

nielsen  
.....

**Bowker**  
a ProQuest affiliate

**BAKER & TAYLOR**  
A Follett Company



# NIELSEN BOOK US STUDY: THE IMPORTANCE OF METADATA FOR DISCOVERABILITY AND SALES

# NIELSEN BOOK US 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 BookScan, PubTrack Digital, PubEasy and Pubnet. David also takes the lead on Nielsen Book's metadata products in North America.

## **CONTRIBUTORS:**

Our thanks for contributions and advice go to:

Patricia Payton, ProQuest and Bowker

Sam Dempsey, Baker & Taylor

Brian O'Leary, BISG

Mo Siewcharran, Director of Marketing Communications, Nielsen Book

## **ABOUT NIELSEN**

Nielsen Book is a leading provider of measurement, consumer research, search, discovery and commerce services globally. Nielsen Book is also the world's largest continuous monitoring service for print book POS tracking through its Nielsen BookScan service, including its B&N, Target, and Walmart BookScan dashboards. In addition, Nielsen Book's portfolio includes transactional services for publishers and retailers through its Nielsen Pubnet and Nielsen PubEasy services; consumer research through its Books & Consumers Tracker, which speaks to 72,000 unique US book consumers annually; and information services through its Nielsen BookData range of products. The Nielsen PubTrack Digital, Nielsen PubTrack Christian and Nielsen PubTrack Higher-Education services provides specialist insights for the e-book, Christian, and Higher Education publishing sectors. For more information email [know@nielsen.com](mailto:know@nielsen.com).

ISBN: 978-1-910284-31-5

© Copyright The Nielsen Company US, LLC

Published in the US December 31 2016

# INTRODUCTION

Nielsen Book first conducted analysis on the link between book sales and bibliographic metadata in the UK market 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. In 2016 we have revisited our earlier paper, and for the first time have carried out a parallel study into the US market.

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 US market 2.5 million different books were recorded as having sales in the 12 month period covered by this study (July 2015 to June 2016). A single bookstore 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), BISG (Book Industry Study Group) and BIC (Book Industry Communication) have developed further standards and formats for the provision of data, such as ONIX<sup>i</sup>, 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.



## DELIVERING AND MAINTAINING DATA

Delivering and maintaining the correct metadata takes constant attention, focus and effort – this study aims to provide some quantitative evidence on the value and effectiveness of these efforts. Areas we will cover in this US study include:

- The provision of a set of basic metadata elements
- The provision of descriptive metadata elements
- The provision of keywords

Some caveats: the bibliographic data we have used in our analysis comes from Bowker® Books In Print data – and though Books In Print data is used widely within the US book trade, not all retailers or libraries use this 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, Books In Print data is likely to represent a good measure of the best level of metadata available in the US 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.

## OUR APPROACH AND DATA

Nielsen Book measures retail sales for approximately 85% of the US market through our BookScan panel, providing robust, reliable and granular data on book sales in the US. Our sponsor, Bowker, aggregates bibliographic data from 40,000 publishers to create an extensive database of titles available in the US market, which is then widely used by retailers both for internal systems and on consumer facing websites.

We have combined these two data sets to undertake this study, which focuses on the top 100,000<sup>ii</sup> best-selling titles over a 12 month period (July 2015 to June 2016<sup>iii</sup>). While this is a relatively small proportion of the total ISBNs recording sales during that time period (around 4%) our data set represents approximately 86% of total book sales over the period. Analyzing the metadata for those titles allows us to identify the correlation between metadata and sales at a high level. Our key measure is average sales 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.

It is also important to note that we are only carrying out a quantitative analysis – looking at the number of metadata elements that are present in comparison to an ideal ‘complete’ record. We are not measuring the quality of the metadata – either the accuracy of attributes attached to the product record, or the effectiveness of the more descriptive data elements or keywords. Such an analysis would likely present further interesting and valuable findings, but is outside the scope of this study.

## PRODUCT DATA BEST PRACTICES

The Book Industry Study Group (BISG)<sup>iv</sup> takes a leading role in coordinating and promoting metadata best practices for the US book market. BISG contributes to the development of ONIX by feeding into EDItEUR’s ongoing activities, and manages the BISAC classification scheme. In addition to this, BISG produces best practice guidance such as their *Product Metadata Best Practices*.

In all of these activities, BISG brings together organizations from all parts of the publishing industry – importantly, including downstream data partners such as wholesalers and retailers who are using the data to help get books through the supply chain and into the hands of customers. More information on BISG’s metadata practices is available from the BISG website.

## BASIC DATA ELEMENTS

Our first measure of the completeness of a title record’s metadata is the presence of a set of basic data elements. These may be described as the objective attributes of the book as a tradeable product, rather than the more descriptive data elements, which we will examine in the next section. The data elements we have grouped together to represent this basic level of completeness include the following:

- ISBN
- Title
- Format/Binding
- Publication Date
- BISAC Subject Code
- Retail Price
- Sales Rights
- Cover image
- Contributor

Analyzing our data set by this measure gives the results below. We clearly see the positive correlation between the completeness of this basic set of metadata and sales, with titles meeting this level of completeness seeing average sales 75% higher than those that don’t.



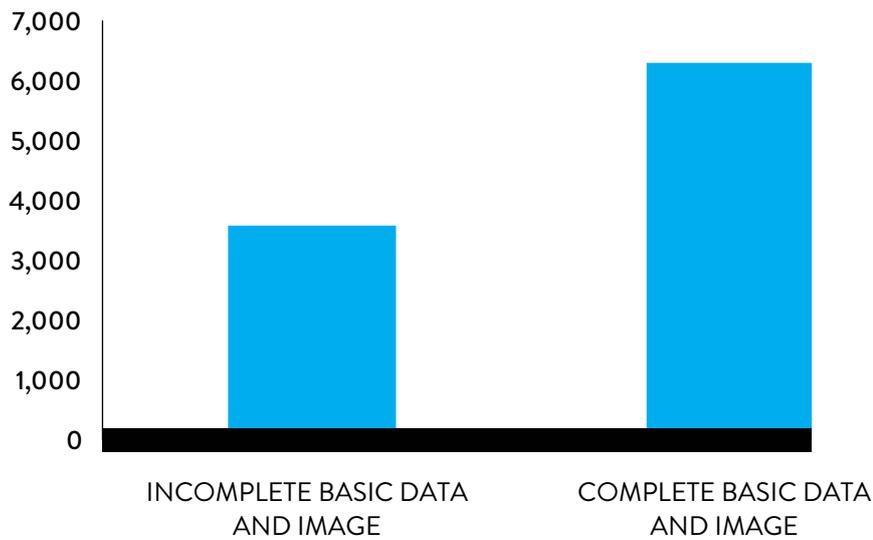


Fig 1.1 Average unit sales per ISBN for records holding complete basic data and a cover image

To drill down a level further, we looked again at this measure in terms of the broad genres of fiction, non-fiction and children’s. The graph below shows that we see the same positive correlation between complete basic data and sales – with the strongest correlation observed for fiction titles, where average sales are 170% higher for titles meeting the criteria than those which don’t. Non-fiction and children’s both sees average sales 55% higher for titles meeting the criteria. We will see consistently through a number of measures that fiction tends to see the highest correlation between the completeness of metadata and sales.

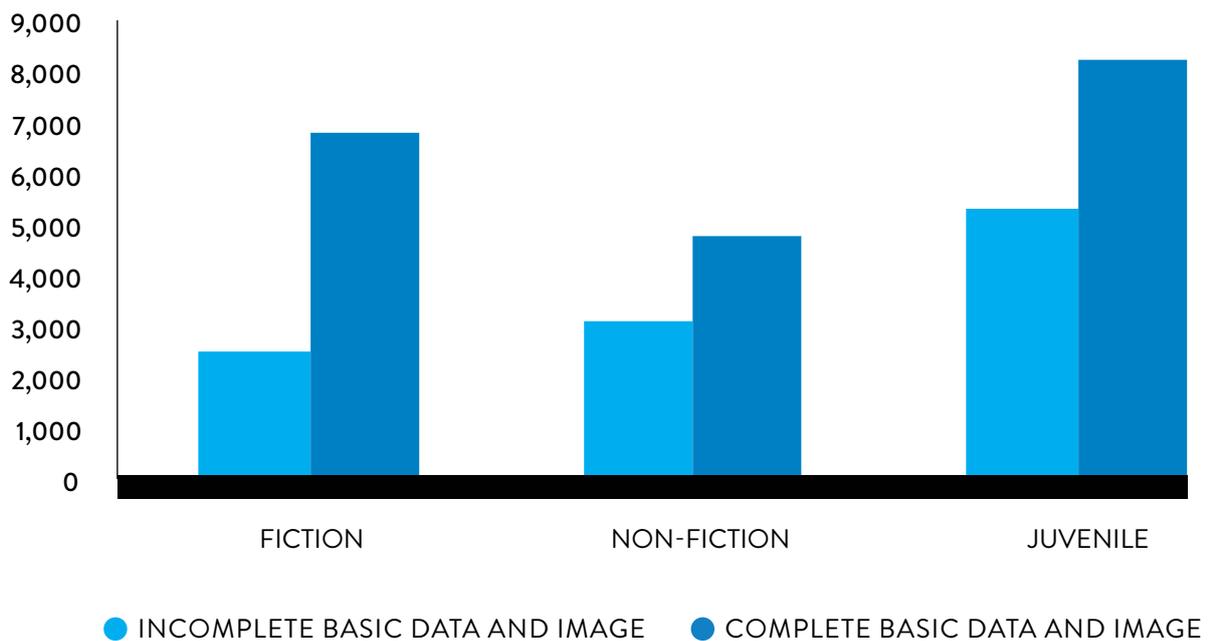


Fig. 1.2 Average unit sales across broad genres for titles with complete or incomplete basic data and a cover image

Taking the presence or absence of a cover image in isolation, we see that much of the positive correlation we saw for titles meeting the basic metadata requirements can be attributed to the cover image. The graph below illustrates this, with titles holding a cover image correlating with sales 51% higher than those which don't.

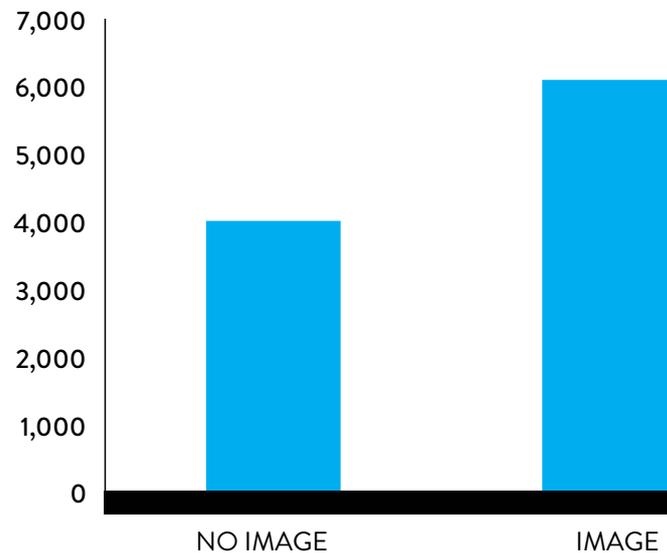


Fig. 1.3 Average unit sales for titles with or without a cover image

Splitting this out into broad genres shows that this is consistent – and once more that the strongest correlation between the presence of this element and sales is found for fiction titles.

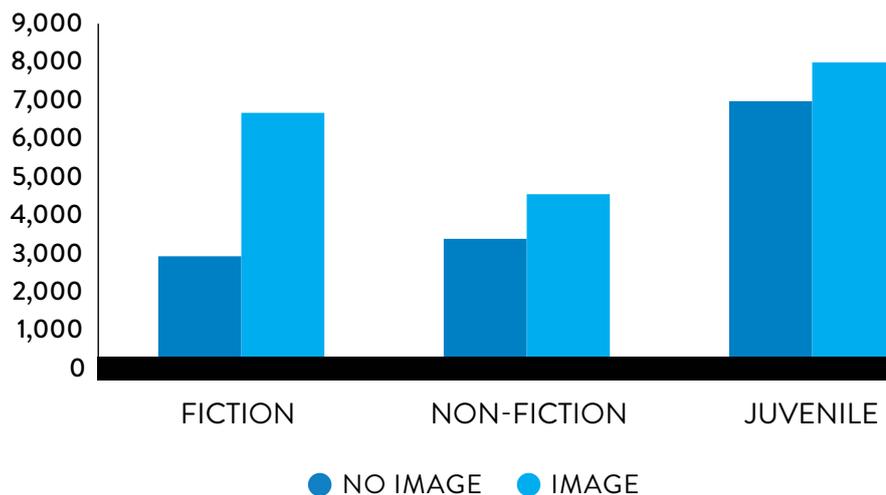


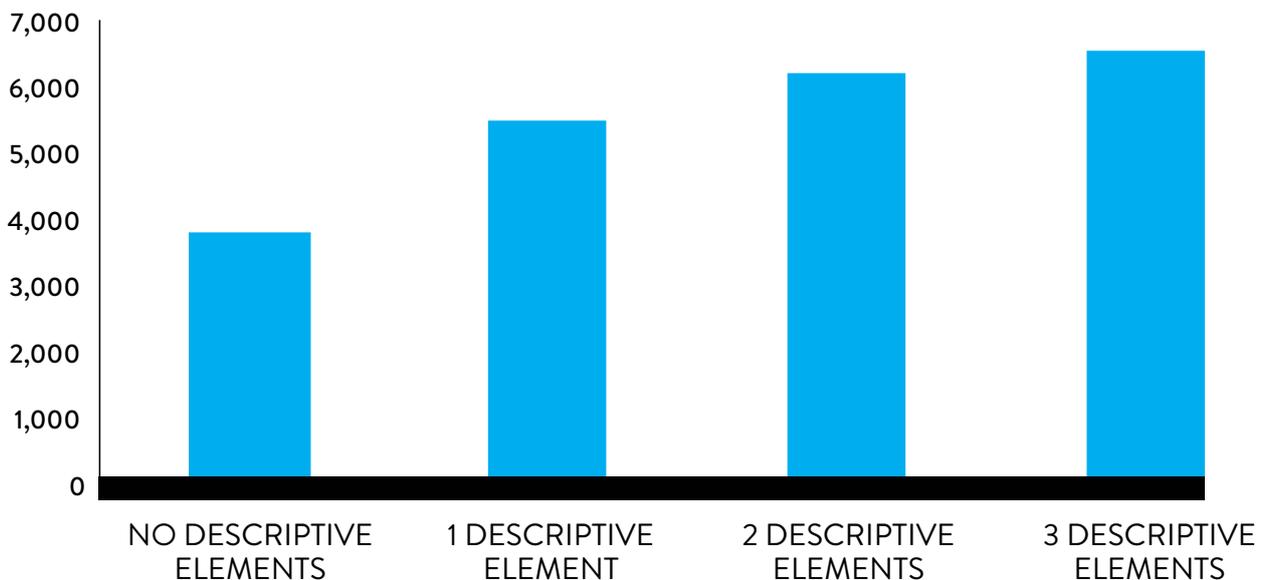
Fig. 1.4 Average unit sales across broad genres for titles with or without a cover image

Through these simple measures we already see a positive correlation between the completeness of metadata and sales. This is consistent with what we found in our 2012 UK metadata white paper, and saw again in our recent UK study.

## DESCRIPTIVE DATA

In addition to the basic data needed to identify a title and help it move through the supply chain, descriptive data adds to the completeness and richness of the data, and should translate into increased discoverability both for book trade buyers and consumers.

Within our data set we have included the title description, author biography and review, and have analyzed these data elements and the correlation with resultant sales. The graph below shows titles grouped into those that hold zero, one, two or all three of the descriptive data elements in our data set. We clearly see that, as the number of descriptive data elements for the titles increases, the resultant average sales are higher. Those titles holding all three descriptive data elements see average sales 72% higher than those with no descriptive data attached.



*Fig. 2.1 Average unit sales for titles with varying levels of descriptive data*

Breaking this down into broad genres shows a similar pattern, but with one anomalous result for children's titles, where titles holding one descriptive element see average sales higher than those with 2 or 3 descriptive elements. Looking at the children's titles that hold just one descriptive data element, we find the bestselling among them are board books, coloring books and classics – where descriptive data is less relevant, as the titles are already known to the consumer. This echoes what we saw in our UK study, where annuals and branded product skewed the figures observed for children's books.

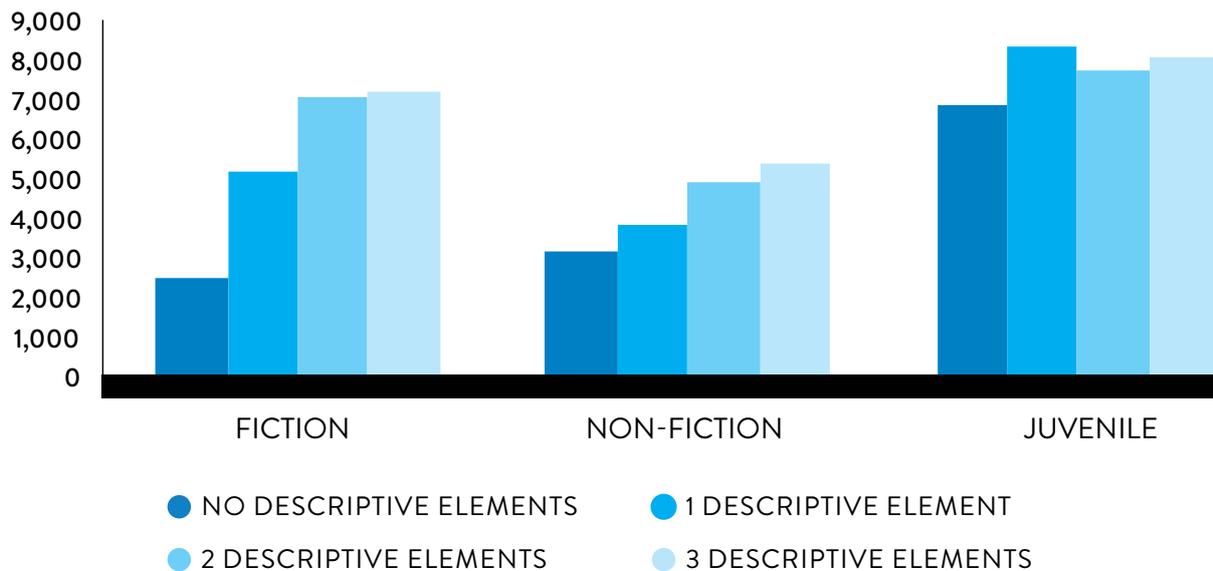


Fig. 2.2 Average unit sales for titles with varying levels of descriptive data, across broad genres

We also see the starkest difference in average sales for the fiction genre. This can be seen as an indication that fiction is the genre most reliant on customer browsing, and therefore more reliant on the presence of descriptive metadata to assist browsing.

## KEYWORDS

Keywords can be added to a title record to supplement the other descriptive data available. Where a title description, review or author biography are intended to be readable, intelligible blocks of text, keywords are simply a list or collection of terms that can be associated with the title and used by search engines and other applications.

The aim of keywords is explicitly to increase a title's likelihood of discovery when searched for. Keywords can include elements such as:

- Character names, locations or associated organizations
- Broader descriptive terms where the title may straddle more than one classification
- Additional information on themes covered in the book
- Related titles or authors

The above list is by no means comprehensive. In adding keywords to a title record, the data supplier is attempting to second-guess what search terms a consumer may use in a search engine or retailers website, and include those terms to maximize their hit rate.

BISG have produced a very informative guide to keywords, which provides further useful guidance<sup>9</sup>.

Analyzing our data according to the presence or absence of keywords produces the results seen in the graph below. Titles which hold keywords see average sales 34% higher than titles with no keywords.

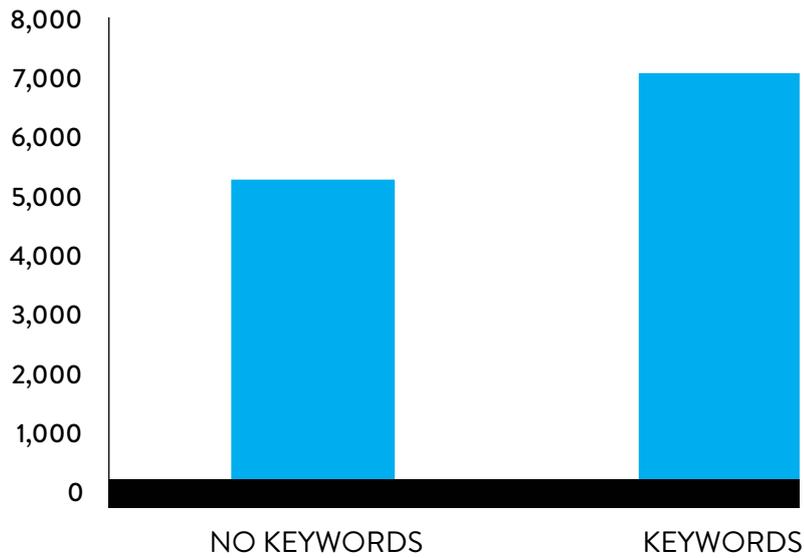


Fig. 3.1 Average unit sales for titles with and without keywords

Looking at keywords across broad genres, we once more find that the strongest positive impact of increasing data completeness is for fiction titles.

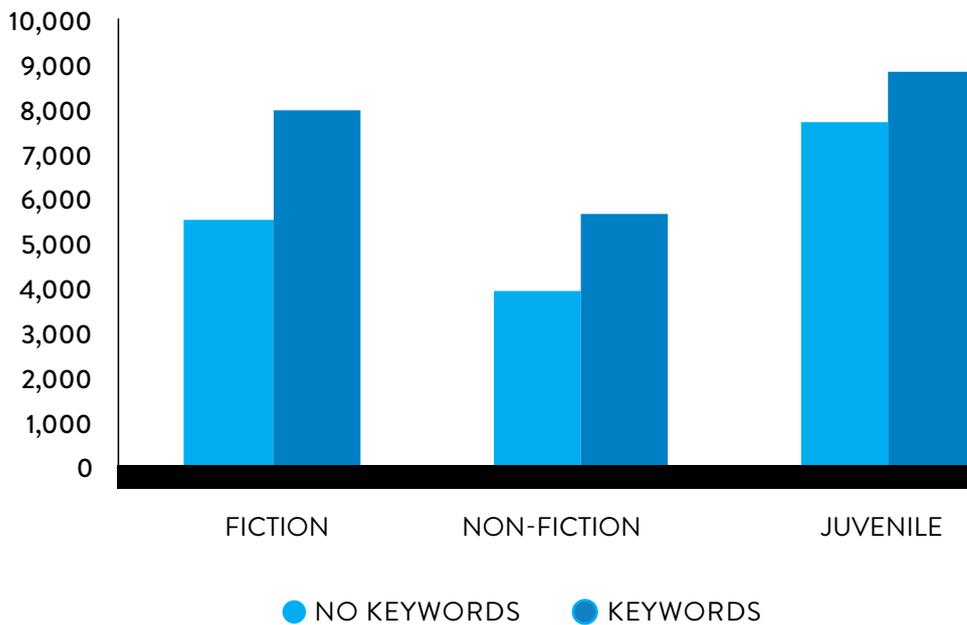


Fig. 3.2 Average unit sales for titles with and without keywords across broad genres

Combining our data for titles with varying numbers of descriptive elements and keywords allows us to look at the titles which hold the optimal level of descriptive data – i.e. all three descriptive data elements and keywords. This is shown in the graph below, and split across broad genres. All three broad genres show that titles with the optimal level of descriptive data see the highest average sales, with once more fiction titles seeing the strongest correlation.

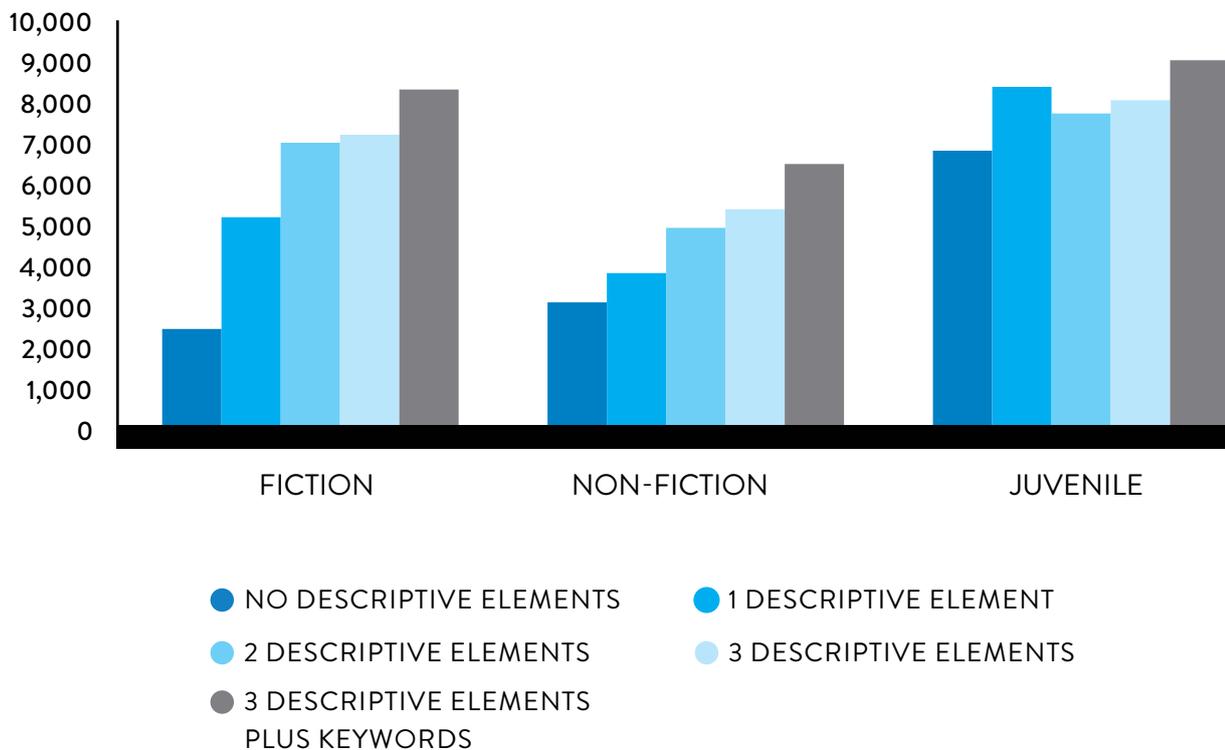


Fig. 3.3 Average unit sales for titles with varying levels of descriptive data and keywords, across broad genres

## ADDITIONAL FINDINGS FROM THE UK METADATA STUDY

Nielsen's 2016 UK Metadata study covers much of the same ground as this US study – with findings very much in line with what is presented here (even down to the anomalous results we see for Children's titles and descriptive data). However, there are some additional measures we have carried out in the UK study that were not possible for the US due to differences between the data sets, and we will summarize these briefly here.

## DATA TIMELINESS

Nielsen Book's UK bibliographic data records, not just what data elements are received but when they are received. As part of the BIC Basic and ONIX Compliance standards for data supply in the UK, there is a timeliness requirement which stipulates that the data should be supplied 16 weeks, or 112 days, ahead of publication. We were therefore able to analyze our UK data based on this timeliness criteria, to judge how this correlates with sales.

The graph below illustrates how, in addition to supplying the appropriate metadata for products, supplying the data sufficiently far ahead of publication correlates with higher average sales. Providing data early ensures that downstream book trade partners can effectively plan their ordering and stock management of titles, and consumers browsing for titles will be able to find what they are searching for, even in advance of publication.

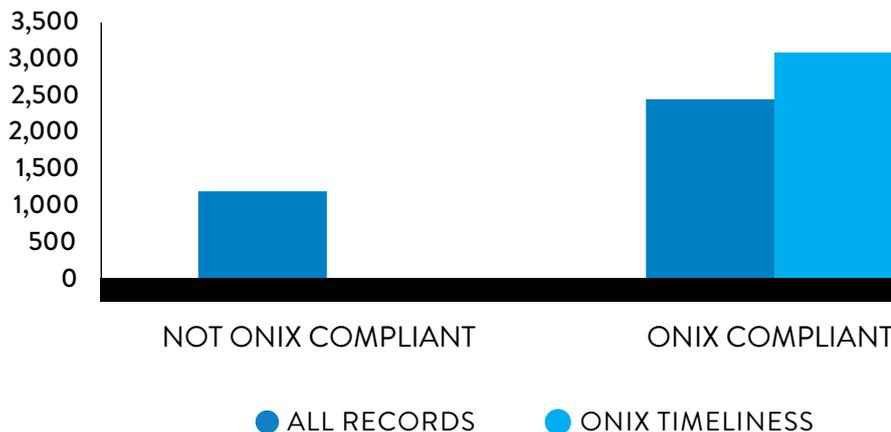


Fig. 4.1 Average UK unit sales per ISBN for records which are not ONIX Compliant, those which are ONIX Compliant and those which also meet the ONIX Compliance timeliness requirement

## LIBRARY BORROWINGS AND METADATA

As well as measuring book sales in the UK, Nielsen Book also measure public library borrowings by aggregating data from 70 public library authorities via our Nielsen LibScan service. We can therefore analyze library activity in a similar way to sales, and judge the value of metadata for the library sector.

The graph below shows average borrowings for titles with varying levels of descriptive data. Those titles carrying the full complement of descriptive data elements see average borrowings over twice the level of those that carry no descriptive data. This shows that descriptive data plays a key role in the sourcing and discovery of books in the library sector, just as it does for book sales.

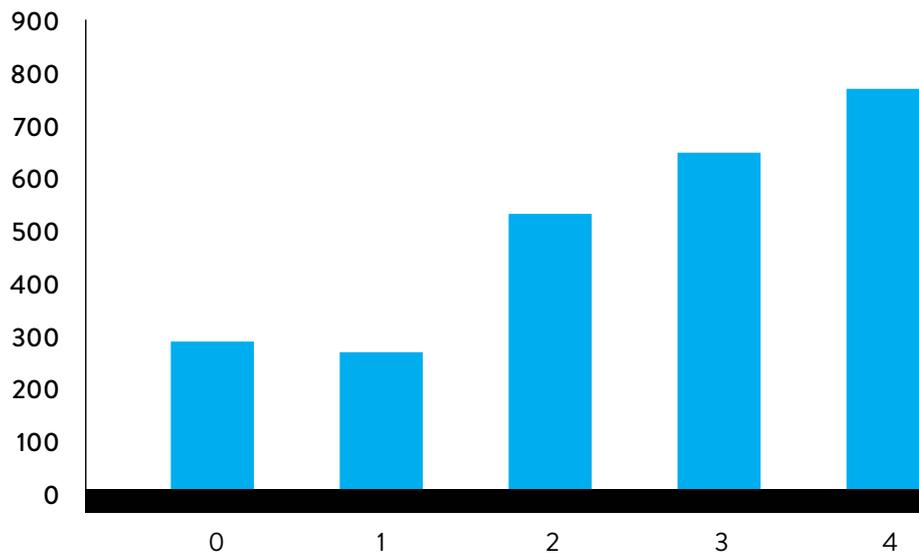


Fig. 4.2 Average UK public library borrowings per ISBN for records holding zero to four descriptive data elements – short description, long description, author biography and review

## SUMMARY

Through our various measures we have consistently seen that increasing completeness of metadata correlates with higher sales on average. This holds true for basic data and cover images, textual descriptive data and keywords. These findings also reaffirm what we have seen in our two UK metadata studies, adding further credence to the results.

## KEY FINDINGS INCLUDE:

- Titles carrying the full complement of basic data elements and a cover image see average sales per ISBN 75% higher than those which do not hold this complete data
- The presence of a cover image alone correlates with average sales 51% higher than titles which do not hold a cover image
- The presence of descriptive data elements on title records correlates with higher average sales – titles holding the 3 descriptive elements we examined saw average sales 72% higher than those with no descriptive data attached
- The addition of keywords shows a correlation with higher sales again – compared to those titles which hold all 3 descriptive data elements, those that also carry keywords see average sales 28% higher

While 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 maximize sales.

## NOTES/KEY:

<sup>i</sup>ONIX 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 originally created by EDItEUR ([www.editeur.org](http://www.editeur.org)) and the Association of American Publishers – it has since been developed by EDItEUR jointly with BIC ([www.bic.org.uk](http://www.bic.org.uk)) and BISG ([www.bisg.org](http://www.bisg.org)), and is now maintained under the guidance of a broad international steering committee.

<sup>ii</sup>There are some titles within the top 100,000 sellers from Nielsen BookScan for which the data is not available for output. These are generally retailer exclusive editions, and we have not included these records in our data set. This reduces the total number of records we have used for our analysis to 97,397.

<sup>iii</sup>More specifically, the sales data used is from 19<sup>th</sup> July 2015 to 17<sup>th</sup> July 2016. This equates to Nielsen BookScan week 29 of 2015 to week 28 of 2016.

<sup>iv</sup>BISG previously administered a product data certification program to help organizations measure and improve the quality of their metadata. While the program is currently inactive, BISG are seeking to offer this again in the near future.

<sup>v</sup>BISG's *Best Practices for Keywords in Metadata* is available for free download from their website ([bisg.org](http://bisg.org)).

EDItEUR: The trade standards body for the global book, e-book and serials supply chains which develops, supports and promotes standards including ONIX, Thema and EDItX, and provides management services for the International ISBN and ISNI Agencies.

**Sponsors:** Our thanks to our sponsors for supporting this report which is an essential tool for the book industry. This US 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 US Study: The Importance of Metadata for Discoverability and Sales* will be used, as the UK edition, extensively to promote metadata provision and best practice for suppliers of bibliographic data in the US and globally.



#### 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](http://www.baker-taylor.com)



#### ABOUT BOWKER:

Bowker is the world's leading provider of bibliographic information and management solutions designed to help publishers, authors, and booksellers better serve their customers. Creators of products and services that make books easier for people to discover, evaluate, order, and experience, Bowker is the official ISBN Agency for the United States and its territories and Australia. A ProQuest affiliate, Bowker is headquartered in New Providence, New Jersey with additional operations in the United Kingdom and Australia. For more information visit [www.bowker.com](http://www.bowker.com)



#### ABOUT FIREBRAND TECHNOLOGIES:

Firebrand Technologies has been helping publishers manage their internal workflows, digital distribution, marketing efforts and business intelligence for 30 years. Our solutions touch the lives of thousands of publishing professionals every day. A unique, community-focused approach gives our clients more than just tools and support, it helps create an optimal atmosphere for innovation and product success. For more information visit [www.firebrandtech.com](http://www.firebrandtech.com)



#### ABOUT ONIXSUITE:

Onixsuite by GiantChair is the most advanced metadata tool on the market today, and integrates seamlessly with the legacy systems of publishers and distributors. Available as a full-service title management and digital distribution platform, as well as an API, Onixsuite is able to store and manage ONIX data in all languages. Publishers and distributors in countries around the world use cloud-based Onixsuite to evaluate, improve and distribute their ONIX. Consulting and data cleaning services are also available. Contact us for further information [sales@giantchair.com](mailto:sales@giantchair.com) or visit our website: [www.onixsuite.com](http://www.onixsuite.com)

## 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](http://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. 16/10666



