

Canaries in a Coalmine: The Index and the Page in Ebooks

MARY COE

The invention of the page is a major milestone in the history of book technology. Consider the transition from the scroll to the codex (the printed book as we know it today) to the electronic book (ebook). Scrolls could be as long as ten metres and were continuous script without page breaks or page numbers.¹ The concept of pagination was introduced with the codex, which is the Latin word for “a plurality of tablets or for multi-leaved tablets,”² and enabled a new reading experience that allowed people to jump into the text at different points and to use a book in a nonlinear way.³ The first paginated books gradually replaced the scroll from as early as the second century.⁴ As page numbers came into common use in the early sixteenth century, book indexes also developed.⁵ However, the development of the ebook is leading us back to the beginning, to the use of a scroll-like format in ebooks without pages. The development of the index is linked to this new phase in book technology.

Throughout this paper, the term *index* refers to an analytic index, which is created by human beings (preferably professional indexers) for the purpose of helping readers find information in the book. Indexes are usually found in non-fiction books because people often use these types of books in non-linear ways; in other words, they might not read from the first page directly through to the last page but rather dip into the book at different places looking for information. Indexes can be a useful tool in an ebook, particularly when combined with other finding aids, such as search boxes or tables of contents. Analytic indexes are different from derived indexes, which might be computer-generated using only terms in the text. Generally, analytic indexes are a more sophisticated tool than a derived index or a full-text search box, which both might be helpful to readers who already know what terms to search for or who are looking for a simple fact or a name. On the other hand, analytic indexes can be particularly helpful for readers who do not already know what is in the text; for example, an index might contain synonyms for terms, might describe concepts that are not readily pulled out of the text with single words or phrases, or might gather information in one place using subheadings.

¹ Martyn Lyons, *A History of Reading and Writing in the Western World* (Palgrave Macmillan, 2010).

² Bill Katz, *Dahl's History of the Book* (Metuchen, NJ: Scarecrow Press, 1995), 69.

³ Lev Grossman, “From Scroll to Screen,” *New York Times*, September 2, 2011.

⁴ Katz, *Dahl's History of the Book*.

⁵ David J. Shaw, “The Book Trade Comes of Age: The Sixteenth Century,” in *A Companion to the History of the Book*, ed. Simon Eliot and Jonathan Rose (Wiley-Blackwell, 2009).

Ebook design is still in its infancy.⁶ Arenas and Barr suggest that current ebook development is similar to the shift from the carriage to the automobile, when people continued to think of cars as a carriage without a horse.⁷ Some ebooks are produced in fixed formats, such as Portable Document Format (PDF), which maintain the concept of a page and retain the same layout regardless of which device they are read on. The use of the codex metaphor in this way provides a comfortable way for readers to approach the new technology. However, other ebooks are created in reflowable formats that are equivalent to a scroll, such as the EPUB format. EPUB 3 is a specification for ebooks that was developed in 2011 by the International Digital Publishing Forum (IDPF) EPUB Working Group, which consisted of IDPF member representatives and Invited Experts, including professional indexers.⁸ Unlike fixed-format ebooks, EPUB ebooks reformat to fit the size of the device and do not rely on the concept of a page.⁹ However, the mental models that readers use in their approach to ebooks can be problematic in the context of these new formats.¹⁰ MacFadyen suggests that because ebook devices do not yet enable readers to focus on the affordances unique to the electronic format, there is a “discontinuity that characterizes digital reading practices.”¹¹ The turning point may come when we start thinking about books for what they can *do* rather than for what they *are*.¹²

⁶ As noted by Geoffrey Little in “The Book Is Dead, Long Live the Book!” *The Journal of Academic Librarianship* 37 (2011): 536–38 and Simon Peter Rowberry in “Ebookness,” *Convergence: The International Journal of Research into New Media Technologies* (2015): 1–18.

⁷ Edilson Arenas and Avron Barr, “The Digital Book in Higher Education: Beyond the Horseless Carriage,” paper presented at the *Electric Dreams: 30th ascilite Conference Proceedings* (Macquarie University, Sydney, 2013).

⁸ Michele Combs and David K. Ream, “Understanding the IDPF EPUB3 Indexes Specification,” *The Indexer* 32 (2014). See also the International Digital Publishing Forum, “Epub 3.0” at <http://idpf.org/epub/30>.

⁹ Glenda Browne and Mary Coe, “Ebook Indexing,” in *Index It Right! Advice from the Experts*, ed. Enid L. Zafran (Medford, NJ: Information Today, Inc., 2014), 29–45.

¹⁰ These mental models have been explored by the following researchers: Catherine Marshall, *Reading and Writing the Electronic Book*, Synthesis Lectures on Information Concepts, Retrieval, and Services, ed. Gary Marchionini (Morgan & Claypool Publishers, 2010); Ellen Rose, “The Phenomenology of on-Screen Reading: University Students’ Lived Experience of Digitised Text,” *British Journal of Educational Technology* 42 (2011): 515–26; Gemma Walsh, “Screen and Paper Reading Research - a Literature Review,” *Australian Academic & Research Libraries* 47 (2016): 160–73.

¹¹ Heather MacFadyen, “The Reader’s Devices: The Affordances of Ebook Readers,” *Dalhousie Journal of Interdisciplinary Management* 7 (2011): 9.

¹² Johanna Drucker, “The Virtual Codex from Page Space to E-Space,” in *A Companion to Digital Literary Studies*, ed. Susan Schreibman and Ray Siemens (Wiley, 2013).

Active ebook indexes

As the technology of the book continues to evolve, indexes will develop along with it. Figure 1 displays an index that was recently published in an EPUB-formatted ebook. This index contains locators that are hyperlinked back into the text. This is known as an ‘active index’, which allows the reader to move from the locator directly to a point in the text (as opposed to a ‘passive index’ without hyperlinks, which is what is usually included in a print book index). While active indexes are becoming more common, the use of the locator ‘*here*’ in this index is unusual and could indicate that this ebook does not have pages as we know them.¹³

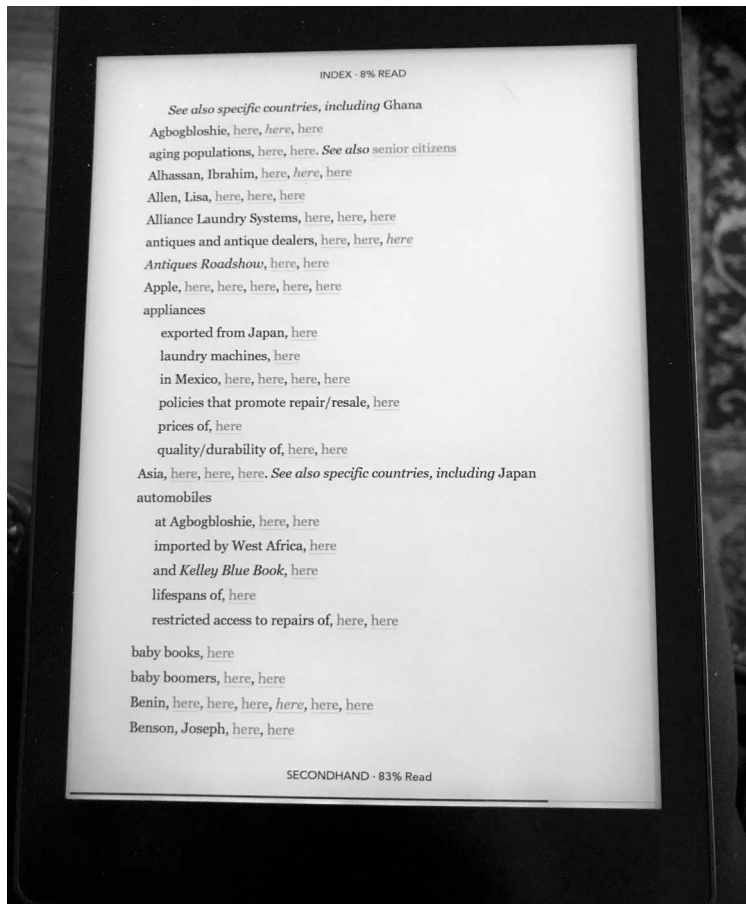


Figure 1. Secondhand: Travels in the New Global Garage Sale by Adam Minter. Published 12 November 2019 by Bloomsbury Publishing (EPUB 3 format presented in a Kobo ereader)

¹³ Please note that the word ‘*here*’ in italics in the index in Figure 1 is used to denote figures in the book. I am using italics in the text of this article to bring your attention to the word ‘*here*’ used as a locator.

There is no reason why ebook indexes should not be included in this reflowable format, but we do need to explore better ways of presenting these indexes and using the technology to its full advantage. For example, how do readers know where the word *here* will lead them or where they are in the text when they complete the jump from the hyperlink? And how do readers distinguish between the locators in a long string, i.e., ‘*here, here, here*’? Based on experience with indexes in print books, readers might assume that the references listed as ‘*here, here, here*’ are arranged in order of appearance in the text, as they would be with page numbers in a traditional print book. But how might we create indexes in reflowable ebooks that offer different functionality? For example, suppose that the indexer arranged the locators in order of perceived relevance, i.e., the most thorough discussion of the topic indexed appears first. Page numbers might not be the most logical choice for creating index locators in this type of arrangement. Is the word ‘*here*’ the best option? What might we use instead?

Book indexers such as myself have been pondering this question and exploring new ways of creating and presenting active indexes in ebooks.¹⁴ In 2012, I considered whether the reign of the page would be coming to an end and what it would mean for book indexes.¹⁵ This is a question that I am still pondering and, most recently, exploring in a current research project entitled ‘a sensemaking approach to active ebook indexes.’

A Sensemaking Approach to Active Ebook Indexes

Research methods

I am inviting university students and staff into a digital library usability lab and asking them to use indexes in ebooks with reflowable formats. With the assistance of Jan Wright, an award-winning indexer and member of the IDPF EPUB Working Group, I selected two ebooks from the Project Gutenberg¹⁶ collection, reformatted them in EPUB, and added active indexes hyperlinked into the text at various levels (to a page, to the beginning of a paragraph, or to a line of text). The books are on general topics that might be of interest to

¹⁴ Glenda Browne and Jan Wright have been particularly active in this area. See the following: Glenda Browne and Mary Coe, “Ebook Navigation: Browse, Search and Index,” *The Australian Library Journal* 61 (2012): 288–97; Glenda Browne and Mary Coe, “Ebook Indexing,” in *Index It Right! Advice from the Experts*, ed. Enid L. Zafran (Medford, NJ: Information Today, Inc., 2014), 29–45; Glenda Browne and Jan Wright, “Epub3 Indexes and the Future of Indexing,” *The Indexer* 31 (2013): 110–21.

¹⁵ Mary Coe, “The Tyranny of the Page,” *The Indexer* 30 (2012): 2–5.

¹⁶ Project Gutenberg is a library of over 60,000 free books that have been digitized and proofread by volunteers. The focus is on older books for which copyright has expired. The project can be found at <https://www.gutenberg.org/>.

students and staff across all university departments. Tasks were then devised with the aim of engaging participants in using the indexes. For example, the first ebook that participants are presented with is a biography of Charles Sturt, a British explorer of Australia and the namesake of the university.¹⁷ The first task they are given is to find out what Wagga Wagga (the city where the university campus is located) looked like when Charles Sturt explored the area.

As participants are using the ebooks and indexes, their interactions are recorded by eye tracking equipment. The eye tracking device observes the path of participants' eyes and how long they look at something. The device also records participants' movements within the ebook, such as between index entries and the text or when using finding aids such as the search box. The eye tracking equipment is situated in front of a normal computer monitor and is unobtrusive. In most cases, it is a simple procedure to set up and calibrate the device, and research participants do not find it uncomfortable or distracting. Once participants are settled with the ebooks and the eye tracking equipment, the recording session is started, and I leave the room to allow them to work through the tasks uninterrupted. Consequently, participants are able to work with the ebooks and indexes as naturally as possible without a researcher looking over their shoulder.

After the participant finishes with the tasks for a book, I return to the room, sit down with her,¹⁸ and ask her to talk me through what she was doing and thinking as I replay the eye tracking recording. I ask many questions in an interview style in an effort to dig into not just what she did in that immediate moment but what influenced her actions—for example, how she uses print book indexes (if she does), how she uses books for studying, and/or how she uses hyperlinks in the online environment.¹⁹ It is these other influences that are part of the process of sensemaking.

Sensemaking

Sensemaking has been defined as, “the process of constructing an understanding when facing complex situations.”²⁰ The word ‘constructing’ is important here, as information seeking in the context of analytic index use is often a process of construction. In other words, information is not always something that readers

¹⁷ *Charles Sturt – His Life and Journeys of Exploration* by J. H. L. Cumpston (published 1951). Available from Project Gutenberg Australia at <http://gutenberg.net.au/ebooks07/0700391h.html>.

¹⁸ I am using the feminine pronoun for convenience. Recruitment of participants for this study will include both men and women. Participants will be asked if they identify with a particular gender (or not) but this information will be used only to determine whether a suitably diverse sample has been included.

¹⁹ Use of hyperlinks online does have a connection to use of book indexes—consider that we call the parts of websites, *webpages*.

²⁰ Kholod Alsufiani and Simon Attfield, “An Exploratory Study of Computer Supported Sensemaking: Relating Representational Structure to Triage,” *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32* (2018): 1–5.

simply find in a book, but rather they may need to read and understand the index and the text in order to construct the information that they are seeking. Marchionini considers sense making to be “a component of search [that] often serves as a bridge between search and learning.”²¹

Brenda Dervin’s sensemaking methodology provides the framework for my research. Dervin describes this methodology as “a theoretic net, a set of assumptions and propositions, and a set of methods which have been developed to study the making of sense that people do in their everyday experiences.”²² The core assumption in Dervin’s sensemaking approach is the concept of the gap or discontinuity that people face when trying to find information. According to Dervin, “a person in a moment defines the moment as a particular kind of gap, constructs a particular strategy for facing the moment, and implements that strategy with a particular tactic.”²³ In the context of ebook index use, the first gap in knowledge faced by the user might be how to interact with the book as an information system itself. The index user must first make sense of how the hyperlinks between index and text work [define the gap] and whether they lead to text that is relevant to their needs [bridge the gap] before they can use the book to fill other gaps in knowledge. Figure 2 shows a preliminary adaptation of Dervin’s gap metaphor for active ebook index use. It illustrates the paths that readers might take when trying to construct the information that they are seeking in the book. These approaches might include using what they already know about book indexes or what they know about the digital environment. Some readers may not cross the gap—they may fall into it or stop trying to cross it—but investigating what blocks readers from using active indexes is equally important.

Preliminary findings

I have conducted two pilot studies in the usability lab. The first pilot confirmed that the ebooks and tasks designed for the project were appropriate and that the eye tracking equipment could be used as intended. The second pilot focused on the interview techniques that will be used when discussing the eye tracking recordings with participants. Preliminary analysis of findings from the second pilot study have also confirmed that two rich sources of data about active ebook index use will be gathered: eye tracking recordings and interviews. Full findings will not be reported here, but I can provide a few snippets from interviews that

²¹ Gary Marchionini, “Search, Sense Making and Learning: Closing Gaps,” *Information and Learning Sciences* 120 (2019): 74–86.

²² Brenda Dervin, “From the Mind’s Eye of the User: The Sense-Making Qualitative-Quantitative Methodology,” in *Qualitative Research in Information Management*, ed. Jack D. Glazier and Ronald R. Powell (Englewood, CO: Libraries Unlimited, 1992), 61–84.

²³ Dervin, “From the Mind’s Eye of the User: The Sense-Making Qualitative-Quantitative Methodology,” 82.

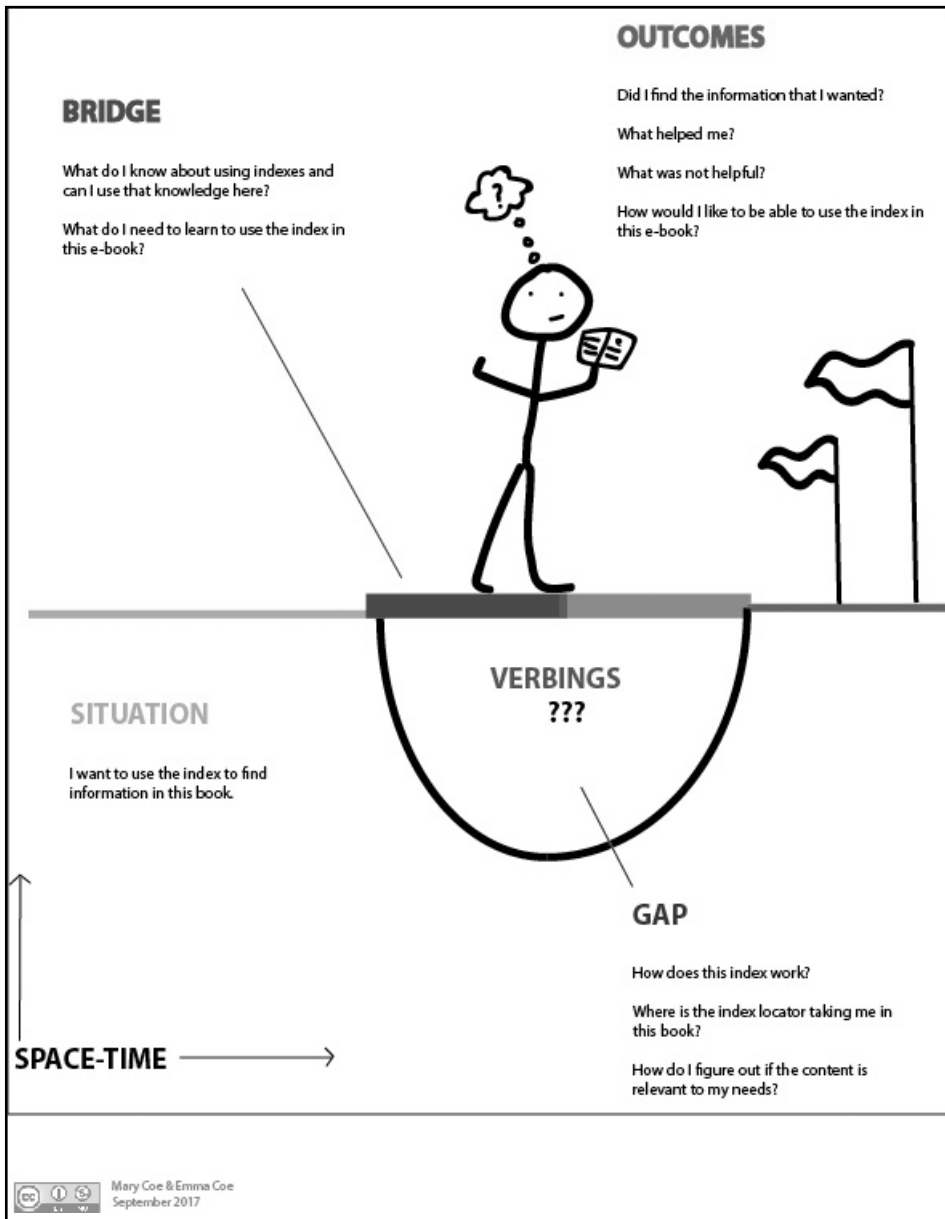


Figure 2. Dervin's sensemaking metaphor adapted to use of active ebook indexes.

will illustrate what the results of this project may reveal in the future.

One participant commented,

I was thinking how far will this go down because I can't see page numbers so I scrolled down further. So, from here, at the moment, I couldn't remember what

the first page number was or what the second page number is. I don't know where in the text I am!

She was not the only one to find herself lost in the text. Her reliance on page numbers—basically her mental model of what a book should be—became a problem for her in this ebook context. While in a print book, she might stop at the bottom of the page and perhaps scroll back up with her eyes to look for the information, in this case she did not know where to stop.

Another participant said,

Well you can see that I'm just up and down and thinking 'where is it?' Just jump out at me please! And I guess that maybe I'm expecting it to do that. I don't know why. Maybe because it's an ebook and I'm looking for a link maybe? I'm expecting it to be easier? I don't know.

In this case, it was the expectation of what would happen in an ebook that might have been an issue. This would be consistent with the prediction corollary, which considers the search process as “a series of choices based on predictions of what will happen if a particular action is taken.”²⁴ This participant and others also searched very specifically for keywords, particularly when they also used the search box in the ebook, hence that comment “Just jump out at me please!” They liked to see keywords highlighted in the text when they used the search box and seemed to be looking for that when using the index also. This is consistent with other studies of visual foraging in highlighted text.²⁵ These practices may also reflect changes in reading behaviour; for example, people are skimming and scanning more as they read on screen and reading comprehension can vary depending on whether they are reading print or on screen.²⁶ The capacity to determine relevance while scanning the text could also affect a reader's ability to construct information from it.²⁷

As Lopatovska and Sessions point out, there have been few efforts to integrate reading explicitly into an information-seeking model.²⁸ While a study of ebook

²⁴ Carol C. Kuhlthau, “A Principle of Uncertainty for Information Seeking,” *Journal of Documentation* 49 (1993): 339–55.

²⁵ Ed H. Chi, Lichan Hong, Julie Heiser, Stuart K. Card, and Michelle Gumbrecht, “Scentindex and Scenthighlights: Productive Reading Techniques for Conceptually Reorganizing Subject Indexes and Highlighting Passages,” *Information Visualization* 6 (2007): 32–47.

²⁶ Luanne Freund, Rick Kopak, and Heather O'Brien, “The Effects of Textual Environment on Reading Comprehension: Implications for Searching as Learning,” *Journal of Information Science* 42 (2016): 79–93; Yiren Kong, Young Sik Seo, and Ling Zhai, “Comparison of Reading Performance on Screen and on Paper: A Meta-Analysis,” *Computers & Education* 123 (2018): 138–49.

²⁷ Gary Marchionini, “Information-Seeking Strategies of Novices Using a Full-Text Electronic Encyclopedia,” *Journal of the American Society for Information Science* 40 (1989): 54–66.

²⁸ Irene Lopatovska and Deanna Sessions, “Understanding Academic Reading in the Context of Information-Seeking,” *Library Review* 65 (2016): 502–18.

indexes may not produce such a model, it may contribute to our understanding of how reading fits into the search for information in books. We also do not know very much about how people use indexes in print books let alone in ebooks²⁹ nor how using an index relates to learning in the context of academic study. If we did, could we use some of that knowledge to improve other finding aids, such as search boxes? They are very popular but are they always useful? What if we could integrate the index into the search box using predictive text or other means? Obviously, there are many ways of using ebooks that could influence how we create them.

Another participant in the study lamented,

It does annoy me in ebooks when they don't have the page numbers. They often do, but they also don't. I want to know what page I'm on, especially if I've been given a reference for something.

Not being able to use a page number for a reference was a problem for this student, and I suspect it may be so for other university students and staff. In this case, readers are very directly missing physical page numbering for its ease in referencing and text navigation, but it could be replaced by some other marker of where they are in the text, such as a paragraph number or section number.³⁰ As well as finding out more about how readers use ebook indexes, I am picking up other small details such as this about how the absence of pages in reflowable ebooks affects readers.

Where to from here?

Even if we do figure out better ways to create ebooks, the elephant in the room will remain the device that is used to read them on and how well the features possible in the ebook format, including active indexes, have been implemented by publishers and ebook reading system developers. Nakajima, Shinohara and Tamura evaluated the ability of ebook reading systems to implement functions required for electronic textbooks in the EPUB format and found that active indexes worked only partially, mainly because the ebook reading devices did not support the most recent specification.³¹ Informal explorations by indexers using active indexes with various tablets and ereaders have also revealed that the device

²⁹ Mary Coe, "Where Is the Evidence? A Review of the Literature on the Usability of Book Indexes," *The Indexer* 32 (2014): 161–68.

³⁰ Glenda Browne, "Section and Paragraph Number Indexing," *The Indexer* 30 (2012): 177–79.

³¹ Toshiya Nakajima, Shun Shinohara, and Yasuhisa Tamura, "Typical Functions of E-Textbook, Implementation, and Compatibility Verification with Use of Epub3 Materials," *Procedia Computer Science* 22 (2013): 1344–53. It should be noted that results for other functions showed similar lack of implementation; for example, the full-text search function did not work at all in several reading devices.

used to access the ebook can have a significant impact on the user experience.³² More recently, Jan Wright and I have been investigating how one of the indexes we created for my sensemaking study is displayed in ereader apps and devices.³³ So far, we are finding this to be inconsistent and often very confusing. So, it may be not only an issue of information seeking behaviour but whether the device or app is sophisticated enough to handle a subtle level of linkage between the index locator and the text.

All of this comes back to the evolution of the book—where should we go from here?³⁴ The unimaginative methods of creating active indexes for ebooks in reflowable formats without pages (see Figure 1) could be the canary in the coalmine warning us that we need to do better if we want to create ebooks that can be used in all the ways that we want to use books.

Charles Sturt University

³² See in particular the following reports: Pierke Bosschieter, “Ereaders: Exploring How Well They Work,” *The Indexer* 31, (2013): 60–61. James Lamb, “Producing an Ebook - How Hard Can It Be?” in *JALamb.com* (2012).

³³ Mary Coe and Jan Wright, “Looking for Needles in a Haystack: How Do Ebook Reader Applications Handle Active Indexes? Part 1 - Ereader and Web Browser Software,” *The Indexer* 37 (2019): 125–40; Mary Coe and Jan Wright, “Looking for Needles in a Haystack: How Do Ebook Reader Applications Handle Active Indexes? Part 2 - Dedicated Ereader Devices,” *The Indexer* 38 (2020): 29–44.

³⁴ Yes, the pun is intended.