THE IMPORTANCE OF METADATA FOR DISCOVERABILITY AND SALES

CONTENTS

INTRODUCTION

In 2012 and 2016 Nielsen Book conducted research into the relationship between physical book sales in the UK and bibliographic metadata. The results were published in two White Papers: *The Link Between Metadata and Sales* (2012) and *The Importance of Metadata for Discoverability and Sales* (2016). Both White Papers concluded that there was strong evidence between the completeness of book metadata and book sales, illustrating that rich metadata assists the discoverability of books.

In this report, *The Importance of Metadata for Discoverability and Sales* (2020), we seek to understand whether this statement still rings true while also, for the first time, shine a spotlight on consumer research data from Nielsen Books & Consumers to understand what elements of book metadata are most influential to consumers when it comes to discovering and buying books in the UK.

Of course, "discoverability"—i.e. the ease with which a book is discovered or found—has been a buzzword in the publishing industry for a decade. However, as this report will show, how consumers discover books is just as relevant now as it was 10 years ago. Arguably, in an age where online retailers are taking an ever increasing share of the UK book market, accurate and rich metadata has never been more important for authors, publishers, printers, retailers, libraries and readers.

When it comes to "discoverability", hearing from the end-consumer is essential and we're delighted to include Nielsen Books & Consumers data into this report for the first time. However, we're also interested in what happens next: after consumers have *discovered* a book, what influences their decision to *purchase* the book. We will be including research on this area, focusing on elements directly related to a book's metadata.

Book-buyers are the most important people in the book industry. However, it is also important to consider discoverability and metadata within the supply chain as a whole. Knowledge of a book's pagination or spine width, for example, may influence the purchase decision of a book-buyer, but this information is just as important to a printer, a distributor, and a retailer, who have considerations around paper stock, packaging, postage and shelf space to name but a few.
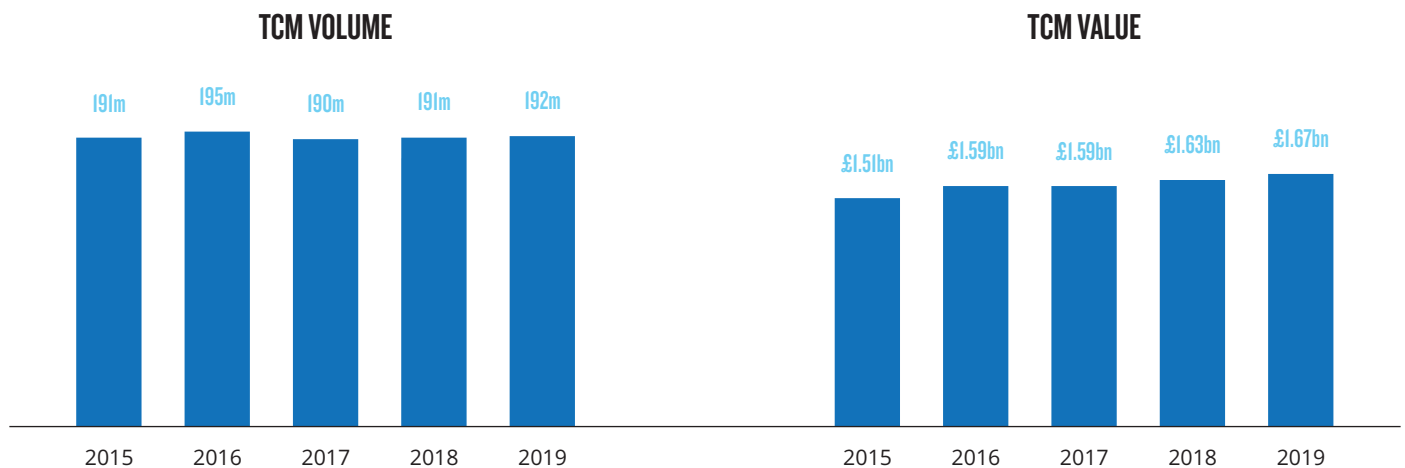
Before diving into an analysis of metadata and book sales, we will first provide some context: what did the UK book market look like in 2019?
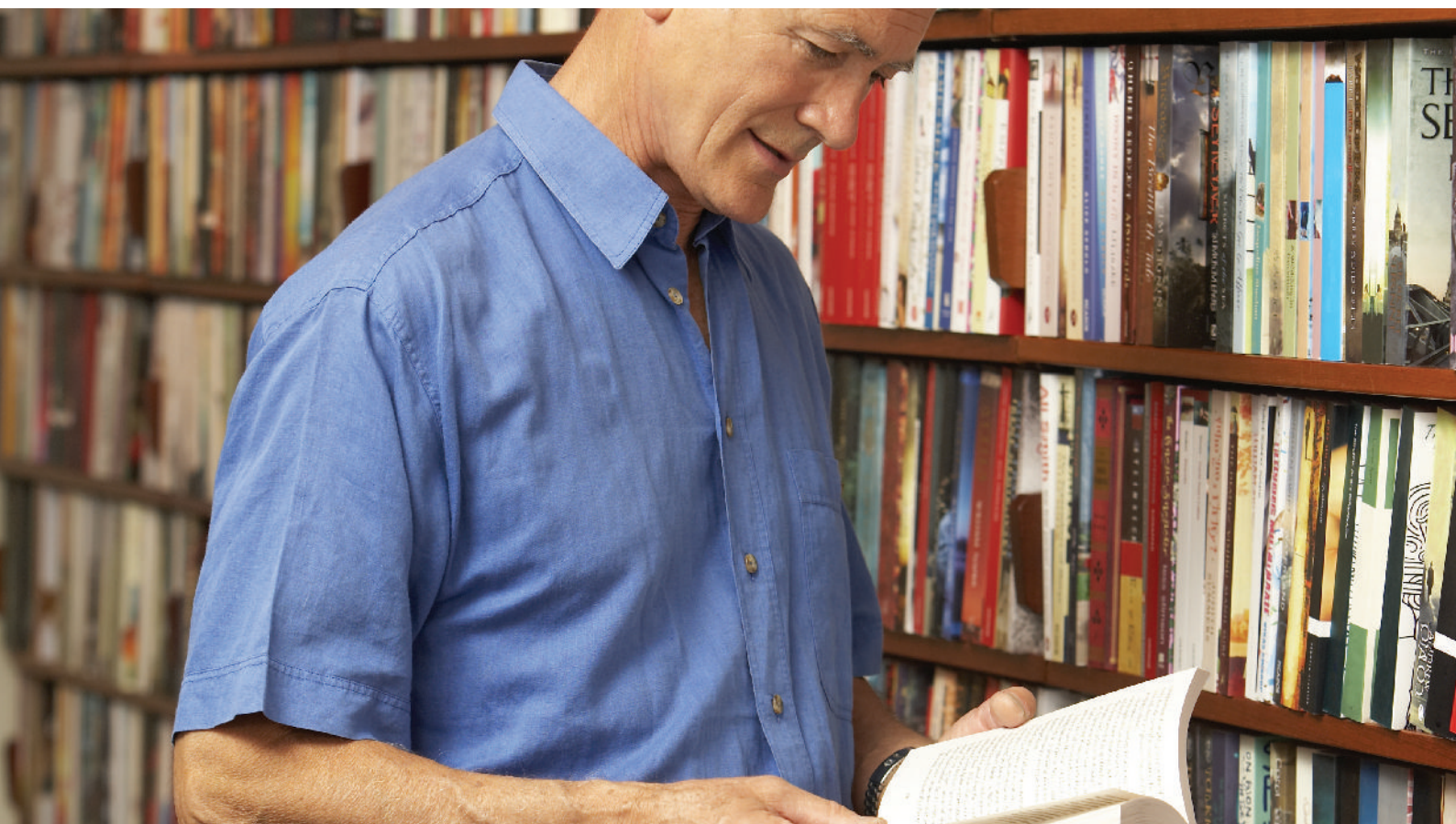
BACKGROUND:
THE UK BOOK MARKET IN 2019

Six printed books were bought every second in 2019 in what proved to be another solid year for physical book sales in the UK. According to data from Nielsen BookScan's Total Consumer Market panel of more than 6,000 book retail outlets in the UK, volume sales grew for the second consecutive year, to 192m, while value sales grew for the fifth consecutive year. In fact, the £1.67bn spent on physical books in the UK in 2019 was the largest amount since 2010 when Jamie Oliver's *30-Minute Meals* and the late Stieg Larsson's *Millennium* series contributed to book-buyers spending a wholesome £1.72bn.

PRINTED BOOK SALES THROUGH NIELSEN BOOKSCAN'S TOTAL CONSUMER MARKET

TCM VOLUME

2015	2016	2017	2018	2019
191m	195m	190m	191m	192m

TCM VALUE

2015	2016	2017	2018	2019
£1.51bn	£1.59bn	£1.59bn	£1.63bn	£1.67bn

But what about the digital market? Whereas Nielsen BookScan tracks physical book sales using point-of-sale retail data, Nielsen Books & Consumers—a monthly survey of circa 3,000 book buyers aged 13-84 that are nationally representative of the UK population—provides the opportunity to gain insights on the UK book market as a whole. Asking buyers a range of questions about the books they have bought not only provides Nielsen the ability to capture insights on sales across all book formats, but allows for focused insights on who is buying books, where from and why.

The Nielsen Books & Consumers top-line reveals that overall, book sales in volume terms across all formats fell marginally year-on-year in 2019 but there was a small rise in terms of consumer spend—to just shy of £2.5bn. Average selling prices climbed 5% year-on-year, by 31p.

VOLUME

355m	352m	▼	- 1%
2018	2019	YEAR-ON-YEAR	

VALUE

| £2.38bn	£2.47bn	▲	+ 4%
2018	2019	YEAR-ON-YEAR	

Spending on hardback books increased 11% year-on-year in 2019 but the value of the ebook market fell by 3%—a factor behind the rise in average selling prices within the market given the average cost of a hardback (£10.20) was more than double the average price paid for an ebook in the UK (£4.12).

Overall, in volume terms ebooks accounted for just under a quarter of all book sales in 2019, and around a seventh of all book sales in value terms.

SALES BY FORMAT:
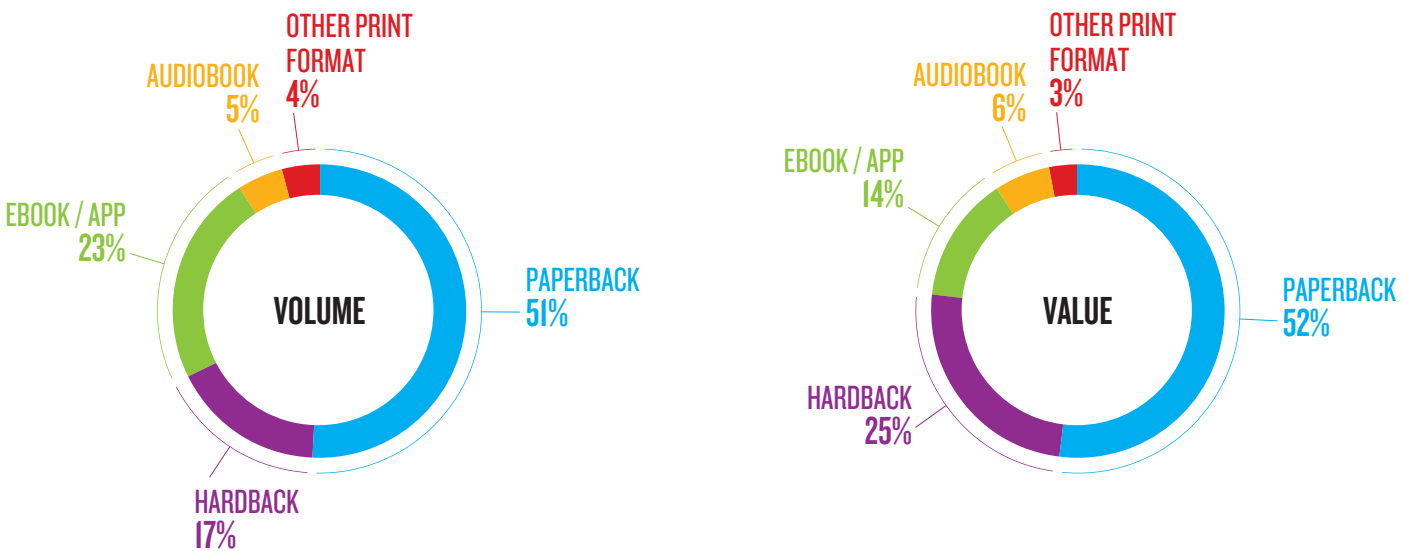
The ebook/print splits differ considerably by broad genre. In volume terms, ebook sales surpassed printed book sales within the Adult Fiction sector in 2019. However, print books continue to dominate within the Adult Non-Fiction and Children's sectors.

VOLUME SALES BY FORMAT BY GENRE:

ADULT FICTION

ADULT NON-FICTION

CHILDREN'S

PRINT 94%

EBOOK 3%

AUDIO 3%
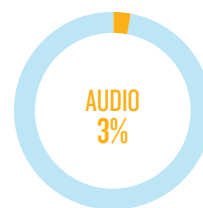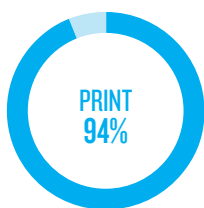
Across all formats, 58% of books were bought online in 2019, while 42% of all print book purchases were made online. Thanks to a high ebook share, 72% of all Adult Fiction sales in volume terms last year were made online, while 59% of all Adult Non-Fiction titles were purchased online.

Comparatively, the majority of children's books were purchased in bricks and mortar stores.

VOLUME SALES ONLINE/ OFFLINE SPLITS

ONLINE 58%

OFFLINE 42%

PRINTED BOOKS

ONLINE 42%

OFFLINE 58%

ADULT FICTION

ONLINE 72%

OFFLINE 28%

ADULT NON-FICTION

ONLINE 59%

OFFLINE 41%

CHILDREN'S

ONLINE 39%

OFFLINE 61%
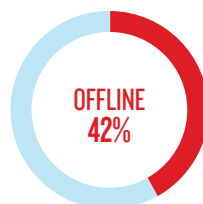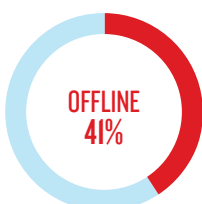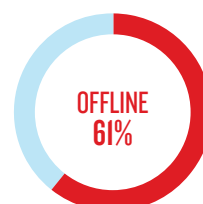
As well as providing an insight into the digital book market, Nielsen Books & Consumers data can be used to provide a demographic analysis of purchases. In 2019, females accounted for 59% of all purchases in volume terms—a marginal increase on 2018 (58%). Females bought more Adult Fiction and Children's books than men in 2019, but men bought marginally more Adult Non-Fiction titles. Almost two-thirds of all ebooks were bought by women in 2019, but audiobook sales were split evenly between men and women.
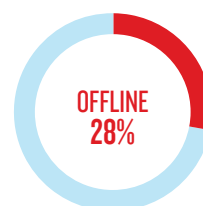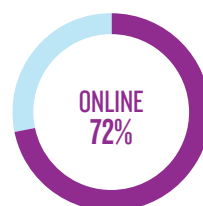
VOLUME SALES GENDER SPLITS:

ALL BOOKS
FEMALE 59%
MALE 41%

ADULT
FEMALE 63%
MALE 37%

ADULT NON-FICTION
FEMALE 48%
MALE 52%

CHILDREN'S
FEMALE 65%
MALE 35%

PRINT
FEMALE 58%
MALE 42%

EBOOKS
FEMALE 63%
MALE 37%

AUDIOBOOK
FEMALE 51%
MALE 49%

In terms of the age of book buyers in the UK, older generations tend to over-index when it comes to Adult Fiction purchases while the younger age-range over-index within the Children's sector.

VOLUME SALES BY AGE-RANGE BY GENRE:

Legend: 13-24 | 25-34 | 35-44 | 45-59 | 60-84

All books
13-24: 18%
25-34: 20%
35-44: 19%
45-59: 23%
60-84: 20%

Adult Fiction
13-24: 14%
25-34: 17%
35-44: 16%
45-59: 26%
60-84: 27%

Adult Non-Fiction
13-24: 17%
25-34: 20%
35-44: 17%
45-59: 26%
60-84: 20%

Children's
13-24: 24%
25-34: 25%
35-44: 25%
45-59: 15%
60-84: 11%
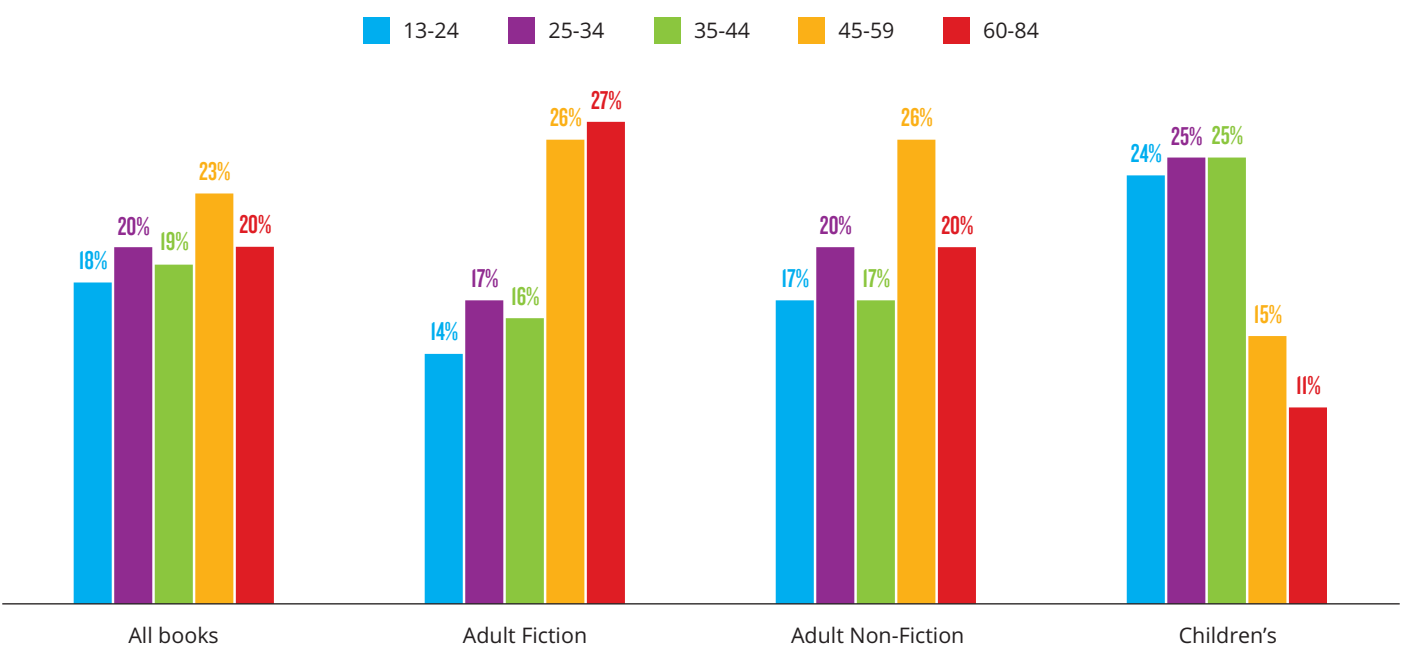
When it comes to formats, printed books tend to be enjoyed evenly across the generations. Within the ebook sector, the 13-24 age band under-indexes while the 60-84 substantially over-indexes versus the average. Within the audiobook market, the 60-84 age band accounts for just 10% of sales, but the 25-34 age band accounts for more than a quarter of the market.

VOLUME SALES BY AGE-RANGE BY FORMAT:

Legend: 13-24 | 25-34 | 35-44 | 45-59 | 60-84

All books
- 13-24: 18%
- 25-34: 20%
- 35-44: 19%
- 45-59: 23%
- 60-84: 20%

Print books
- 13-24: 20%
- 25-34: 20%
- 35-44: 19%
- 45-59: 22%
- 60-84: 19%

Ebooks
- 13-24: 10%
- 25-34: 18%
- 35-44: 17%
- 45-59: 16%
- 60-84: 28%

Audiobooks
- 13-24: 21%
- 25-34: 26%
- 35-44: 21%
- 45-59: 22%
- 60-84: 10%
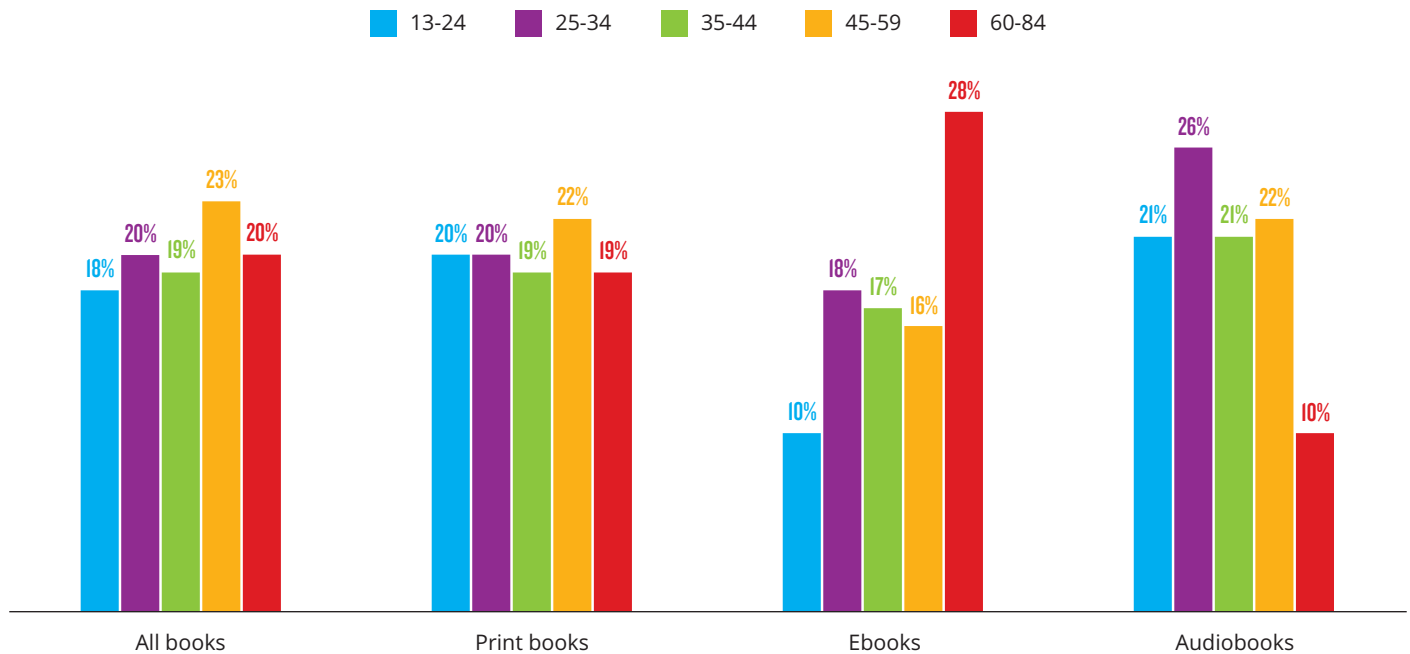
HOW DO PEOPLE DISCOVER BOOKS?

Vital in understanding the importance of metadata is an understanding of how book-buyers discover books and Nielsen Books & Consumers panel data can provide valuable insights in this area.

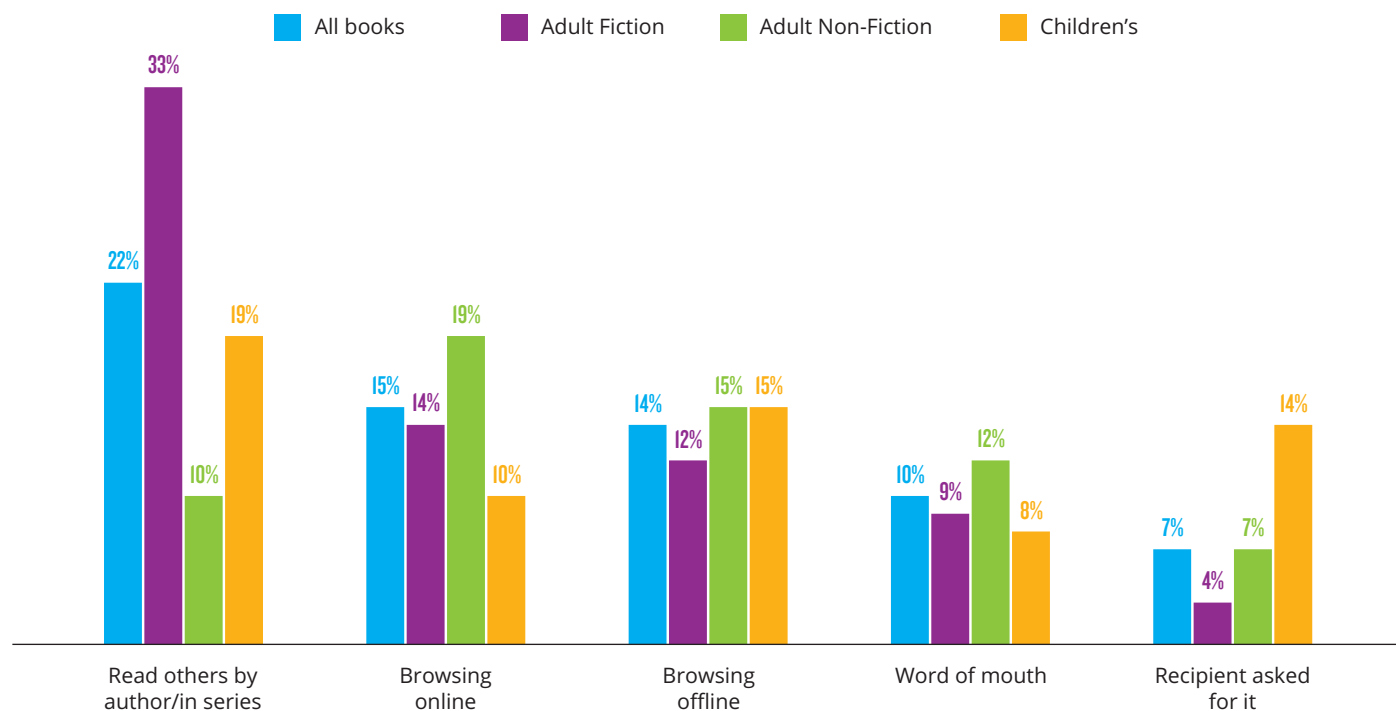There are five key areas that consumers reveal as being major influencers on how they discover books: in-store browsing, online browsing, having read books by the author and/or books in the same series before, word of mouth and because the recipient of the book has specifically requested it.

In 2019, reading the author or series before accounted for 22% of all purchases, but a third within Adult Fiction. 19% of Non-Fiction books were discovered via browsing online while "recipient asked for it" is a significant driver of discoverability within the Children's sector.

DISCOVERY METHODS FOR BOOKS BY GENRE:

■ All books   ■ Adult Fiction   ■ Adult Non-Fiction   ■ Children's

	Read others by author/in series	Browsing online	Browsing offline	Word of mouth	Recipient asked for it
All books	22%	15%	14%	10%	7%
Adult Fiction	33%	14%	12%	9%	4%
Adult Non-Fiction	10%	19%	15%	12%	7%
Children's	19%	10%	15%	8%	14%

With ebooks accounting for almost 50% of all Adult Fiction sales, and author and/or series being a major influencer on the discoverability of Adult Fiction, it is perhaps no surprise to discover that 31% of survey respondents discovered ebooks through having read the author or series before. Similarly, as one might expect, online browsing was a big driver of ebook discoverability while offline browsing was a major driver of printed book discoverability.

DISCOVERY METHODS FOR BOOKS BY FORMAT:

Legend: ■ All books ■ Print ■ Ebook ■ Audiobook

Discovery method	All books	Print	Ebook	Audiobook
Read others by author/in series	22%	17%	31%	24%
Browsing online	15%	12%	25%	19%
Browsing offline	14%	20%	2%	3%
Word of mouth	10%	9%	8%	10%
Recipient asked for it	7%	10%	2%	4%
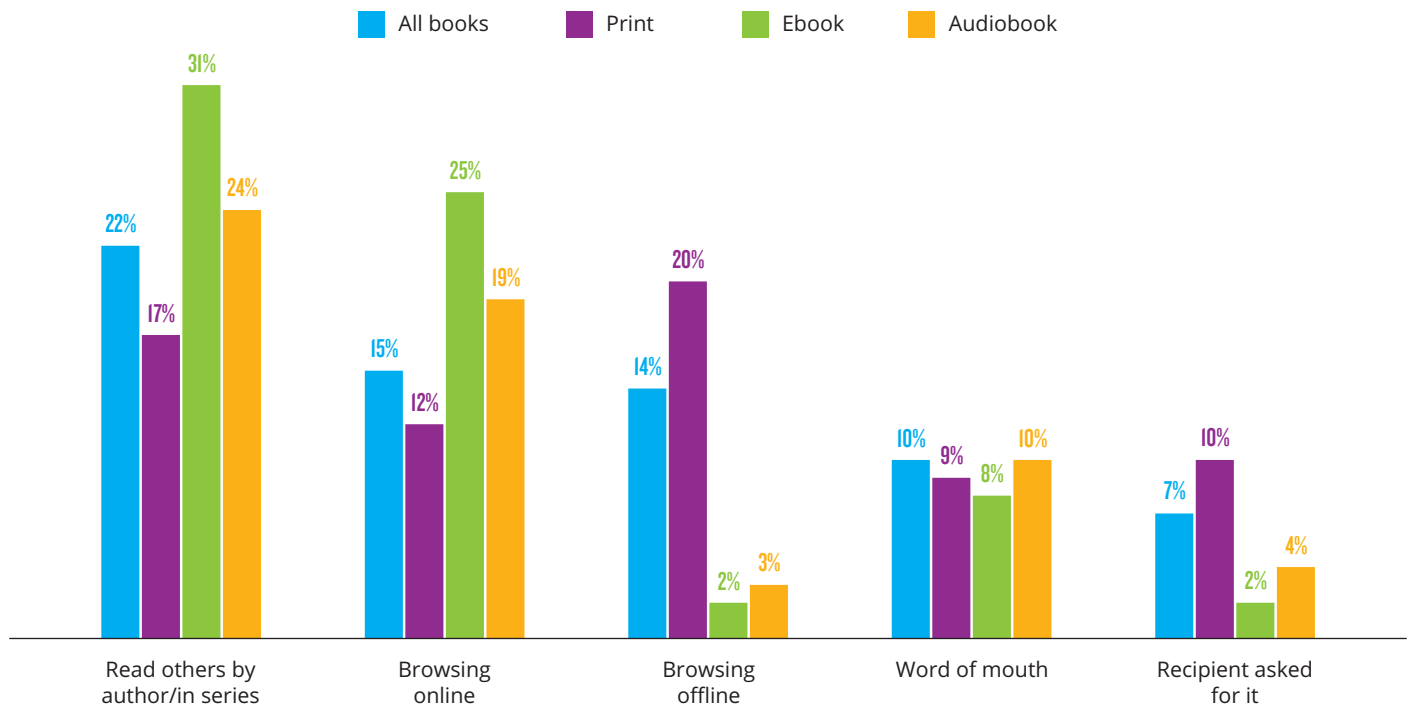
WHY DO PEOPLE BUY BOOKS?

When asked why they bought the book that they did, in 2019 consumers most often replied that they liked the author or were interested in the subject matter. The description of the book and liking the series were other main drivers within the Adult Fiction and Children's sectors.

PURCHASE INFLUENCERS FOR BOOKS BY GENRE	All books	Adult Fiction	Adult Non-Fiction	Children's books
Contained info needed	5.7	1.3	13.4	3.4
Description of the book	17.7	20.7	17.7	12.1
Front cover	11.5	11.3	9.9	14.1
Like the author	26.1	34.4	20.3	18.1
Interested in the subject	28.8	24.1	43.5	17.4
Like the main character(s)	12.6	15.8	7.1	14.1
Like the series	17.4	23.3	6.5	21.1
Low price/on offer	14.1	15.7	12.8	12.8
Read extract/looked inside first	9.5	11	9.3	6.9
Recipient asked for it/likes this kind of book	9.1	4.8	8.6	18.1
Recommendation/review	11.7	11.4	14.7	8.3

The description of the book was a big contributor to ebook sales in volume terms, underpinning the importance of ensuring good descriptive metadata is available for each and every book.

PURCHASE INFLUENCERS FOR BOOKS BY FORMAT	All books	Print	Ebook	Audiobook
Contained info needed	5.7	7.6	3.5	3.9
Description of the book	17.7	15.6	22.1	17.3
Front cover	11.5	13.8	7.6	5
Like the author	26.1	19.8	31.4	29.6
Interested in the subject	28.8	28.4	28.3	24.5
Like the main character(s)	12.6	9.9	14.3	16.6
Like the series	17.4	13.7	22.9	22.6
Low price/on offer	14.1	14.8	18.4	10.1
Read extract/looked inside first	9.5	9.5	10.5	6.3
Recipient asked for it/likes this kind of book	9.1	13.6	2.2	3.9
Recommendation/review	11.7	10.5	12.4	12.4

Based on the data above and Nielsen Books & Consumers estimates on the size of the overall book market in the UK, a book's description influenced more than 60m UK consumer book purchases last year, worth around £400m.

THE RELATIONSHIP BETWEEN METADATA AND SALES

Book Industry Communication (BIC) is the UK book industry's supply chain organisation, committed to improving the efficiency of the trade and library supply chains. BIC created and maintains the BIC Basic standards for the supply of product information and it is this "standard" which underpins the forthcoming analysis in this report.

BIC's Product Data Excellence Awards (PDEA) accreditation scheme is designed to encourage publishers to commit to providing accurate and timely product data to the book supply chain, a simple diagram of which can be seen here:

THE UK BOOK SUPPLY CHAIN

The 2020 instalment of the PDEA, recently refreshed to reflect the increasingly sophisticated data expected by retailers, libraries and consumers, lays out specific mandatory and conditional criteria for all book bibliographic data.

To conduct the study of the relationship between metadata and sales we have analysed a Total Consumer Market top 100,000 bestseller chart for the 52-week period ending 28 December 2019 against the following elements from PDEA BIC Basic scheme:

- Product Identifier (ISBN 13)
- Title Information
- Product Format
- Product Properties (size and pagination)
- Audience
- Subject
- Publisher and Imprint

- Publication Date
- Publisher Status
- Cover Image
- Sales Rights (related to the UK market)
- Supplier
- Supplier Availability
- Pricing including VAT

BIC BASIC

Our analysis reveals a comparable result to the studies conducted in 2012 and 2016: that average sales of books with complete bibliographic records sell on average more than twice the numbers of books with incomplete records. The difference in the average sale in 2019 between physical books with a complete BIC Basic record and an incomplete record was exactly 1,000 copies.

BIC BASIC: COMPLETE vs. INCOMPLETE

1,910
COMPLETE

910
INCOMPLETE AVG. SALE

▲ **+110%**
DIFFERENCE

Compliance—i.e. the number of records meeting the BIC Basic criteria—was 82% within the top 100,000, with around 7% of titles falling foul of just one criterion. Largely this related to product properties (pagination or dimension)—a new "mandatory" element of the refreshed PDEA scheme.

BIC BASIC COMPLIANCE

COMPLIANCE
82%

SALES
90%

In addition to making sure all the BIC Basic elements are available, the timely delivery of this information is also encouraged by BIC. With pre-order sales often a major influencer over a book's success, the provision of accurate data into the supply chain far in advance of a book's publication date is to be encouraged, so that booksellers, printers and distributors are able to plan ahead and consumers can discover forthcoming titles on retailer websites.

The required standard for product records to be available is 16 weeks, or 112 days, ahead of publication. Our analysis of the top 100,000 bestsellers reveals that the average sale of a book meeting this timeliness criteria was 44% greater than the average sale of a book that did not meet the timeliness criteria, or meet the criteria altogether. The relatively high average sales for non-timely publications is in part down to some high-volume, sensitive titles that dropped onto the publication schedules relatively late, plus some strong-selling retailer-exclusive licensed products that also often lack numerous metadata elements.

BIC BASIC TIMELINESS

1,965
PASS

1,360
FAIL AVG. SALE

▲ **+44%**
DIFFERENCE

At the very top end of the 2019 bestseller chart, just one title failed to meet BIC Basic criteria, falling foul for the minor infringement that one piece of physical dimension data was missing. Three other books in the top 200 failed to meet just a single element: one also lacked a piece of physical dimension information; one had no rights information and the other lacked a cover image.

COVER IMAGE

1,740
YES

895
NO AVG. SALE

▲ **+94%**
DIFFERENCE

99% of the top 100,000 bestsellers of 2019 had a jacket image, with the average sales of books with cover images almost double the sales of those without.

BIC CODE

1,730
YES AVG. SALE

470
NO AVG. SALE

▲ **+94%**
DIFFERENCE

One of the biggest influencers on a book's average sale in 2019 from a bibliographic perspective was BIC Code—simply, whether the book had a subject classification record. As we have seen, online browsing is a significant influence on book discoverability so ensuring a book has an accurate genre classification record is essential.

We can see a high compliance here with 99% of records having at least one BIC Code record, while the average sale of a book with at least one BIC Code was more than three times the sales of books without. Just one book without a classification record sold more than 10,000 copies in the UK in 2019.

DESCRIPTIVE METADATA

Although not mandatory in the PDEA scheme at BIC Basic level, a book's description is a major purchase influence. Books & Consumers survey data reveals that in 2019, 17.7% of book-buyers cited "description of the book" as a reason behind their purchase behaviour—greater than a book's front cover (11.5%), whether the book has won an award (2%) or appeared in a bestseller chart (7%), and even whether it was priced low or on offer (14.1%).

There are five key elements of descriptive metadata: keywords, short description, long description, author biography and reviews.

Keywords supplement the standard metadata elements such as title, author, subject, etc., and are words and short phrases that relate to a book's themes or content—the kinds of words and phrases that one might expect consumers would use when searching for a book without having a specific title or author in mind.

On average, books in the top 100,000 chart for 2019 with keyword records sold more than 2,300 copies on average, compared to under 1,000 copies for books without. Compliance was around 57%.

KEYWORDS

2,305	955	▲ +140%
YES AVG. SALE	NO AVG. SALE	DIFFERENCE

Short Description is a 350-character (max) blurb used to describe the content of the book but, interestingly, there was little difference in the average sales of books with it and books without it. A look into the underlying data reveals that many retailers utilise **Long Description** in preference to the Short Description metadata element, only using the latter in exceptional circumstances. A total of 43 different books that sold more than 100,000 copies did not have any "short description". However, all but two of the 43 titles did have a "long description". Overall, within the top 100,000, compliance on "short description" was lower (70%) than compliance on "long description" (83%). Books with a "long description" enjoyed considerably higher average sales than those without.

SHORT DESCRIPTION

1,745	1,685	▲ +4%
YES AVG. SALE	NO AVG. SALE	DIFFERENCE

LONG DESCRIPTION

1,930	790	▲ +144%
YES AVG. SALE	NO AVG. SALE	DIFFERENCE

AUTHOR BIOGRAPHY

2,010
YES AVG. SALE

1,019
NO AVG. SALE

▲ **+97%**
DIFFERENCE

Where author data was available, compliance on **Author Biography** was 73%. The average sale of a book in the 2019 top 100,000 with complete author biography information was almost double the sales of books with an author but no author biography.

REVIEW

**2,380**
YES AVG. SALE

1,300
NO AVG. SALE

▲ **+83%**
DIFFERENCE

Meanwhile, books with **Review** information enjoyed sales that were 83% higher on average than books without.

Broadly speaking, the more descriptive elements a book has the higher its average sales. The average sale of a book in the 2019 top 100,000 with zero descriptive elements was just 756 copies, but books with all five descriptive metadata elements enjoyed average sales more than three times higher (2,392 copies sold).

AVERAGE SALE BY NUMBER OF DESCRIPTIVE METADATA ELEMENTS

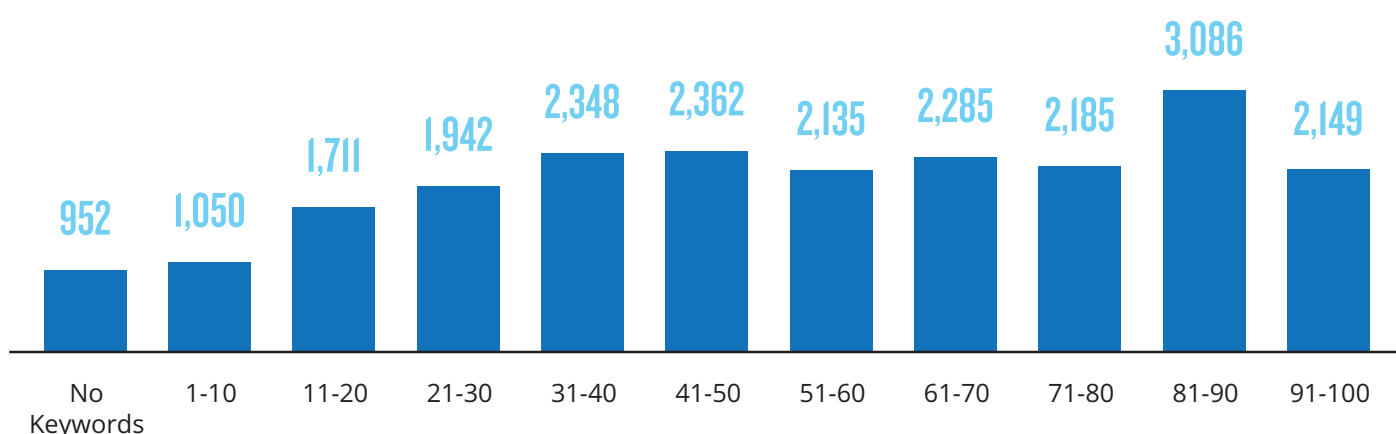| Elements	Average Sale
0	756
1	1,487
2	1,434
3	1,519
4	1,901
5	2,392
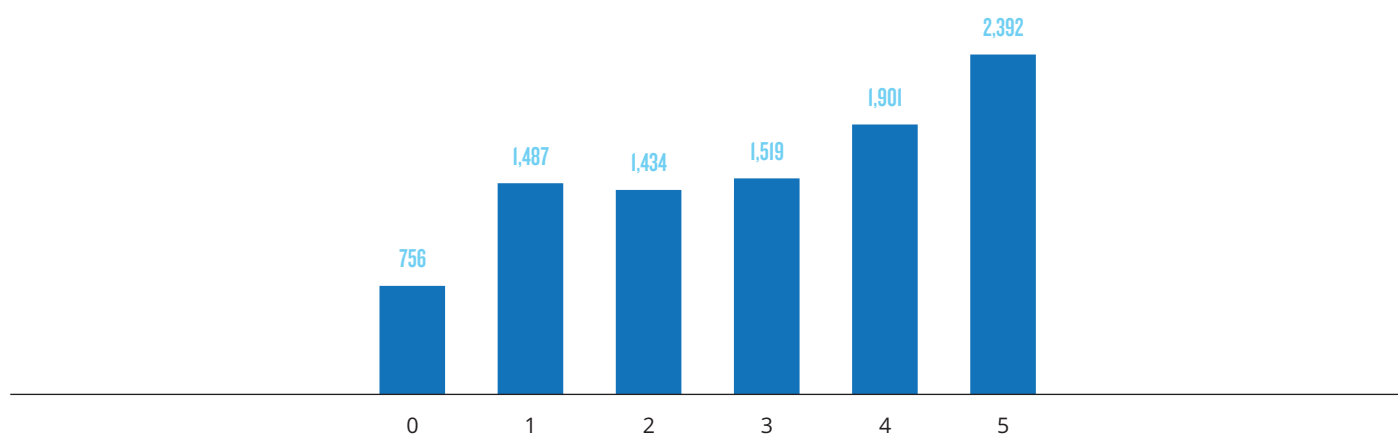
KEYWORDS

A new area of analysis for 2020 is an attempt to discover whether or not there is any relationship between the number of words in a book's "keyword" or "long description" text and its sales. For the purposes of the analysis any set/grouping of less than 1,000 data points (ISBNs) was excluded in an attempt to reduce the influence of outliers as much as possible.

Our analysis reveals that, generally, books with 30+ words in their keyword element (note that words in a phrase are counted individually) enjoy significantly higher average sales than books with fewer than 30 words in their keyword element. Books in the 2019 top 100,000 report with no keywords at all sold fewer than 1,000 copies on average.

The underlying data reveals that many publishers are taking a creative but no doubt informed approach to keywords for their bestselling titles. The findings show that it is possible for publishers to utilise the keyword field's potential not just as a descriptor of an individual book's content but as a way to infiltrate the search results of consumers looking for similar titles or authors. So, for example, the keywords belonging to a book by a popular crime writer might include "thriller", "murder", "crime", and "death" alongside a selection of other key authors for that genre. Similarly, a cookbook by a television star may include keywords such as "cookbook", "menu", "recipe" and "dinner" alongside keywords to other well-known cooking personalities.

AVERAGE SALE BY NUMBER OF WORDS IN KEYWORD METADATA ELEMENT

No Keywords	1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100
952	1,050	1,711	1,942	2,348	2,362	2,135	2,285	2,185	3,086	2,149

The general trend where the long description is concerned is also: the more the better, and certainly more than 200. We have capped the graph below at 500 words in the long description as data points beyond this point reduce and become skewed by outliers.

AVERAGE SALE BY NUMBER OF WORDS IN LONG DESCRIPTION METADATA ELEMENT

0	1-100	101-200	201-300	301-400	401-500
790	1,361	1,617	2,538	3,201	3,336
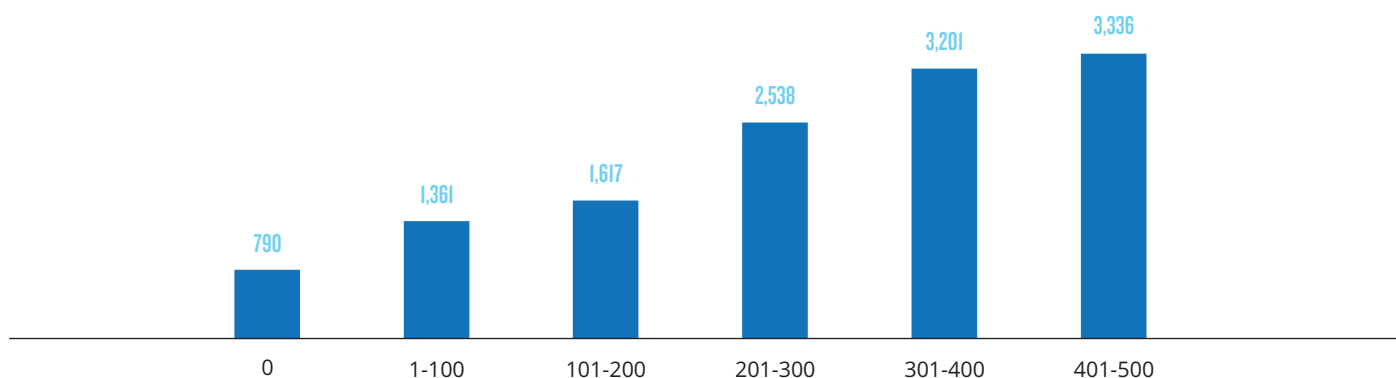
IN SUMMARY

As we have seen in previous White Papers, our latest report illustrates that books with accurate, complete and descriptive metadata sell more copies on average than books with less complete metadata. It is difficult to measure direct causation as different retailers display different metadata elements in different ways and, short of interviewing every book buyer at the point of purchase, there is no way to know what influence each specific metadata element may have on book sales in the UK. However, we hope we have shown through slicing and dicing the bibliographic and sales data in numerous ways, and bringing in Books & Consumers research data on consumer purchase behaviour, that there are solid indicators for the link between metadata and sales.

Our key insights include:

- 15% of all books bought in the UK in 2019 were discovered via online browsing; 14% via offline browsing.

- A book's description influenced 18% of purchases across all formats in 2019; rising to more than 20% for Adult Fiction purchases and more than 22% for ebooks.

- Titles that met the BIC Basic requirement standard enjoyed average sales more than twice as high as books that did not meet the standard.

- Books with a cover image sold almost twice as many copies as those that did not have a cover image.

- Books with BIC category data enjoyed average sales that were 270% higher than those that did not.

- Books with keyword information enjoyed average sales that were 140% higher than those that did not.

- Books with five descriptive data elements on average experienced sales that were three times higher than titles with no descriptive data elements.

SUPPLYING METADATA

HOW TO SUPPLY METADATA TO NIELSEN BOOK

Nielsen Book collects metadata from over 60,000 publishers worldwide and we currently hold over 42m titles on our database. We disseminate this title information to over 3,500 booksellers, internet retailers and library customers in over 100 countries, making your titles discoverable to thousands of potential book buyers.

We have two levels of services we offer for supplying your metadata:

Our free Basic Listing Service allows you to include the bibliographic data required to meet the minimum Book Industry Communication (BIC) Basic Standard. We recommend publishers use our Title Editor service to add new titles, view and edit information and add cover images. The site is free to use and is the quickest way of supplying data to our database.

Our Enhanced Service allows publishers to add, where appropriate, enriched metadata information including descriptions, table of contents, reviews, promotional information and author biographies. It is an annual subscription service based on the number of titles published and a minimum charge applies.

For more information:

Basic Listing Service:
General enquiries: pubhelp.book@nielsen.com
Register for Title Editor: www.nielsentitleeditor.com/titleeditor/

Enhanced Service:
More information and pricing details: www.nielsenisbnstore.com/Home/EnhancedServices
General enquiries: sales.book@nielsen.com

ABOUT NIELSEN BOOK

Nielsen Book provides a range of services to the book industry internationally, aiding the discovery and purchase, distribution and sales measurement of books. We are proud to run the ISBN and SAN Agencies for UK & Ireland as well as providing search and discovery services for booksellers and libraries. Our electronic trading solutions, including Nielsen PubEasy, help everyone involved in the book supply chain trade more easily and our Research services provide retail sales analysis for both print and ebooks alongside research from our Books and Consumers Survey. For publishers we offer a range of services from assigning an ISBN to your book to adding your metadata to our database and promotional tools to help market your book. If you would like to know more visit: www.nielsenbook.co.uk

ABOUT NIELSEN

Nielsen Holdings plc (NYSE: NLSN) is a global measurement and data analytics company that provides the most complete and trusted view available of consumers and markets worldwide. Nielsen is divided into two business units. Nielsen Global Media, the arbiter of truth for media markets, provides media and advertising industries with unbiased and reliable metrics that create a shared understanding of the industry required for markets to function. Nielsen Global Connect provides consumer packaged goods manufacturers and retailers with accurate, actionable information and insights and a complete picture of the complex and changing marketplace that companies need to innovate and grow.

Our approach marries proprietary Nielsen data with other data sources to help clients around the world understand what's happening now, what's happening next, and how to best act on this knowledge.

An S&P 500 company, Nielsen has operations in over 100 countries, covering more than 90% of the world's population. For more information, visit: www.nielsen.com