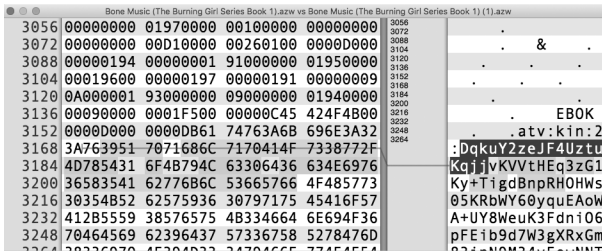
If the citation should identify a specific copy for provenance purposes, the EXTH headers provide additional copy-specific data. For example, figure 12.1 shows two copies of the same file downloaded within a minute of each other. As these material differences occur between different copies of the same file, copy-specific reference may also rely on checksums such as MD5 to demonstrate the unique aspects of a copy.¹² In the example shown in the figure, an MD5 calculation performed through the Mac OS command line demonstrated material differences:

```
1: ad93a79acfb043ed2e7dfc033ea92627
2: a89114eddb084d60c00f9c8c66a7doc
```

This level of copy specificity is not necessary in most cases and has not been applied to the bibliographical specification of digital objects in this book. If such specificity is necessary for analysis, I suggest referring to the MD5 hash for the title.

I have documented best practice for Kindle-based bibliographies based on the sociotechnical conditions of Amazon’s platform. Some of the principles here will not transfer to other ebook platforms, while others can easily be adapted. For example, iBooks stores relevant metadata in “Books.plist,” a file contained in the same folder as the raw EPUB files. The document lists important information including the unique iBooks “itemID” and “publication-version.” From these metadata, it is possible to compile a comparable reference for J. J. Abrams and Doug Dorst’s *S.* for iBooks:

J. J. Abrams and Doug Dorst (2013). *S.* [iBooks 1.5 (Mac OS X 10.11.6)]. EPUB. 726968790. Canongate. Version 3612477.



12.1 Ebook signatures for two otherwise identical copies of Christopher Rice’s *Bone Music* [Kindle 3, 3.4.2], AZW, B07354S1K7 (Thomas & Mercer, 2018), loc. 6395. Detail from Hex Fiend. Screenshot by the author.

I have purposefully omitted the page number information, despite internal consistency with *S.*, since the text is presented in a fixed layout, because by design iBooks pages are dynamic. This is a common problem for EPUB, as the format has no normative specification for in-text locations beyond attempts by the IDPF to create EPUB Canonical Fragment Identifiers (CFI), which remained recommended in the EPUB 3.1 specifications. The specification provides an example of CFI implementation:

```
book.epub#epubcfi(/6/4[chap01ref]/4[body01]/10[para05]/3:10)13
```

The system used XPath, an XML-based navigation syntax, to identify a precise location based on the “spine” of an EPUB file. The technique would be more difficult to implement in encrypted Kindle files and can only identify a passage for a machine. Therefore, while the specification offers an additional level of technical specificity, the trade-off between machine- and human-legible forms of citation is important in identifying citations for the broadest possible audience. While granular citation remains an open issue, I recommend using locations for Kindle ebooks and the proprietary standards for other platforms with direct quotations where possible to ensure text can be identified.

This is a section of [doi:10.7551/mitpress/11985.001.0001](https://doi.org/10.7551/mitpress/11985.001.0001)

Four Shades of Gray

The Amazon Kindle Platform

By: Simon Peter Rowberry

Citation:

Four Shades of Gray: The Amazon Kindle Platform

By: Simon Peter Rowberry

DOI: 10.7551/mitpress/11985.001.0001

ISBN (electronic): 9780262369114

Publisher: The MIT Press

Published: 2022

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

© 2022 Simon Peter Rowberry

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Filosofia OT by Jen Jackowitz. Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Names: Rowberry, Simon Peter, author.

Title: Four shades of gray : the Amazon kindle platform / Simon Peter Rowberry.

Description: Cambridge, Massachusetts : The MIT Press, [2022] | Series:

Platform studies | Includes bibliographical references and index.

Identifiers: LCCN 2021013279 | ISBN 9780262543507 (paperback)

Subjects: LCSH: Kindle (Electronic book reader) | Electronic book readers.

| Electronic books.

Classification: LCC Z286.E43 R689 2022 | DDC 004.1675—dc23

LC record available at <https://lcn.loc.gov/2021013279>

10 9 8 7 6 5 4 3 2 1