NIelsen Book
UK Study: The
Importance of
Metadata for
Discoverability
And Sales
NIELSEN BOOK UK STUDY: THE IMPORTANCE OF METADATA FOR DISCOVERABILITY AND SALES

AUTHOR:

David Walter, Senior Director, Client Solutions

David heads up Nielsen Book’s Research and Commerce Solutions business in North America, including products such as Nielsen BookScan, Nielsen PubTrack Digital, Nielsen PubEasy and Nielsen Pubnet. David also takes the lead on Nielsen Book’s metadata products in North America via its range of Discovery and Information services.

CONTRIBUTORS:

Andre Breedt, Director Nielsen Book Research International
Clive Herbert, Head of Publisher Services, Nielsen Book Discovery Services
Mo Siewcharran, Director of Marketing Communications, Nielsen Book

ABOUT NIELSEN


ISBN: 978-1-910284-30-8
Published in the UK 30 November 2016
Printed in the UK by Ingram Content Group
INTRODUCTION

Nielsen Book first conducted analysis on the link between book sales and bibliographic metadata in 2012. The results of that White Paper, *The Link Between Metadata and Sales*, illustrated a strong link between the completeness of the appropriate metadata and the resultant sales. Providing complete and appropriate metadata aids the tradability and discoverability of titles – and our previous analysis added some quantitative measures to back up this notion.

When we talk about ‘tradability’ we are referring to the ease with which products can be identified and traded, and move through the book supply chain. The book trade has some unique complexities. Many of these arise from the fact that there are millions of individual, separately tradable products available in the global market at any one time, potentially being supplied by tens of thousands of different publishers. In the UK market 1.5 million different books were recorded as having sales in the 12 month period covered by this study (July 2015 to June 2016). A single book shop may carry tens of thousands of titles, and is likely to hold only one, or a few copies of many of those titles. This means that ordering and stock replenishment in the book trade, with the exception of bestsellers and new releases, is generally on a little and often basis.

Add to this the traditional sale-or-return model between publishers and booksellers, and the flow of a huge range of products to, and sometimes back from, retailers quickly grows to significant complexity. These factors mean that creating a sustainable supply chain for the book trade needs attention, planning and cooperation between all parties. The ISBN (International Standard Book Number) provides the foundational key for many of the book trade’s supply chain efficiencies, accurately identifying a unique item, for which a record can be created listing key attributes. Industry bodies such as EDItEUR (The trade standards body for the global book, e-book and serials supply chains), BIC (Book Industry Communication) and BISG (Book Industry Study Group) have developed further standards and formats for the provision of data, such as ONIX, the accompanying code lists, and classification schemes. Providing accurate data on properties such as publication date, price, supplier and physical attributes aids booksellers in planning their stock management, from scheduling future orders, to planning shelf space or storage allocations, to...
ensuring shipments are made on the most economical terms (through referencing physical attribute data). Maintaining an efficient supply chain ensures that booksellers can focus on selling books – and maximizing sales for publishers and themselves. Where this valuable supply chain data isn’t available to the bookseller, at best they will need to carry out additional work (leading to decreased efficiency) and at worst they may not order the product due to an inability to plan for it effectively.

Discoverability has been somewhat of a buzz word in the book industry for several years now. In essence, the quality of discoverability is the ease with which a particular product can be found. This can either relate to trading partners within the book trade, or end customers purchasing a title – to booksellers or libraries searching for titles to stock, or consumers searching on a website and relying on the metadata available. It can relate to the discovery of a specific title, where the individual searching knows what they are looking for and needs to find the appropriate information or product record; or where an individual is using more general criteria to browse, then identify a title that meets their needs or taste.

Both of these qualities, the ease with which books can be discovered and the ease with which they can be traded, rely heavily on the provision of appropriate, accurate and timely metadata.
HOW DO CONSUMERS DISCOVER BOOKS?

Based on Nielsen’s Books and Consumers™ data from 2015, seeing a book while browsing/searching in a library, shop or catalogue, on a website or device, was still the top discovery method as in previous years. Books discovered through this method accounted for 29% of all purchases - nearly half of these were books discovered in a bookshop, with just over a third (37%) being via a book website, and ‘other online’ accounting for most of the rest. Browsing therefore remains a key path to purchase for book buyers – and relies on the appropriate metadata being available to support online browsing, discovery and purchase.

Awareness of the author or series was the next most common way of discovering books – for nearly one in five purchases – with recommendation/review not far behind. Of books discovered through recommendation/review, nearly half (44%) were down to word of mouth recommendation, with discovery online accounting for around two in five of these purchases, and newspaper/magazine reviews or recommendations around one in ten. Once again we see that information available online, such as a review, plays a key role in book sales – and as we will see in our study, an increase in the number of descriptive metadata elements attached to a title correlates with higher resulting sales.

DELIVERING AND MAINTAINING DATA

Delivering and maintaining the appropriate metadata takes constant attention, focus and effort. Revisiting our earlier study with up to date data provides an opportunity to focus again on the importance of metadata for the book trade. Areas previously covered that we will examine again include:

- The provision of a set of basic metadata elements.
- The provision of descriptive metadata elements.
- How beginning to add descriptive metadata elements translates into increased sales.

As well as refreshing our study with more current data, we have added some further elements to our analysis this time. These include:

- The timeliness of data delivery.
- An overview of library borrowings and how these correlate with product metadata.
Some caveats: though Nielsen Book’s data is used widely within the UK book trade, not all retailers or libraries use Nielsen Book as their data source. Therefore we cannot draw a direct line between the data we have used for this study and the data used by all retailers. However, Nielsen Book’s data is likely to represent a good measure of the best level of metadata available in the UK book trade.

Another limitation is that the metadata we have used is only a snapshot, taken just after the period of the sales we refer to in the study. Titles published at the start of the 12-month period (i.e. July 2015) may have had inadequate metadata at the start of their lifespan, which has subsequently been improved before we have taken our snapshot of the data. If anything, the consequence of this is that we are understating the extent of the link between complete metadata and sales. We have attempted to partially counteract this by also reporting on the BIC Basic and ONIX Compliance timeliness values in our data, which measure the ‘when’ of metadata supply, on top of the ‘what’.

**OUR APPROACH AND DATA**

Nielsen Book has access to both bibliographic data (through our book database) and sales and borrowing data (through our Nielsen BookScan Total Consumer Market and LibScan public library borrowing data). Nielsen Book aggregates data from over 60,000 suppliers in the UK and overseas to create an extensive and comprehensive source of bibliographic data. We also measure retail sales for over 90% of the UK book market, and library borrowing data from 70 UK public library authorities.

To conduct this study we have focused on the top 100,000 selling titles over a 12 month period (July 2015 to June 2016) which represent approximately 93% of total book sales for that time period. To examine the link between metadata and library borrowing, we have similarly used data for the top 100,000 most borrowed titles for this time period. Extracting and analysing the metadata for those titles allows us to identify the correlation between metadata and sales or library borrowing at a high level. Our key measure is average sales or borrowings per ISBN – we are not looking at absolute numbers, rather grouping titles which have a similar level of metadata completeness, and comparing these to other groups using the average sales per ISBN as a measure.
BIC PRODUCT DATA STANDARDS

BIC (Book Industry Communication) is the UK book industry’s independent supply chain organisation, committed to improving the efficiency of the trade and library supply chain. As part of their work BIC developed and maintains the BIC Basic and BIC Excellence standards (commonly referred to as ONIX compliance) for the supply of product information. Each standard has a related set of data elements and it is against these data sets that we are measuring the completeness and timeliness of metadata in this first part of our study.

These data element sets comprise part of the BIC Product Data Excellence Award Scheme. The scheme is designed to encourage publishers to appreciate the contribution made by accurate and timely product data to their sales; to enable them to demonstrate their commitment to making that data available in the supply chain; and to enable them to benchmark their performance against their competitors. Publishers can seek an award under this accreditation scheme.

Full details of the BIC Product Data Award Scheme together with a list of the required data elements are available on the BIC website.

BIC BASIC

The BIC Basic standard remains a key benchmark for the supply of bibliographic data. The BIC Basic elements include data that is essential to identify and trade the product, as well as a cover image. Specifically, the BIC Basic elements are:

- ISBN
- Title
- Product Form
- Main BIC subject category
- Imprint name
- Publication date
- Cover image
- At least one supplier name
- Availability status
- GBP retail price including VAT
- Statement of rights relating to the UK

Looking at the results from our most recent data set, we see similar results to our earlier study. The graph below shows all titles grouped into those that hold all of the necessary BIC Basic data elements or not, and those which hold a cover image or not.
We can clearly see the importance of supplying a cover image – even for titles that do not hold all of the required BIC Basic data elements, the presence of a cover image correlates with average sales over twice as high. However, providing both a complete set of BIC Basic data elements and an image correlates with an even higher average sales per ISBN value.

As an overview and to what extent these criteria are met within our data set, the chart below shows what proportion of titles fall into each of these categories. Whilst a healthy majority of titles do meet the BIC Basic standard and carry a cover image, a not inconsiderable percentage (15%) fall short of even these basic requirements. It is worth remembering that our data set uses the top 100,000 selling titles, so we are not referring just to long tail publications here – these are active titles in the market.
BIC BASIC - TIMELINESS

In addition to measuring what data elements are present, meeting the BIC Basic standard also mandates that data should be supplied sufficiently far in advance of the publication date. The key motivation behind supplying data in a timely fashion is ensuring that booksellers have access to the data they need to discover, order and manage forthcoming titles so that they can be made available to the market effectively upon publication. Providing sufficient metadata well in advance of publication also ensures that the relevant consumer facing data on internet retail sites will be available. Titles typically enjoy their strongest sales in the weeks immediately following publication, when marketing and promotion activities are at their height. Failing to maximise sales in this early period can mean missing an opportunity which cannot be subsequently compensated for.

The required standard is for all BIC Basic data elements to be present on the product record 16 weeks in advance of the publication date (112 days). As this is measured only once, at the time a record first becomes compliant with the BIC Basic standard, our data is not subject to any changes made to the data subsequently – we therefore have a true measure of timeliness to analyse.

The chart below shows how sales are affected for titles where the BIC Basic data and an image are present when this data is also provided in a timely fashion – meeting the 112 days before publication standard. Timeliness correlates to another increase in the average sales per ISBN for those titles, of approximately 6%. Based on our data set around 70% of titles that met the BIC Basic standard also met the timeliness requirement.

Fig. 2.1 Average sales per ISBN for records with complete BIC Basic data, with and without a Cover Image, and those which also meet the BIC Basic timeliness requirement
A further measure of the quality of bibliographic metadata is the ONIX Compliance criteria. In addition to the BIC Basic elements, ONIX Compliance requires that data is supplied in a valid ONIX format, and includes at least one descriptive data element for the title record. As with the BIC Basic standard, ONIX Compliance includes a measurement of timeliness, again set at 16 weeks, or 112 days, ahead of publication.

The chart below shows the average sales per ISBN for both ONIX Compliant and non-compliant titles, and also the affect that the timely supply of this data has. ISBNs that meet the ONIX Compliance requirement show average sales over 100% higher than those that don’t. For ISBNs also meeting the timeliness requirement, average sales are a further 25% higher.

Approaching our data through the lens of the BIC Basic and ONIX standards and timeliness requirements shows a clear progression in average sales per ISBN as the completeness and timeliness of the data improves. Whilst the majority of titles in our data set do meet the requirements, there are a significant number that don’t – wherein lie opportunities for improvement.
EFFECTS ACROSS BROAD GENRES

Looking across broad genres, we can see some differences in the extent to which having complete and timely data correlates with sales. Using the four broad Product Class definitions from our Nielsen BookScan data we can see similar results to our previous study – this time with the added metric of data timeliness. As shown in the graph below, Fiction sees the strongest positive impact of complete metadata, followed by Trade Non-Fiction, Children’s then Specialist Non-Fiction. Adding the timeliness measure shows once again a positive correlation between this additional factor and higher sales.

![Graph showing average sales per ISBN across broad genres](image)

**Fig. 4.1** Average sales per ISBN across broad genres for records which do not meet the BIC Basic requirement or hold an image, those which do meet the BIC Basic requirement and hold an image, and those which also meet the BIC Basic timeliness requirement.
Similarly, looking at ONIX Compliance and timeliness shows the strongest growth in the correlation between metadata and sales for Fiction – and once more the timeliness element correlates with the maximum average sales per ISBN.

Fig. 4.2 Average sales per ISBN across broad genres for records which are not ONIX Compliant, those which are ONIX Compliant, and those which also meet the ONIX Compliance timeliness requirement.
DESCRIPTIVE METADATA

The ONIX Compliance standard requires the presence of at least one descriptive data element, additional descriptive elements are valuable for most titles, and add to the completeness and richness of the data. This translates into increased discoverability both for book trade buyers and consumers.

The graph below shows the average sales per ISBN for title records with varying numbers of descriptive metadata elements. For this part of our analysis we have used only those records from subscribers to the Nielsen BookData Enhanced Service, where these descriptive data fields are output in our data feeds to retailers, libraries and other parties. Titles from publishers who do not subscribe to the Nielsen BookData Enhanced Service do not have their descriptive data output, and therefore it would not be appropriate to include them in our comparison here. We have also used four out of the five main descriptive elements in our analysis – these are the short description, long description, author biography and reviews. We do hold data on the table of contents where this is present – however, given that it is only appropriate for certain titles to hold a table of contents (i.e. academic and other non-fiction works, anthologies) and the relatively low number of these present in our data set, we have not included this in our analysis.

Whilst there is not a clear linear progression of increasing sales as more metadata elements are added, what is clear is that the titles that have four descriptive metadata elements see the highest average sales per ISBN.

Fig. 5.1 Average sales per ISBN for records holding zero to four descriptive data elements – short description, long description, author biography and review
Drilling down into the titles with no descriptive data elements helps to explain why we see the anomalous result above. There are only 1,019 records in our data set that carry no descriptive data elements. The best-selling of these are children’s annuals and museum guidebooks. For the museum guidebooks, sales will be driven by footfall on site at the museum, so the metadata available (or lack thereof) for those titles is of little relevance to the sales outcome. We can interpret the relatively high sales of the children’s annuals, which hold no descriptive metadata, to the power of the characters and brands involved, and the seasonal nature of these titles, whereby their presence at retailers drives unplanned gift purchases. These factors, combined with the low number of titles in this set, skew our figures, leading to the unexpectedly high average sales per ISBN for titles holding no descriptive metadata elements.

Looking at the proportion of titles within our data set with the different levels of descriptive data, we see the percentages below. A huge majority of records carry two or more descriptive data elements.

Looking at how the presence of descriptive data elements correlates with sales across different genres, Fiction shows the greatest differences between titles with varying levels of descriptive data. We again see the anomalies of titles with no descriptive elements outperforming those with one descriptive element present for Fiction and Trade Non-Fiction – however, these averages are based on very small sample sizes of 1% or less of the titles in those categories, so a handful of outlier titles are driving these results.

Fig. 5.2 Proportion of records holding zero to four descriptive metadata elements
Fig. 5.3 Average sales per ISBN across broad genres for records with zero to four descriptive metadata elements

ADDING ENHANCED METADATA TO TITLES

As a further measure of the effect of including descriptive metadata in product records, we have analysed sales for publishers who previously did not use our Nielsen BookData Enhanced service but took up a subscription between July 2015 and June 2016. This gives us the opportunity to review publishers whose title records did not include descriptive data, but have subsequently added it.

The graph below shows the annual sales for these publishers in the year previous to their including descriptive data, and the year when they began including this data. The increase in total sales for these publishers in the year they begin including descriptive data is 13%. As some of these publishers may have been newly established around the period of our data, we have also reviewed publishers who were selling at least 100 copies in the year prior to adding descriptive data. With this additional filter, we see an increase in total sales of 10% year-on-year.
Nielsen Book also run a measurement service for the UK public library sector, Nielsen LibScan. Running on similar principles to Nielsen BookScan, LibScan aggregates borrowing data from a panel of participating libraries to provide borrowing data for the overall market being measured. Nielsen Book’s bibliographic data is also used throughout the library sector so we can carry out analysis based on Nielsen LibScan and our bibliographic data - rather than investigating average sales per ISBN, for the library sector we will review average borrowings per ISBN.

Analysing this data based on the BIC Basic and Timeliness requirements we used previously, we get the results shown in the chart below. There is a similar pattern for library borrowings as was seen for retail sales. Meeting the BIC Basic requirements and including a cover image correlate with higher sales than when these criteria are not met. Meeting the timeliness requirement correlates with further increased borrowings for titles meeting the BIC Basic requirements and carrying a cover image. We see an anomalous result for the timeliness requirement for titles which meet the BIC Basic requirements but do not carry an image – only a small number of records fall into this category (584) which is the likely cause of this result.
Analysing the correlation between library borrowings and the number of descriptive data fields present again echoes what we found when analysing sales. The chart below shows the average number of borrowings per ISBN for titles with zero descriptive metadata elements, up to those titles carrying all four descriptive elements. Titles with all four descriptive metadata elements show average borrowings 167% higher than those with no descriptive elements.

fig. 7.1 Average library borrowings per ISBN for records with incomplete or complete BIC Basic data, with and without a Cover Image, and those which also meet the BIC Basic timeliness requirement

Analysing the correlation between library borrowings and the number of descriptive data fields present again echoes what we found when analysing sales. The chart below shows the average number of borrowings per ISBN for titles with zero descriptive metadata elements, up to those titles carrying all four descriptive elements. Titles with all four descriptive metadata elements show average borrowings 167% higher than those with no descriptive elements.

fig. 7.2 Average library borrowings per ISBN for records holding zero to four descriptive data elements – short description, long description, author biography and review
SUMMARY

Revisiting our previous study has shown that our original findings still hold true – that titles with complete, appropriate and descriptive metadata sell more copies, on average, than titles with less complete metadata. Additionally, we have seen a correlation between the timely supply of basic and descriptive data with increased sales, and a positive correlation between the completeness of metadata and the level of library borrowings.

We will repeat our caveat – that we cannot measure direct causation, but just identify the correlation between the presence of metadata and sales. However, as we see this correlation when using different measures of a title’s metadata, and segmenting the data in several different ways, we believe that there are clear indicators for the link between metadata and sales.

KEY FINDINGS INCLUDE:

• Titles meeting the BIC Basic requirements and carrying a cover image see average sales over twice as high as those of titles that do not meet that standard.
• Titles which also meet the BIC Timeliness requirement see average sales a further 6% higher.
• Titles meeting the ONIX compliant requirement see average sales over 100% higher than titles do not meet that standard.
• Titles which meet the ONIX Compliance timeliness requirement see average sales a further 25% higher.
• Titles which carry all four descriptive metadata elements used in our analysis see average sales 65% higher than titles which don’t carry any descriptive elements.
• Publishers that begin to add and disseminate descriptive data for their titles via the Nielsen BookData Enhanced service saw sales grow over 10% year-on-year.
• Average library borrowings per ISBN similarly correlate positively with the completeness and timeliness of BIC Basic data, and the extent of descriptive data.

Whilst many titles do show best practice in meeting the various measures we have used of metadata quality, there are still a significant proportion of titles that fall short of this. There is still, therefore, an opportunity to make a positive impact on the tradability and discoverability of titles – to fully exploit supply chain efficiencies, and to maximise sales.
NOTES/KEY:

1ONIX for Books is a standard of XML message which is used for representing and communicating book industry product information in electronic form. ONIX for Books was developed by EDItEUR (www.editeur.org) jointly with BIC (www.bic.org.uk) and BISG (www.bisg.org), and is now maintained under the guidance of a broad international steering committee.


3Nielsen Book collates data from thousands of sources (in over 70 countries) – publishers, distributors, wholesalers and other data aggregators, for example – to compile a comprehensive database of English-language titles. The database currently holds over 27m records.

4The Nielsen BookScan UK Total Consumer Market (TCM) aggregates sales from approximately 6,500 book retailers in the UK, including high street chains, independent bookshops and internet retailers. Point of sale data is collected and processed on a weekly basis with checks and balances in place to ensure that the TCM is an accurate, objective and independent measure of book sales. As Nielsen BookScan aims to measure sales to end consumers, institutional and bulk sales are not included in the UK TCM figures.

5Nielsen LibScan aggregates data from 70 public library authorities. Using similar methodology to BookScan. Participation in the panel is free and data is collected and processed 4-weekly and combined to produce an overall picture of library borrowings across the whole panel.

6There are some titles within the top 100,000 sellers from Nielsen BookScan for which we either do not hold data, do not output data, or that are exclusive to one retailer – we have not included these records in our analysis. We have also limited our data set to book product only – ISBNs may be used on other book related products which would not carry the same types of metadata as books. This reduces the total number of records we have used for our analysis to 91,741.

7More specifically, the sales data used is from 12 July 2015 to 16 July 2016. This equates to Nielsen BookScan week 29 of 2015 to week 28 of 2016, or period 8 of 2015 to period 7 of 2016.

8Contact details for BIC can be found at www.bic.org.uk

9The Nielsen BookScan Product Class classification is a proprietary system that places an emphasis on simplicity and usability when analysing sales data. An individual title may carry several BIC codes – but for the purposes of analysis it is clearer that each title only belongs to one chart or segment, rather than appearing multiple times.

10Our data for this part of the analysis is therefore a subset of 77,223 records from subscribing publishers, from our initial data set of 100,000.

118,557 records from our data set carried a table of contents.

Sponsors: our thanks to our sponsors for supporting this report which is an essential tool for the book industry. The Study provides evidence to support the strong belief that good data helps to promote and sell books. There is undoubtedly an underlying link between the provision of good metadata and book sales, and the Nielsen Book UK Study: The Importance of Metadata for Discoverability and Sales will be used, as the previous edition, extensively to promote metadata provision and best practice for suppliers of bibliographic data in the UK and internationally.

ABOUT BIC:
BIC is an independent, not for profit members’ organisation working at the heart of the UK book industry. Established in 1991, it is sponsored by The Booksellers Association, The British Library, The Chartered Institute of Library and Information Professionals and The Publishers Association to promote supply chain efficiency in all sectors of the book world through e-commerce, best practice, training, events, and the application of standard processes and procedures. BIC is governed by an Executive and an Operational Board, both of which are comprised of members across the entire book industry. Visit BIC at www.bic.org.uk and follow at @BIC1UK.

ABOUT BOOKSONiX:
BooksoniX is a cloud based publishing management service which automates and streamlines ONIX data delivery to multiple recipients worldwide. BooksoniX can work with ONIX 3, ONIX 2.1 or spread sheets and is renowned for high quality data and many data aggregators recommend BooksoniX.

BooksoniX is easy to use; you can create your own company central bibliographic database which is available 24/7. You can instantly produce AI sheets, price lists, catalogues, etc. You can tracks foreign rights, contributors/contacts and include production/workflow management and email alerting. Available options include website integration and royalty modules. BooksoniX services are vital resources for publishers, distributors and sales agents on five continents. For more information, visit www.booksonix.com

ABOUT BAKER & TAYLOR:
Baker & Taylor is the premier worldwide distributor of books, digital content and entertainment products from approximately 25,000 suppliers to over 20,000 customers in 120 countries. The company offers cutting-edge digital media services and innovative technology platforms to thousands of publishers, libraries, schools and retailers worldwide. Baker & Taylor also offers industry leading customized library services and retail merchandising solutions. For more information, visit www.baker-taylor.com

ABOUT INGRAM CONTENT GROUP
Ingram Content Group is a trusted leader in providing comprehensive, global content distribution services to publishers, retailers, libraries and educational institutions.
ABOUT NIelsen

Nielsen Holdings plc (NYSE: NLSN) is a global performance management company that provides a comprehensive understanding of what consumers watch and buy. Nielsen’s Watch segment provides media and advertising clients with Total Audience measurement services for all devices on which content — video, audio and text — is consumed. The Buy segment offers consumer packaged goods manufacturers and retailers the industry’s only global view of retail performance measurement. By integrating information from its Watch and Buy segments and other data sources, Nielsen also provides its clients with analytics that help improve performance. Nielsen, an S&P 500 company, has operations in over 100 countries, covering more than 90% of the world’s population.

For more information, visit www.nielsen.com.

Copyright © 2016 The Nielsen Company. All rights reserved. Nielsen and the Nielsen logo are trademarks or registered trademarks of CZT/ACN Trademarks, L.L.C. Other product and service names are trademarks or registered trademarks of their respective companies. 16/10451