

Miha Kovač, Angus Phillips, Adriaan van der Weel and Ruediger Wischenbart

What is a Book?

Miha Kovač, Angus Phillips, Adriaan van der Weel, and Rüdiger Wischenbart

Abstract: The aim of this paper is to reach a level of conceptual clarity about what we call a book.

The motivation for this exercise lies in the desire to chart the trajectory of the book as a cultural phenomenon in light of the gradual move to shorter textual expression that is taking place alongside the delivery of stories in other forms besides text. For this purpose the article takes a historical perspective without, however, attempting to chart all the phases in the development of the book. Concurrently with the move to shorter textual expression, in the digital reading environment the basic elements of the 1964 UNESCO definition of the book (printed, a minimum number of pages) have had to be left behind. Alongside the arrival of new publishing business models, the entire notion of the book is in jeopardy. This set of developments calls for a fundamental reconsideration of how we define a book in relation to other book-like objects and text forms. The approach taken is iterative, moving closer towards a definition of the book whilst acknowledging the arrival of offspring such as the ebook and audiobook.

Key words: audiobook, book, ebook, book market, book definition, reading research, publishing statistics.

As we have noted elsewhere (12), the definition of the book is especially important for publishing and reading statisticians, as one of their main units of measurement becomes hard to grasp. With this paper, we therefore try to conceptualize and close a gap that opened up in previous debates on reading research and publishing statistics that took place as part of the COST E-READ Action between

2014-2019, at the By the Book conferences in Florence in 2017 and 2018 and at the WIPO book statistics workshop in Geneva in 2019.

It should be noted that we regard the issues around book statistics as far from being of interest only for the book industry. As discussed below, we read printed books in a different way as compared to textual content on the web (1, 13, 15, 23): printed books (and their digital equivalents) remain one of the main bastions of deep or immersive reading whilst web content is predominantly skimmed. When viewed alongside other media statistics, publishing and reading statistics therefore tell us a lot not only about changes in the tools we use to produce, store, retrieve and digest information, but also indirectly about the changing ways we read and think as a result of using these tools.

The discussion about what is a book has at least three layers. First, it makes a vital contribution to the debate about the importance of linear long-form reading in contemporary society. Second, it underlies the regulatory policy mechanisms, such as lower VAT rates and fixed prices that came into being because of the special status accorded to books for the last two centuries. Thirdly, the debate gives us a better insight into the historically contingent nature of the phenomenon of books and reading in human civilization.

For a start, we will need to question why we call objects as diverse as clay tablets and codices 'books'. The ambiguities and uncertainties regarding the definition of the book go way back in history and the problems we have with the definition today mirror problems with the definition of a book as an historical object.

A historical controversy: were we too liberal in calling different textual objects 'books'?

Visually recognizing ancient scrolls as predecessors of printed books requires a leap of the imagination or at the very least a certain amount of historical knowledge. Besides presenting a text on a paper surface, ancient scrolls and today's printed books share few common physical features. The same goes for the even older ancestors, clay tablets, compared to their newly born descendants, ebooks.

The rationale behind an all-embracing historical perception of the book that encompasses all of these very different objects is most clearly expressed in the statement 'that the book performed similar tasks (they were read, bought, sold, collected ...) in different periods even if the physical form of early books was different from today's' (9; see also 11 and <https://en.wikipedia.org/wiki/Book>).

The viewpoint that these various incarnations of the book may look very different, but that at a certain level of abstraction they are nevertheless the same because they perform the same – or at least similar – tasks, however, neglects not only physical differences among textual objects but also the vastly different cultural and societal context in which they come into being and the function they perform. In the same way as, for example, chairs were symbols of high office and status in ancient Egypt, Greece and Rome, ancient scrolls were symbols of status and authority too. As centuries passed by, with the exception of a few royal thrones, chairs lost their aura of authority and evolved from seats for dignitaries to mass-market products, without significantly changing their visual physicality as four-legged carriers of human bodies in a sitting posture. On the other hand, scrolls and long before them clay tablets, almost entirely disappeared, and with the advent of Christianity the codex as a physically entirely different carrier of texts came to prominence. This was the start of centuries of modifications of its physical appearance and of its social and cultural status, similar to

the transformations of the chair. Just as chairs evolved from thrones to mass-market products, so did codices evolve from objects of intellectual and religious authority to predominantly mass-market products of the entertainment industry, even though their appearance did not radically change. An early printed and illuminated Bible bought by a rich Church luminary in the fifteenth century performed a vastly different social task than a paperback copy of a Dan Brown novel bought by a commuter in a subway kiosk in contemporary Stockholm or Moscow. Nevertheless, if he was taken by time machine from medieval Mainz to a contemporary bookstore, Gutenberg would likely recognize a contemporary paperback as a distant grandchild of his Bible. It is far less likely that a similar time travel operation would allow a Babylonian administrator to recognize the functional relationship between a paperback, Gutenberg's Bible and a clay tablet.

It would be an exaggeration of course to say that the social and cultural status of books changed so dramatically between the fifteenth and twenty-first centuries that books entirely transformed from objects of religious, spiritual and intellectual authority to parts of the entertainment industry. Even in the first centuries of print, there was a significant number of lowbrow books (5) and today there remain many books that command intellectual authority. Nevertheless, through centuries of technological development, books slowly ceased to be the only medium for the organization and dissemination of knowledge. Meanwhile growing literacy, the development of printing technology and the growth of consumer publishing and bookselling firmly established books in the high street as mass-market entertainment products. The final blow to the central status of books in dissemination of knowledge took place with the arrival of computers as universal machines that allowed the production and consumption of textual and other cultural artefacts on the same gadget: screens rapidly became the most common substrate for the dissemination of information. All in all, by the turn of the millennium the 'order of books' as the main organizing principle of knowledge ran out of steam and the book lost its privileged position in the media landscape (2, 20).

The long tradition of using the term 'book' for describing such physically and historically different textual objects as clay tablets, scrolls and printed paperbacks could therefore be seen as a residue of an 'order of books' mentality that - without much reflection - projected to the past a perception of the book as a medium of intellectual expression as it came into being in the golden era of print. That is, declaring all textual artefacts from the pre-codex era as predecessors of books delivered a long and respectable tradition that affirmed their position as symbols of intellectual authority and as the main knowledge-organizing principle. Yet besides the physical differences between clay tablets and codices, the role of knowledge and its dissemination in Mesopotamia could hardly be compared with processes like the scientific revolution that took place in Europe in the era of print. Clay tablets were invented to record commercial transactions and administrative accounts, and only a small proportion of surviving clay tablets contain any form of text that we would regard as book content today. Why then declare clay tablets to be predecessors of books and not of, for example, accounting sheets?

If we further examine such ambiguities around the use of the word book, the present only appears to be a continuation of the past. Just as there are no visual similarities between Sumerian clay tablets and Alexandrian scrolls, there are none between a modern paperback and the Chinese online reading platform Tencent. More importantly, these four textual substrates were produced in different historic, technological and cultural circumstances where they performed different social roles and their different formats evoked different physical interactions with the text and consequently different modes of reading. They make varying assumptions regarding their readers' literacy in the wider sense of understanding the function and social significance of textual information. The only common trait is that they are all text-based, but that seems hardly enough to call them all books, just as we would not call an email message or the neon lettering of a pharmacy sign books.

To make things more complicated, many contemporary objects we call books are not text-based.

Why is a lavishly illustrated gardening guide a book but not the latest issue of the *New York Review of Books*? Why do we call a silent book, i.e. a book for children based only on artwork without text (title and imprint aside) a book? Furthermore, why categorize colouring books for adults – a great sales hit around the world a few years ago – as books although they are not reading objects and have no narrative and no text? What about an audiobook, which we listen to but do not *read*? And when and why does a digital textual document become something that we call a book? Further, if we again take a historic perspective, if clay tablets were the first medium for transfer of written information, why do we consider them the ancestors of books but not ancestors of newspapers or letters? Isn't calling them books far too narrow a nomenclature? Why not acknowledge them as ur-grandmothers of the entire media family, i.e. not only of books and papers, but also of radio, television and the web?

The answer is simple: if we consider clay tablets and scrolls not as books but as the predecessor of the entire media family, over time the book changes into one of its many descendants and loses its historical and cultural centrality. This centrality was the result of the order of the book in which the book as a commercially exploitable commodity was indissolubly linked through an entirely contingent historical accident to the intellectual content of the text. The current challenge is that we need to move to an understanding of how different formats of books may have never been forms of the same cultural object, recognising that such a broad definition of the book was the result of our failure to analyse a specific social and cultural mentality that made us define a book in this way.

Yet before we attempt to redefine the book as a concept, why do we need a definition at all?

Obviously, as we suggested above the reasons are not only academic. The 'order of books' brought to life a set of legislative measures such as lower VAT, fixed book prices and subsidies to ensure the

privileged status of books as carriers of knowledge and culture in the media market. These persist today and any ambiguities in definitions of books have practical consequences for present-day policy-makers, librarians, publishers and booksellers, as the concept of the book continues to underpin legal, fiscal and business rules that govern book markets in terms of availability and price. Put simply, if we cannot identify a book, how can the book market regulatory mechanisms be applied sensibly? Further if we agree that there are no longer cultural differences between books and other media, is there any longer the need to apply such mechanisms? Or do such differences still persist? And if yes, what is their gist?

The debate about 'What is a book' therefore covers both the academic concern to chart the trajectory of the book as an important phenomenon in cultural history as well as very practical issues of cultural and economic policy. Let us start this exploration with the issue of book identifiers.

The growth of book identifiers

In 1964 UNESCO published the following definition:

A book is a non-periodical printed publication of at least 49 pages, exclusive of the cover pages, published in the country and made available to the public.

The essence of the definition was in the word 'printed': back then, book printing technology made financially viable print runs larger than 500-1,000 copies containing more than three printing sheets, effectively making impossible self-publishing as we know it today and excluding all booklets that had fewer than 48 pages. (Note that also excluded were illustrated publications for children with fewer than 48 pages.) By using the word 'non-periodical' daily papers, magazines and scientific journals were excluded from the realm of books. A journal published annually in a book format was considered a periodical publication and a novel by Eleanor Hibbert, who for fifty years under a

variety of pen names published more than one book per year, was not. Simply put, UNESCO defined books as those textual artefacts published by commercial publishers whose codex-like format was visually different from other print media such as magazines and daily papers, and whose business model was the copy sale of a single item (3).

In the second step, after 1967, every publication that could be described as a book received an International Standard Book Number (ISBN), making the cataloguing of the ever-growing quantities of newly published titles much easier. The system worked (and still does) in that an ISBN was assigned to each edition and variation (except identical reprints) of a book, clearly showing the commercial purpose of the ISBN identifier. A hardback and paperback of the same title have different ISBNs. In following years, having an ISBN became a key form of identification for the book.

Any contradictions found in the UNESCO definition were pragmatically dealt with by the definition produced by the United States Postal Service:

A bound publication having 24 or more pages, at least 22 of which are printed and contain primary reading material, with advertising limited only to book announcements.

The artistry of this definition is that on the one hand it included illustrated books for children, with little text and fewer than 48 pages, and at the same time excluded commercial catalogues, newspapers and magazines containing advertising. Yet a periodical without ads, such as a scientific journal, could be considered to be a book too. In 1971, this issue was resolved with the introduction of a separate identifying number for periodical publications, the ISSN (International Standard Serial Number). A publication with an ISSN is therefore not a book although it might look very much like one - a good example of such a book-like periodical is *Book History*, published annually by the Society for the History of Authorship Reading and Publishing (SHARP).

With the arrival of digitization, these definitions and identifiers started to look shakier, as all digital texts 'arrive to the reader in the same material form, which suggest that all texts are leveled, reduced to the same physical experience, encouraging the same cognitive-emotional experience. When *Hamlet* and *Gilligan's Isle* and a video of a funny cat and an email from your boss and a text from your boyfriend arrive in the same material device, they all become part of the same experiential flow, with fewer markers to remind us of their real differences' (16). When a teacher lectured in the 1960s on the differences between textual media, they could show copies of magazines, newspapers, scientific journals and books. No such thing can be done with digital media today: digital objects have no physicality and their screen appearance may be identical. In brief, the technology behind digital media is not framing them in the same way as printing technology framed books, magazines and papers as distinct genres when UNESCO coined its definition of the book in 1964.

The closest fit to older definitions of the book - in terms of content it is largely its digital equivalent – is the vanilla ebook (17). It has no fixed pages and the text adapts to the size of the screen. Hence, its appearance is different from its printed cousin although the content is exactly the same. As long as such textual continuity persisted, publishers and librarians considered ebooks as one more book format and assigned to them ISBNs alongside the hardback and paperback editions of the same title.

The situation started to change when a growing number of ebooks were published only in digital form. Without the constraints of print, an ebook can be 2,000 or even 1m words. As a result some publishers and authors published short stories as ebooks, and academic and educational publishers offered their e-textbooks and e-monographs for sale in chunks, such as chapters. As online publishers and booksellers found that the attention span of an average reader was becoming shorter, this stimulated ebook authors to write shorter texts (1). On the other hand, advances in

printing made possible low print runs of books with a small number of pages, so the trend towards shorter extents moved into paper-based publishing. The system of ISBNs now encompassed short digital and printed publications with low print-runs and sales, compromising the statistical record.

A further complication in book statistics was the introduction of the Amazon Standard Identification Number (ASIN). The system works in a way that 'each product sold on Amazon.com is given a unique ASIN. For books with a 10-digit ISBN, the ASIN and the ISBN are the same ... However, the Kindle edition of a book will not use its ISBN as the ASIN, although the electronic version of a book may have its own ISBN' (Wikipedia, accessed 14 August 2018). The consequences of such developments are sometimes confusing. Kristen Roupenian's *Cat Person*, for example, was published first as a short story in the *New Yorker* (4 December 2017); after it went viral, the story was published as a small format paperback with 52 pages (and with a lot of white space around the text) by Jonathan Cape in 2018. The book had an ISBN like a real book – but the text on the cover announced that the author's 'first book was to be published in 2019', suggesting that *Cat Person* was not a book even though it has an ISBN. Further, when sold as an ebook and audiobook on Amazon, it has an ASIN, and another ISBN when on sale as an ebook on Kobo. To confuse matters further, as 'global expansion has changed things so that ASINs are only guaranteed unique within a marketplace', it could be the case that 'the same product may be referred to by several ASINs, and different national [Amazon] sites may use a different ASIN' (Wikipedia, accessed on 14 August 2018). Technically, this means that a single short story such as *Cat Person* has at least five different identifiers and statistically counts as at least five book titles – whilst in the pre-digital era it would most likely have remained a short story in a magazine until collected in a book with other stories.

New business models

New publishing business models play havoc with the implied connection between sales and reading, whilst self-publishing questions the concept of the 'book industry'. Before the rise of online physical sales and the advent of ebooks, publishers mainly sold books to booksellers, and booksellers sold books on to readers or libraries: the main business model was to ensure that books were dispersed through as many retail channels as possible, where as many readers as possible would find and buy them. Neither booksellers nor publishers knew who bought a given title and whether they actually read it; also unknown were their reading tastes and the size of their home library. There were other business models of course, such as book clubs and direct sales, but in terms of the number of books sold, their market share was in most countries less important than sales via bookstores.

With the arrival of digital technologies, direct transactions between the publisher or the bookseller and the customer started to grow. As a result, online booksellers harvested data about the purchasing habits of their customer and in the case of ebooks, data on reading pace, reading time and geographical location. Further, in order to put an ebook on sale on Amazon or Kobo, authors do not need publishers: self-publishing a book online is a simple process and if needed, editorial and marketing services once executed by publishers can be purchased online by the author. As a result, self-published authors became publishing entrepreneurs; besides investing their time in writing and promoting their books as they did in pre-digital times, they started to invest their own money in the production and marketing of the books. Further, in an echo of the commercial libraries that were part of the Western book landscape between the eighteenth and mid-twentieth centuries, the early twenty-first century saw the rebirth of the subscription model. In 2018 the Chinese ebook subscription service owned by Tencent had 192m users and offered 5m books, many of which were self-published. Similarly successful was Amazon's Kindle Unlimited model in USA and Europe (14).

Throughout these processes, the way books were authored changed too: *Fifty Shades of Grey* was first published as a piece of fan fiction and the author, E. L. James, communicated with her readers, receiving advice about the text and its development (Kovaç and Wischenbart, 2018). As ebooks evolved attributes that printed books do not have, such as reading metrics and the sharing of reader notes, comments and highlights, both reading and authorship became more social (10) and book statistics became ever more complicated: are all the pre-versions of *Fifty Shades of Grey* stand-alone book titles too or just drafts reviewed by an unusually high number of peers?

The method of payment to authors for books in subscription schemes further contributed to such statistical confusion. For centuries, the backbone of authors' fees was the number of printed copies sold. However, for books that are part of a subscription scheme, Amazon introduced payment according to the number of pages read, an unimaginable methodology in the print era. From the authors' point of view this means that a 30-page story that is read by 100,000 readers earns more money than a 700-page novel that 100,000 readers start to read but which most of them stop reading on page 15; in print publishing the situation is reversed. Such a system encourages authors to write shorter stories: if successful, readers will buy them as a series and at a certain point they could be collected together in one volume in print. Meanwhile each story is called an (e)book and each gets its own ASIN or ISBN.

Things become even more complicated on self-publishing platforms such as Wattpad, which had 75 million users in 2018. There, most successful texts operate as series, with the same protagonists and a recognizable storyline. In 2017, Wattpad started to work with book publishers to publish the most-read series of stories on the platform in a print or in an audiobook format. In January 2018, they proudly declared that 'hundreds of Wattpad stories have been published as traditional books - many as best-sellers – and dozens have been adapted or licensed for TV, film and digital video, appearing

on screens around the world' (wattpad.com, accessed 16 August 2018). In January 2019 the company announced the launch of its own book publishing division.

As a result, the text that was not considered a book when originally available on Wattpad is turned into a book with an ISBN when publicly available in print and sold in bookstores. Moreover, to complicate further the issue of identity, after being published in book format, a Wattpad story series remains freely available on the platform. Technically they are a book when available on Amazon or Kobo as they have an ASIN and/or an ISBN, regardless of the fact that they are not when read digitally on the Wattpad app, even though the text looks pretty much the same in both cases. Therefore, the text as an object changes depending on when, where and how we observe or measure it. The book has its quantum moment.

Towards a definition

The unsuitability of existing book definitions in the digital realm was noted already in 2008, when Rüdiger Wischenbart observed that 'the UNESCO definition of the book did not bother to say much about the originator or the industry that produced the book, nor its realm, distribution or economy; nor does the definition discuss the book's audience or the process of reading' (22). Absence of these aspects from twentieth-century definitions of the book indicates that in the 1960s they were still self-evident. As described in this paper, they ceased to be so when the rapid growth of screen media disrupted both book industries and reading modes.

Does this mean that the book as a clear-cut definable object has already died without anybody noticing it? And again, do we need a definition of the book at all? The answer to this question is twofold. If we want to differentiate media activities in the digital and analogue realms, understand the changing scope of the book industry, and measure and research the trajectory of long-form

reading, then the answer to this question is yes, whilst the answer is no if there is little interest in such areas. In the latter case, publishers and booksellers can continue to assign an identifier to any object sold through channels that require this type of identification, regardless of its book or non-book nature, as is current practice. In this case, the ISBN will remain a useful trade tool and will further decline in utility as a statistical tool. Consequently, our understanding of the nature of book production will continue to decline.

Even more, if the *differentia specifica* of a book disappears, people might start answering 'yes' to the question whether they read books not only when they listen to audiobooks or podcasts, but also when they play with story apps or skim online fiction platforms such as Wattpad. If this happens, this will be an indicator that the reading of books as embodiments of long form texts ceased to have a special cultural meaning that needs to be protected from raw market forces. Fixed book prices, lower VAT rates, cheaper postage, subsidies for publication, translation and bookshops will disappear and the need to define a book at all will follow suit. This is not a plea for their preservation, but an observation that the very foundation on which we can consider the desirability of such measures might no longer be solid. Such policies will go south if long-form reading ceases to be seen as a training for thinking, through the use of working memory and the growth of background memory via depth and length of vocabulary (13). We can therefore speculate that the survival of the book as a medium with special status depends on whether society continues to recognize long-form reading as an activity of sufficient cultural importance to keep supporting the publishing industry through market regulation, despite the decline of the book's importance as the bearer of textual knowledge, culture and information.

In sum, the debate about the definition of the book should also be a debate about the importance of reading and not only about the physicality of the book. A simple update of the UNESCO definition could be along the following lines:

A book is a non-periodical printed or digital publication primarily designed to be read.

But does this go far enough or too far? And what kind of reading is involved?

Digital disruption has given fresh momentum to reading research, offering up some surprising findings: in 2017 to 2019, three meta-studies covering more than 200 earlier reading studies with more than 200,000 participants, showed that when reading the same long informational text, understanding differs depending on whether text is read from screen or from paper (19, 6, 4). Perhaps for the first time in media history, it became clear that different substrates evoke different ways in which we read and understand written information.

Kovač and Van der Weel (13) summed up the correlations between reading substrates and reading modes through a triad of skimming (online short textual, audio and video content, predominantly accessed on screen); immersive reading (long and medium-sized linear texts, read on screen or paper, such as genre fiction); and deep reading (short and long-form cognitively demanding linear texts such as poetry, literary fiction and serious non-fiction, predominantly read on paper).

Consequently, the mode of reading and the substrate from which it is read ought to become one of the attributes that defines a book. From the reading research point of view, a book could then be described as:

A linear long-form text that can be read on paper or screen and evokes deep or immersive reading.

However, from the perspective of publishing and bookselling, such a 'reading' description is not inclusive and precise enough:

- a) it still assumes a cognitive equivalence of paper and screen, which does not do justice to the fact that deep or immersive reading tends to be better served by paper than by screens;
- b) it banishes many objects that we are used to calling a book such as audiobooks, silent books, colouring books and gamified book-based story apps;
- c) the categories deep, immersive and long-form are subjective and cannot be measured, defined and standardized in a single way for each and every category of books and for each and every reader – at least not in a similarly exact way as UNESCO described the book through the frame of printing technology.

If defining a book solely as a long-form reading object is not good enough, what about the organizational properties of linearity in book format that exist in printed books? Could they be seen as the main attributes of the book even in the case of the absence of the text or in digital formats? A variety of organizational tools came to flourish in printed books in the late sixteenth century and have been with us for so long that they have become part of every book reader's conscious or unconscious expectation. As shown by Elisabeth Eisenstein in her seminal work (8), the information explosion of the early print era required tools to navigate the rising sea of information, and page numbers, chapters, alphabetical indexes, title pages, author's names and alphabetical organization of texts (in dictionaries and encyclopedias) could come into being precisely because printed text was fixed to a page and multiplied in identical copies. Cope and Phillips (18) identified such tools as the main elements of the 'information architecture' of the printed book:

- Title
- Cover
- Title page
- Author
- Linear structure: page numbers, chapters, alphabetical order
- Body text [and images]

As seen in Table 1, by comparison with twentieth-century definitions of the book, the combined concepts of (a) long-form, deep reading and (b) the information architecture of a book brings us closer to the textual objects that book industries have traditionally sold and distributed. Yet, bearing in mind the fluidity of digital media, we should add a note of caution: we consider books as narrative objects that can be *predominantly* accessed by reading (because of course we can access the same book content not only by reading but also by listening to a narrator reading the text; also a silent book is telling a story through pictures). Also not all books necessarily have a complete set of the elements of the information architecture of the book – we propose that if an artefact is to be considered a member of the book family, it should have the *majority* of them. Last but not least, a book becomes a book when made public – compare the novel sitting in manuscript form in the bottom drawer of the writer’s desk.

| Book | UNESCO | UPS | ISBN | Long-form, deep reading | Information architecture – linear structure, textual elements |
|----------------------------------|---------------|------------|-------------|------------------------------------|--|
| <i>The Great Gatsby (print)</i> | Yes | Yes | Yes | Yes | Yes |
| Ebook of <i>The Great Gatsby</i> | No | No | Yes | Yes | Yes |
| Kindle self-published title | No | No | No | Yes | Yes |
| <i>Cat Person (print)</i> | Yes | Yes | Yes | No | Yes |
| <i>Logos</i> publishing journal | No | Yes | No | Yes | No |
| Audiobook | No | No | Yes | No | No |
| Wattpad fiction | No | No | No | ? | No |
| Colouring book | Yes | Yes | Yes | No | ? |
| Silent book | ? | Yes | Yes | No | Yes |

| | | | | | |
|--|-----|-----|-----|-----|-----|
| <i>Solar System</i> app | No | No | No | No | No |
| Wikipedia | No | No | No | No | No |
| Encyclopaedia Britannica (print) | Yes | Yes | Yes | Yes | Yes |

Table 1 Book definitions in practice

If we consider information architecture and long-form, deep reading as the defining attributes of the book, only printed books and vanilla ebooks count as true representatives of the species. However, as we have shown, many ebooks are shorter, involving texts that would not have been issued as separate paper book titles twenty years ago and they have evolved a set of attributes such as sociability that printed books do not have (10). Listening to an audiobook is (strictly) not reading as you are not using the part of your brain that decodes the meaning residing in the graphic signs that make up the text.

In short, book offspring differ: some have more and some have fewer of the full set of attributes of the printed book. If we borrow a metaphor from biology, books are not a single organism but a genus within which separate species came into being.

| Artefact | Information architecture | Long-form, deep reading | Additional attributes |
|------------------------|---------------------------------|--------------------------------|------------------------------|
| Printed book | Yes | Yes | None |
| Ebook | Yes | Yes | Flowable text; social |
| Comic/illustrated book | Yes | No | None |
| Silent book | Yes | No | Narrative through pictures |
| Audiobook | No | No | Narration of text of book |
| Colouring book | Yes | No | Colouring |

Table 2 The species of the book

Although we have attempted to provide a set of considerations for a definition of the book, the task still seems daunting. Indeed others have made similar attempts and failed. Some years ago Zoran

Velagić (21) came to the verdict that 'consensus does not exist "even at the level of a basic definition"', citing Robert Escarpit, who reached the same conclusion in 1966, writing that 'no one has yet been able to provide a complete and final definition' (7, p. 19).

The only way out of the predicament appears to be the 'quantum approach', respecting the viewpoint of the observer. The Unesco and US postal definitions remain relevant definitions as they include what we have identified as central book attributes. The minimum length, together with the emphasis on textual content and boundaries, enable us to identify items that require long-form (even if not necessarily deep) reading. To these criteria we have added the fourth criterion that the content should be ordered as an information architecture (possessing a linear structure and key textual elements such as a title and chapters). The following criteria make up our definition of the book:

- A minimum length
- Emphasis on textual content
- Boundaries to its form
- Book information architecture

The artefact that meets all four criteria is the printed book. Ebooks remain the closest to printed books: they are meant for reading, have a long-form narrative and apply recognizable organizational tools to linear content (and adding some new ones, such as the percentage of text read). Of course, ebooks also feature attributes that printed books lack: they are more social than printed books, allowing comments and highlights to be shared and discussed, and they measure and benchmark

reading dynamics (10). Audiobooks have a longer history as they became a serious market proposition with the arrival of the cassette tape in the 1970s. They were narrated replicas (or condensed versions) of printed books and as such were linear and could be said to adhere at least to an extent to the same information architecture as their printed parents (with a cover sleeve and a sequence of tapes to be listened to). With the advent of the Kindle and then subscription services, ebooks and audiobooks converged as downloadable content, with the facility for the user to switch between reading and listening. Audiobooks should be considered as more distant relatives of books as they involve no reading in the sense of decoding characters for semantic meaning. Silent books are based only on artwork and have little by way of text. Whilst textual reading is largely missing, they have more book-like information architecture than audiobooks. The most distant offspring of the book (and as such, the most questionable member of the genus) is colouring books as they have only a small set of elements of book architecture such as pagination but little narrative (unlike a silent book) and no text.

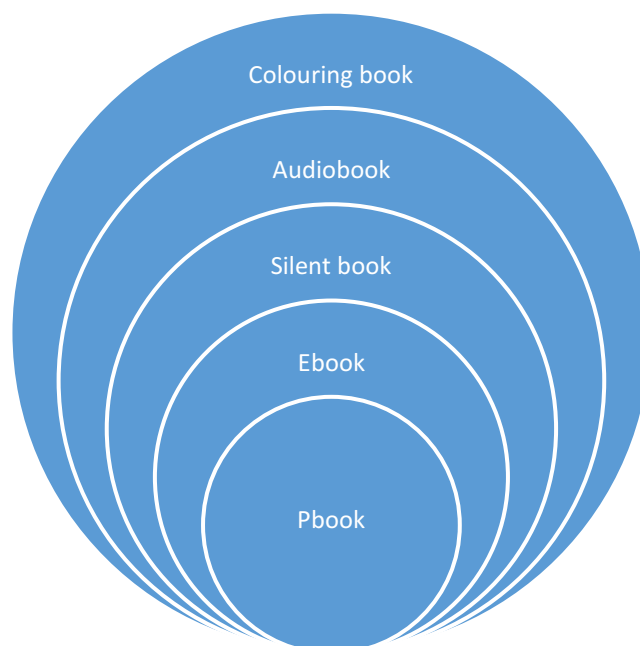


Figure 1 The hierarchy of the book

Our overall model is a hierarchy of the book (Figure 1), with the printed book at the core and the colouring book as the layer furthest removed. Such an onion-like model should also have a quantum element. Not only do texts become books depending on when and where in their lifecycle they are observed (e.g. not in manuscript form, nor when posted online), but their long-form nature depends on genre. A 36-page illustrated book can be regarded as long-form in the children's category but not so in the adult one.

Conclusion

Through an iterative process this paper has identified four attributes of the book: length, textual content, boundaries to its form, and book information architecture (e.g. linear structure and key textual elements). The application of these attributes to particular species of the book, from the printed book to the colouring book, has produced a hierarchy of the book. The species of the book genus differ: some have more and some have fewer of the full set of attributes of the printed book. The purest definition of the book demands the presence of all four identified attributes.

In addition to its layers, the model has a quantum element allowing variation according to the nature of the observation. A text on Wattpad is not a book but becomes one as a printed publication or ebook. The specified length will vary from fiction to children's books. This model enables us to take account of a variety of examples where the term *book* is used, whilst reflecting the feeling that the printed book is the purest form in existence. The book as a business model has sought to take in colouring books, diaries, standalone short stories ... as shown by the widespread use of the ISBN in the supply chain.

If we want to measure the fortunes of the book and long-form text, we can see a way forward by use of the hierarchy of the book to bring greater precision.

Sources

1. Baron, N. (2014). *Words on Screen*. New York: Oxford University Press.
2. Chartier, R. (1994). *The Order of Books*. Palo Alto: Stanford University Press.
3. Clark, G. and Phillips, A. (2019). *Inside Book Publishing*, 6th edition. Abingdon: Routledge
4. Clinton, V. (2019). 'Reading from paper compared to screens: A systematic review and metaanalysis'. *Journal of Research on Reading*, 2019.
5. Darnton, R. (1996). *The Forbidden Bestsellers of Pre-Revolutionary France*. New York: W. W. Norton.
6. Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: A meta-analysis on the effects of reading media on comprehension. *Educational Research Review* 25, 23-38. DOI: [10.1016/j.edurev.2018.09.003](https://doi.org/10.1016/j.edurev.2018.09.003)
7. Escarpit, R. (1966). *The Book Revolution* (Paris and London: Harrap and UNESCO)
8. Eisenstein, E. (1979). *The Printing Press as an Agent of Change*. Cambridge: Cambridge University Press
9. Howard, N. (2009). *The Book: The Lifestory of a Technology*. Baltimore: Johns Hopkins University Press.
10. Johnson, M. J. (2019). 'What is a Book?'. *Publishing Research Quarterly*, 1-11. DOI: [10.1007/s12109-018-9622-z](https://doi.org/10.1007/s12109-018-9622-z).
11. Johnston, C. (ed.) (2019). *The Concept of the Book*. London: School of English Studies, University of London.

12. Kovač, M., Phillips, A., van der Weel, A., Wischenbart, R. (2017) 'Book Statistics: What are they good for?'. *Logos*, 4: 7-17.
13. Kovač, M. and van der Weel, A. (2018). 'Reading in the post-textual era'. *First Monday*. 23 October.
14. Kovač, M. and Wischenbart, R. (2019). Globalization and Publishing. *Handbook on Publishing* (eds. Michael Bhaskar and Angus Phillips) New York: Oxford University Press
15. Mangen, A. et al. 2013 'Reading Linear Texts on Paper Versus Computer Screen: Effects on reading comprehension', *International Journal on Education Research*, 58: 61–8
16. McLaughlin, T. (2015). *Reading and the Body: The physical practice of reading*. New York: Palgrave McMillan
17. Phillips, A. (2014). *Turning the Page: The evolution of the book*. Abingdon: Routledge.
18. Phillips, A. and Cope, B. (eds.) (2006). *The Future of the Book in the Digital Age*. Oxford: Chandos.
19. Singer, L.M. and Alexander, P.A. (2017). 'Reading on paper and Digitally: What the Past Decades of Empirical Research Reveal'. *Review of Educational Research*. DOI: 10.3102/0034654317722961.
20. Van der Weel, A. (2011). *Changing Our Textual Minds*. Manchester: Manchester University Press.
21. Velagić, Z. (2014) *The Discourse on Printed and Electronic Books: analogies, oppositions and perspectives*. *Information Research* 19/2.
22. Wischenbart, R. (2008). 'Ripping off the Covers'. *Logos* 19/4, 196-202
23. Wolf, M. (2018). *Reader come Home*. New York: Harper.