



White Paper: The Link Between Metadata and Sales

By

Andre Breedt, Head of Publisher Account Management

David Walter, Research and Development Analyst

25 January 2012

The Link Between Metadata and Sales

Introduction

Perhaps more than any other industry the book trade is reliant on sharing product data in order to function efficiently. With the massive variety of products available the ISBN has been central to the industry for generations, and attaching accurate and appropriate information to the ISBN has become essential for identifying products both for trading partners and customers.

Supplying sufficient and appropriate metadata is actively encouraged within the book industry, but until now the effect of supplying (or not supplying) this metadata has not been measured. Being at the heart of the global book industry, Nielsen has unique access to both bibliographic data supplied by publishers and used by retailers, and sales data supplied to us direct from book retailers. Combining these data sets we can assess the impact that supplying good quality, appropriate metadata has on sales.

As the internet becomes the consumer's primary source of information, it's likely that customers will increasingly discover and learn about books via the information they can find about that book, rather than the physical product itself. With this in mind we have also examined sales through the online channel to see if the link between metadata and sales is greater and whether customers are even more reliant on the data supplied.

Finally, we have looked at the broad genres within the book market to assess whether providing accurate and appropriate metadata is even more crucial in certain genres, and to ascertain what might be the very most important metadata elements to supply.

Our Approach and Data

We have examined the link between metadata and sales by taking three different approaches.

The first of these is to look at the level of population of basic metadata elements in the UK top selling 100,000 titles of 2011, and to identify any links we can make between sales and the level of metadata population. We have used the BIC Basic standard as our measure of the completeness of this metadata. Although over 1 million unique ISBNs were sold in the UK in 2011, these top 100,000 titles represent 91% of the total volume sales, and 87% of total value sales in 2011.

The second way in which we have looked for links between metadata and sales is to analyse a number of the enhanced metadata elements which can be added to a product record. These are the more descriptive elements such as the long and short descriptions, author biography and review. These data elements are only sent out in Nielsen BookData's outgoing data feeds to retailers for those publishers who subscribe to our Nielsen BookData Enhanced Service, and we have therefore used only records from subscribing publishers for our analysis in this section.

Finally, we have looked at sales for publishers who took up our Nielsen BookData Enhanced Service during 2010 thereby releasing the enhanced, descriptive metadata elements for their records for the first time, and tried to measure any impact on sales.

BIC Basic – achieving the basic level of metadata requirements

Book Industry Communications (BIC)ⁱ aim to promote supply chain efficiency in the book industry through the application of standard processes and procedures. BIC has a set of standard data attributes which all book product records should have, and this is called BIC Basic. There are 11 required metadata elements to meet the BIC Basic standard:

- 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 UK

Within our bibliographic data Nielsen Book holds all of these metadata fields, and has a 'BIC Basic flag' which indicates when a record has all of the required BIC Basic fields. We have a separate flag which indicates the presence of an image for the record, and using these two indicators can identify titles that meet all of the BIC Basic criteria.

Looking at the top selling 100,000 titles from 2011ⁱⁱ we analysed the volume sales for titles where either the BIC Basic or image flag was missing, and compared these with titles where one of the flags were missing and titles where both the BIC Basic and image flags were present, indicating that the BIC Basic standard was met. Figure 1.1 shows the average sales per title for these four different sets of records.

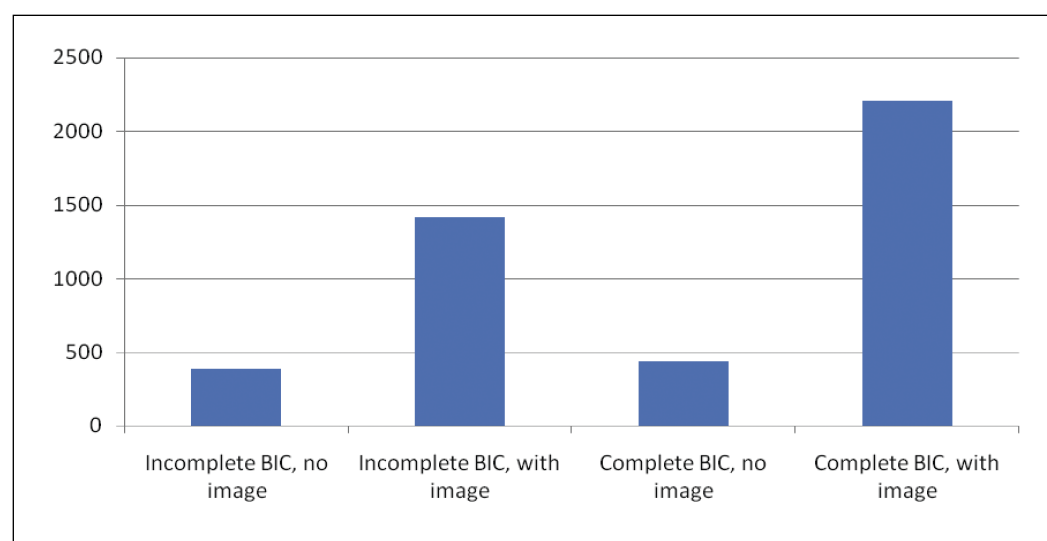


fig. 1.1 Average sales per ISBN for records with complete or incomplete BIC Basic data and an image

The positive impact of supplying complete BIC Basic data and an image is clear. Records without complete BIC Basic data or an image sell on average 385 copies. Adding an image sees sales per ISBN increase to 1,416, a 268% boost. Records with complete BIC Basic data but no image have average sales under 437 copies, but when we look at records with all of the necessary data and image requirements, average sales reach 2,205. This represents an increase of 473% in comparison to those records which have neither the complete BIC Basic data elements or an image. Figure 1.2 shows a direct comparison between all records with insufficient data to meet the BIC Basic standard, and those that meet the requirements.

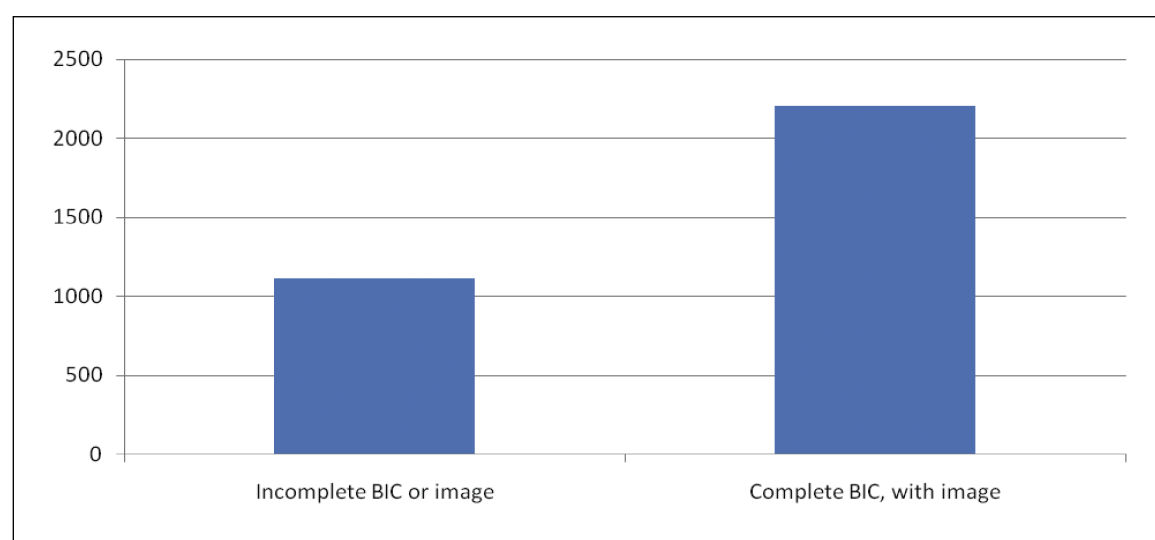


fig. 1.2 Average sales per ISBN for records with complete or incomplete BIC Basic data and an image

Here, the average sales across all records with incomplete BIC Basic elements are 1,113 copies per title, with the complete records seeing an 98% increase in average sales.

We can further break this down to compare the effect seen for online sales with those seen through physical book shops, which we will refer to as offline sales. Figure 1.3 shows the change experienced across these two channelsⁱⁱⁱ.

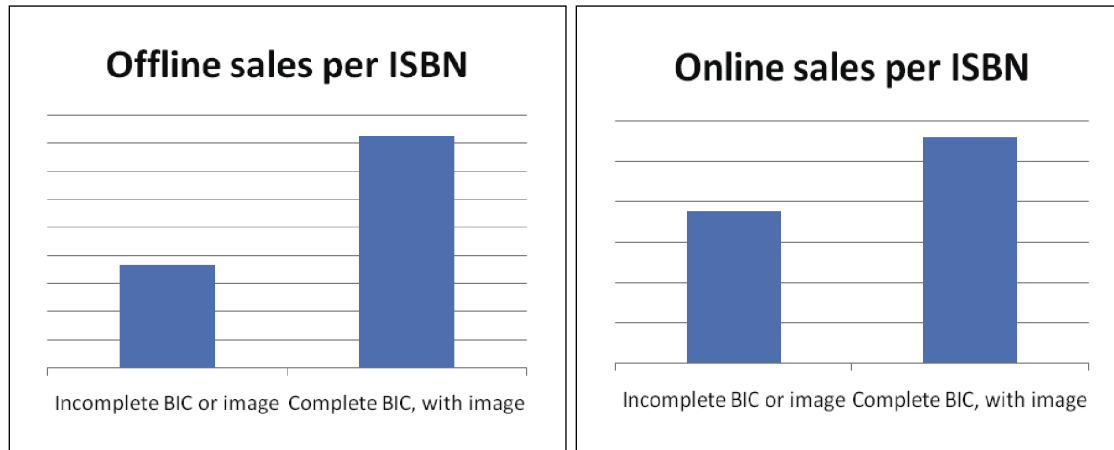


fig. 1.3 Average sales per ISBN for records with complete or incomplete BIC Basic data and an image in offline and online book retailers

We see here that there is a more marked difference in average sales for offline retailers than online retailers, where we might have expected the opposite to be true. The offline retail channel sees sales rising 124% for titles meeting the BIC basic standard, whereas online retail sales see growth of 48%.

Looking at the effect of supplying this basic level of metadata by genre shows differences in the degree to which average sales are affected, but in all cases the records with more complete metadata have average sales significantly higher than those with incomplete data. Figure 1.4 shows this across the four broadest genres used in Nielsen BookScan information.

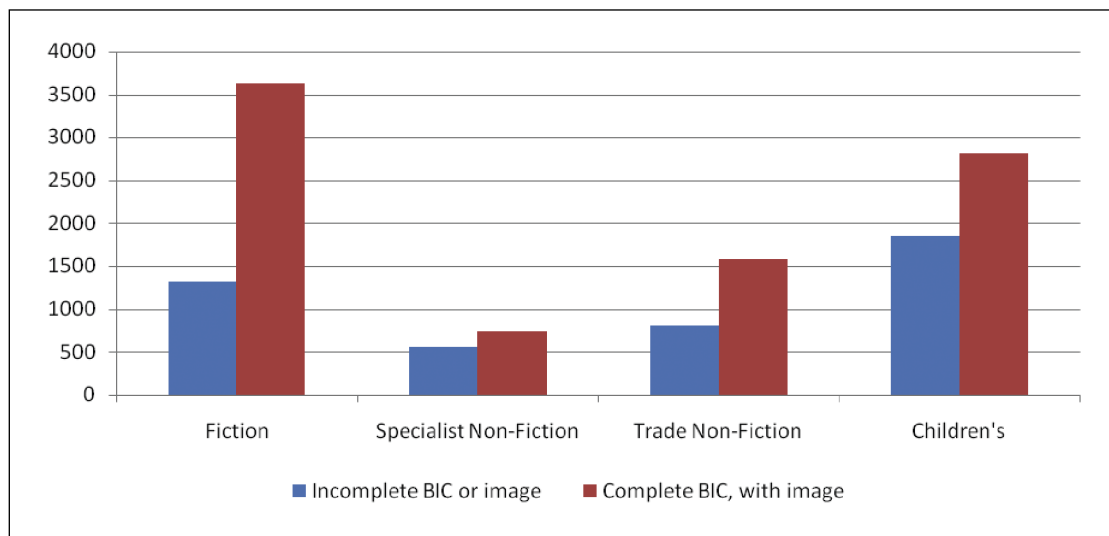


fig. 1.4 Average sales for records with complete or incomplete BIC Basic data across different genres

Fiction titles see the most dramatic improvement in average sales for records with complete BIC Basic data and an image. Records with incomplete data have average sales of 1,326 but this grows by 173% for titles with complete BIC Basic data and an image, which have average sales of 3,624. Specialist Non-Fiction titles show the smallest increase but even so see average sales rise by 33% for titles with complete data and an image. Trade Non-Fiction average sales per ISBN grow by 97% when all BIC Basic elements are present, and Children's titles see average sales rise by 52%.

Overall, we can see a clear relationship between the completeness of basic metadata and sales.

Enhanced Metadata – the additional value of descriptive data

In addition to the BIC Basic metadata fields there are further 'enhanced' metadata elements that can be added to title records. These are the more descriptive elements – the short description, long description, review, author biography and table of contents. These elements are only held and sent out by Nielsen for subscribers to our Nielsen BookData Enhanced Service, and we have therefore used this set of records from our initial top 100,000 of 2011 for this section of the analysis^{iv}.

Our primary analysis showed that the table of contents was only present on a relatively low number of titles^v, and is actually only applicable to certain types of book – i.e. anthologies, academic works and textbooks. We have therefore not analysed our data set in terms of the presence or otherwise of a table of contents on the title records. Conversely author biography, although not applicable to all titles (Guinness World Records and children's annuals are two notable examples) is appropriate for the majority of titles^{vi} and we have therefore included this in our analysis.

Our initial approach to analysing the effect that these enhanced metadata fields have on sales was simply to look at the number of enhanced elements present on each record – with the 'ideal' record having all four elements: short and long descriptions, review and author biography. Figure 2.1 shows the average sales per ISBN for records holding zero to all four enhanced metadata elements.

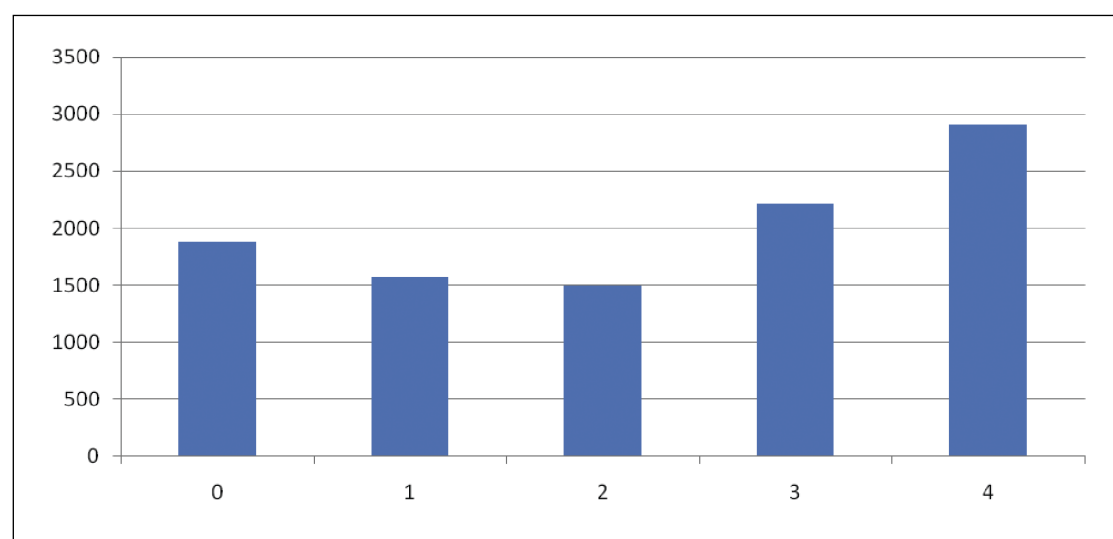


fig. 2.1 Average sales for records with varying levels of enhanced metadata

We can clearly see the high impact on sales that having a data rich product record has. Titles which hold all four enhanced metadata elements sell on average over 1,000 more copies than those that don't hold any enhanced metadata, and almost 700 more copies than those that hold three out of the four enhanced metadata elements. In percentage terms, titles with three metadata elements see an average sales boost of 18%, and those with all four data elements 55% when compared to titles with no enhanced metadata elements.

We do see some anomalies here, in that records with one or two enhanced metadata elements sell fewer copies on average than those with no enhanced metadata elements. Looking at the titles which fall into this category gives some indication why this may be the case – the highest selling titles within this group are mostly children's activity books, annuals and well established brands^{vii}.

Splitting sales out into online and offline sales channels shows some clear differences in the effect that having enhanced metadata on product records has on sales. Figure 2.2 shows title records with varying levels of enhanced metadata split into offline and online sales.

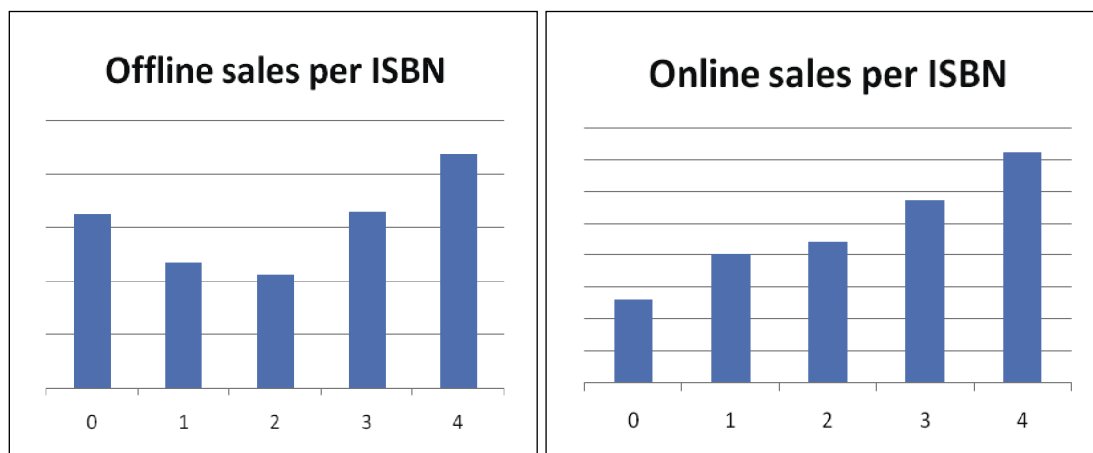


fig. 2.2 Average sales for records with varying levels of enhanced metadata, split into online and offline channels

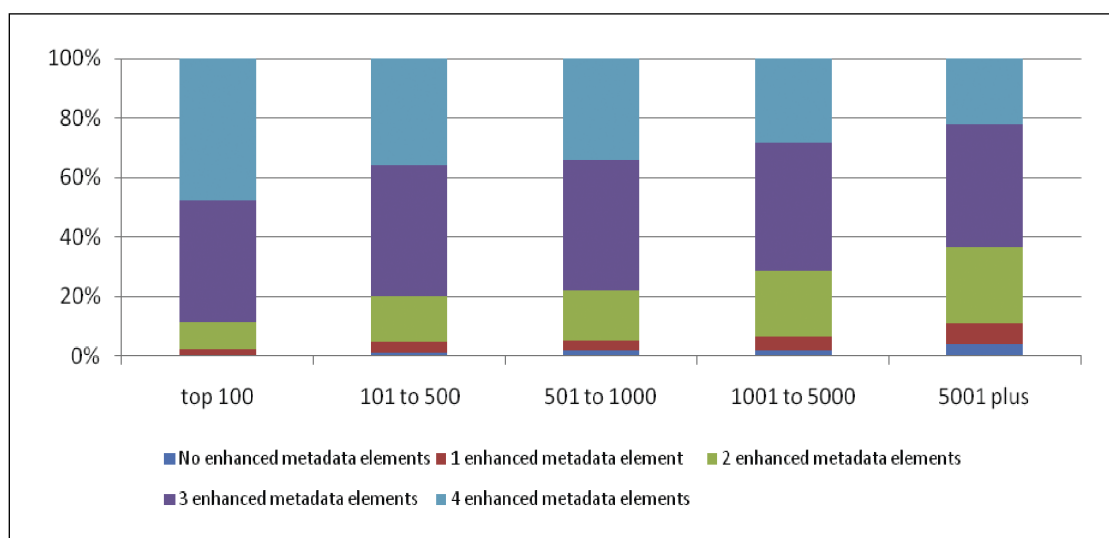
Offline sales reflect the anomalies we previously saw to a greater extent, with records having one or two enhanced metadata elements selling fewer copies than those with no enhanced metadata elements. The overall percentage increase in sales for titles with all four elements in comparison to those with none is 35% in the offline channel.

Online sales show a marked contrast to this – records with progressively increasing amounts of enhanced metadata see progressively increasing average sales. Records with just one enhanced metadata element see an increase of 55% in comparison to those with none; those with two enhanced metadata elements see an increase of 71%; those with three increase 120%; and those with all four enhanced metadata elements see average sales increase by 178%.

This indicates in the strongest terms that supplying enhanced metadata for product records is even more essential for sales through online retailers. It is a logical assumption that online buyers will be reliant on the information supplied to inform their purchases, and our research bears this out.

An alternative way of looking at how the level of enhanced metadata correlates with sales is to look at the difference in levels of metadata within different bands of the top selling titles – i.e. the top 100 titles, 101 to 500 etc. Having analysed our data in this way, we find that titles towards the top of the chart have the highest degree of enhanced metadata population, with levels decreasing as we move through the lower bands of the chart. Figure 2.3 shows the proportion of titles within each chart band with the varying levels of enhanced metadata.

fig. 2.3 Proportion of titles within chart bands with varying levels of enhanced metadata



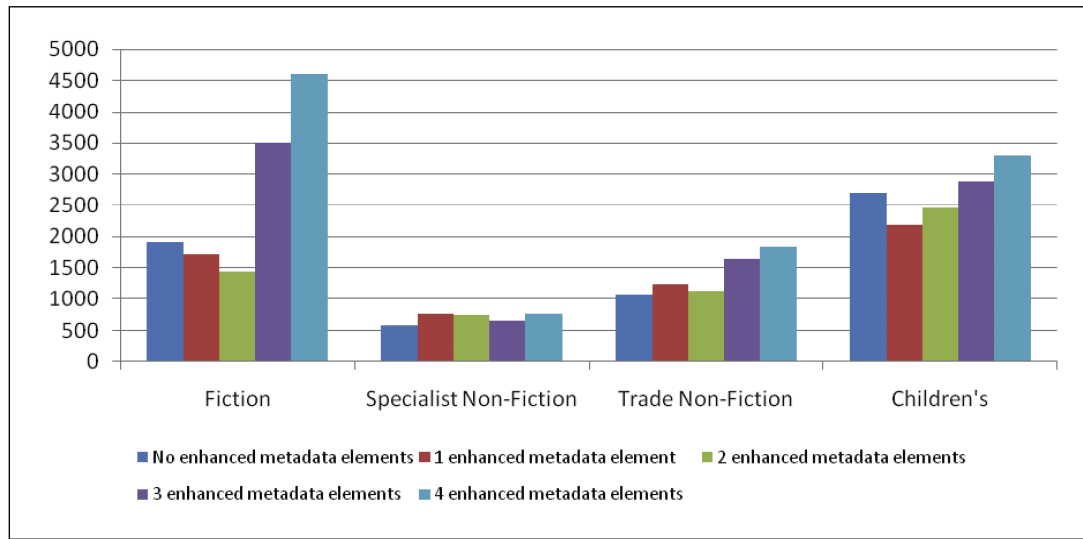
We can clearly see that a greater proportion of titles within the top 100 have all four enhanced metadata elements present (48%), with this proportion decreasing substantially for the titles in positions 101 to 500 (36%), and continuing to decrease as we move down the chart positions. Although it is difficult to separate cause and effect when looking at chart positions, what we see here

reinforces our earlier analysis – just as we’ve seen that higher levels of metadata correlate with higher average sales, we see that the very bestselling titles have a higher level of metadata than those that come further down the sales rankings.

Breaking the overall sales down into broad genres, we see that Fiction experiences the sharpest rise in average sales when the metadata is more complete – as we saw with the level of BIC Basic completeness. Fiction titles with all four enhanced metadata elements see average sales 140% higher than those with no enhanced metadata, as illustrated in figure 2.4. Specialist Non-Fiction titles see an increase of 33%, Trade Non-Fiction titles an increase of 71% and Children’s titles an increase of 22%.

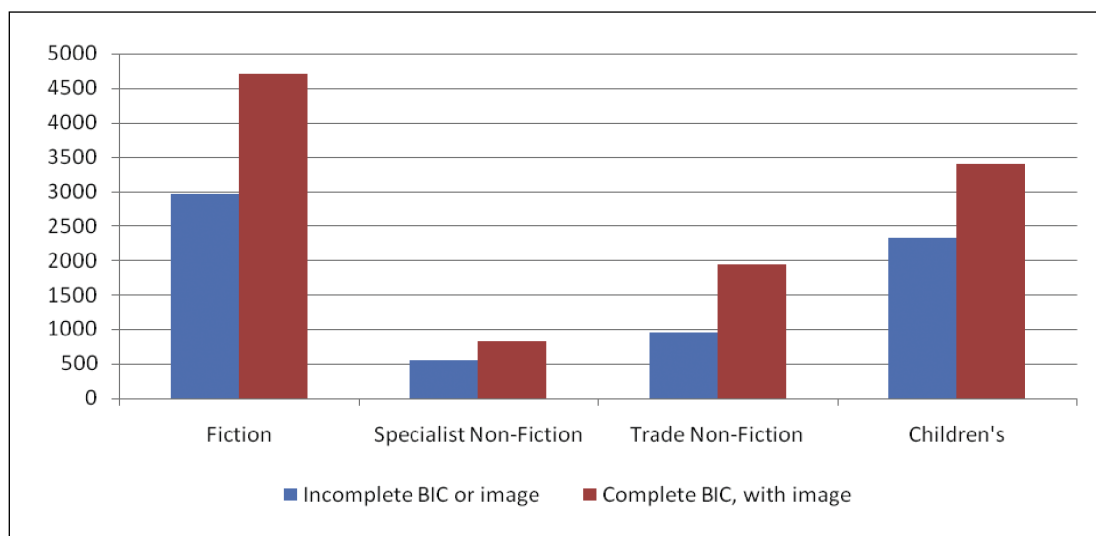
fig. 2.4 Average sales for records with varying levels of enhanced metadata, within broad genres

Having enhanced metadata on a product record doesn’t necessarily mean that the more fundamental BIC Basic criteria are being



met. Therefore, our ideal product records would meet all of the BIC Basic requirements (including image) as well as having all four enhanced metadata fields populated. Figure 2.5 shows the difference, by genre, between those records which meet the BIC Basic criteria and those that don’t – these are the average sales for titles which have all four enhanced metadata elements.

fig. 2.5 Average sales for records with all four enhanced metadata, split by compliance to BIC Basic standard

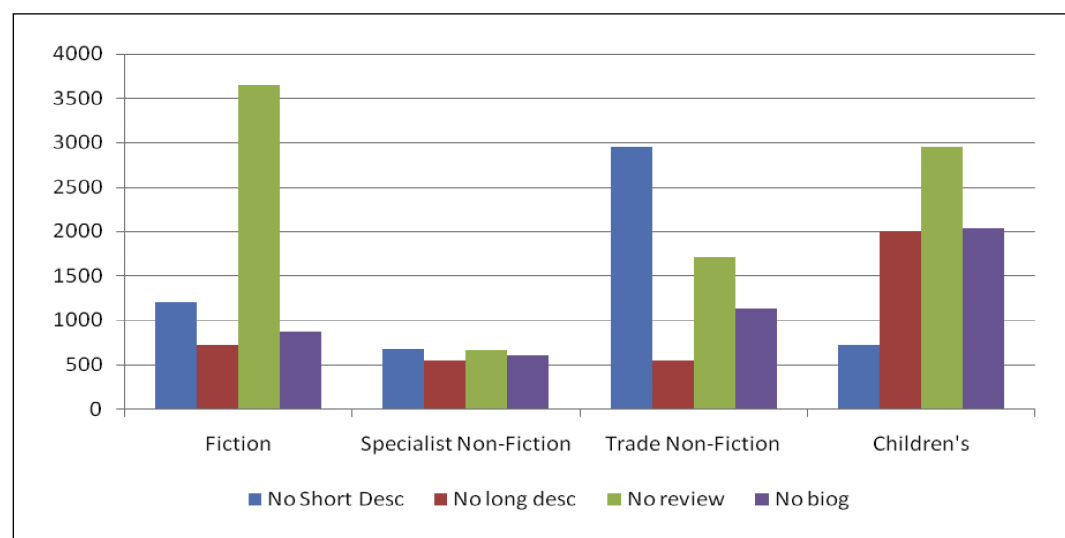


We see here that Trade Non-Fiction shows the greatest proportional rise in average sales, with a 103% jump when the BIC Basic requirements are met in addition to all four enhanced metadata elements being present. Fiction sees a 59% rise, Children’s a 45% rise and Specialist Non-Fiction a 52% rise.

Carrying out further analysis, we have attempted to identify which enhanced metadata element has the greatest effect on sales. To do this we looked at the average sales per ISBN for the groups of records which have three out of the four enhanced metadata

elements, divided up according to which of the elements is missing. Figure 2.6 shows this by broad genre.

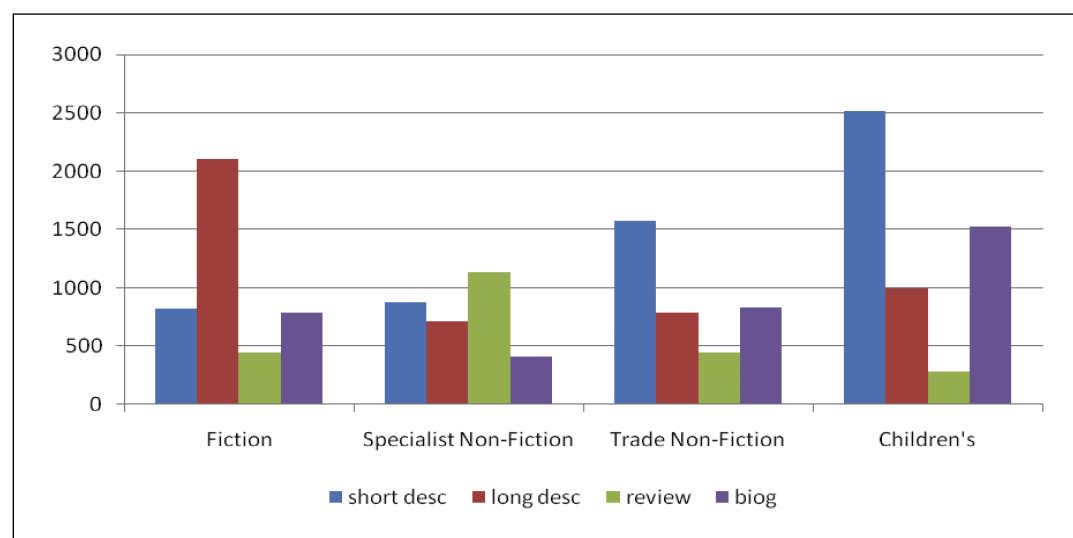
fig. 2.6 Average sales for records with three enhanced metadata elements, by genre



Fiction, Specialist Non-Fiction and Trade Non-Fiction all see the greatest negative impact on sales when the long description is omitted from the title record. Children's titles, however, see average sales fall the most sharply when the short description is omitted. Review appears to be the least significant indicator of increased sales within Fiction, Specialist Non-Fiction and Children's, whereas the short description appears to be the least significant factor for Trade Non-Fiction.

We have also analysed records where only one metadata element is present to judge which single element may have the greatest impact on improving average sales, and this mirrors what was seen in figure 2.6 to a certain extent.

fig. 2.7 Average sales for records with one enhanced metadata element, by genre



We see clearly that the long description appears to be the most significant factor for Fiction titles, and the short description for Children's titles. The picture for the Non-Fiction genres is somewhat less clear, but this is perhaps not surprising given the range of categories within these broad genre splits.

Overall we see clear indications that supplying a set of full enhanced metadata for product records helps to maximise sales, and that this relationship between enhanced metadata and sales is even stronger for the online retail sector.

Releasing Enhanced Metadata – the impact on sales of adding enhanced metadata to records

The final way in which we have analysed the relationship between sales and metadata is to look at publishers who took up a subscription to our Nielsen BookData Enhanced Service during 2010, thereby releasing their enhanced data to book retailers. We have looked at annual sales for the year preceding the subscription and the annual sales for the year following subscription for all new subscribers' titles in the Nielsen BookScan sales data^{viii}. Figure 3.1 shows the combined volume sales for these publishers before and after taking up their subscription.

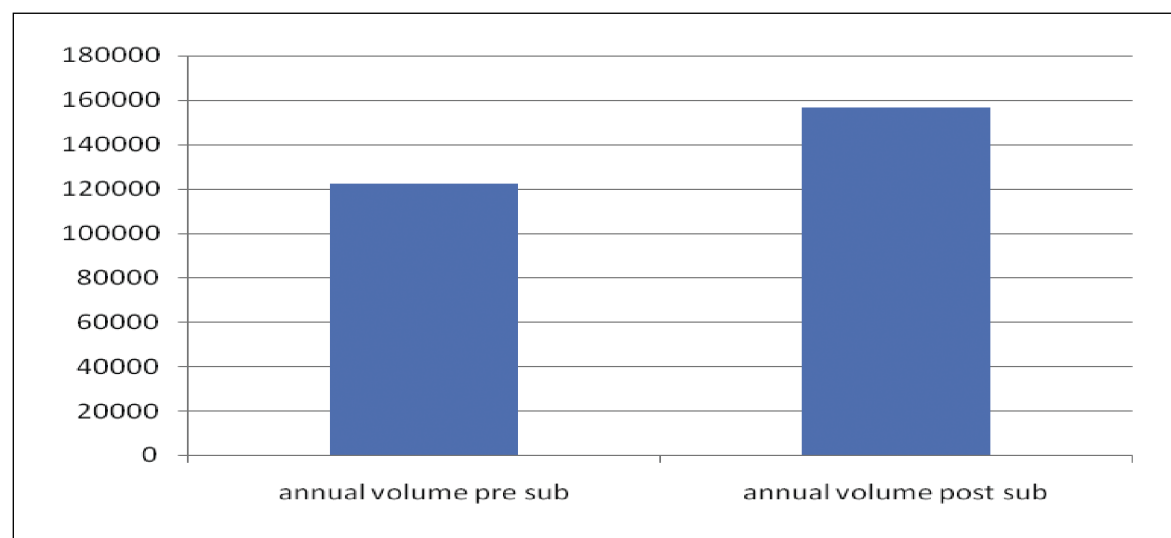


fig. 3.1 Annual sales for publishers for the year preceding and then following subscription to the Nielsen BookData Enhanced Service

We can clearly see the growth in total volume sales, which represents 28% year-on-year growth. Of the 156 publishers analysed 129 saw positive growth the year after starting their subscription.

As it is likely that many of these publishers may have been new publishers, we have removed any publishers which were selling less than 10 units in the year before subscription was taken up. This is shown in figure 3.2.

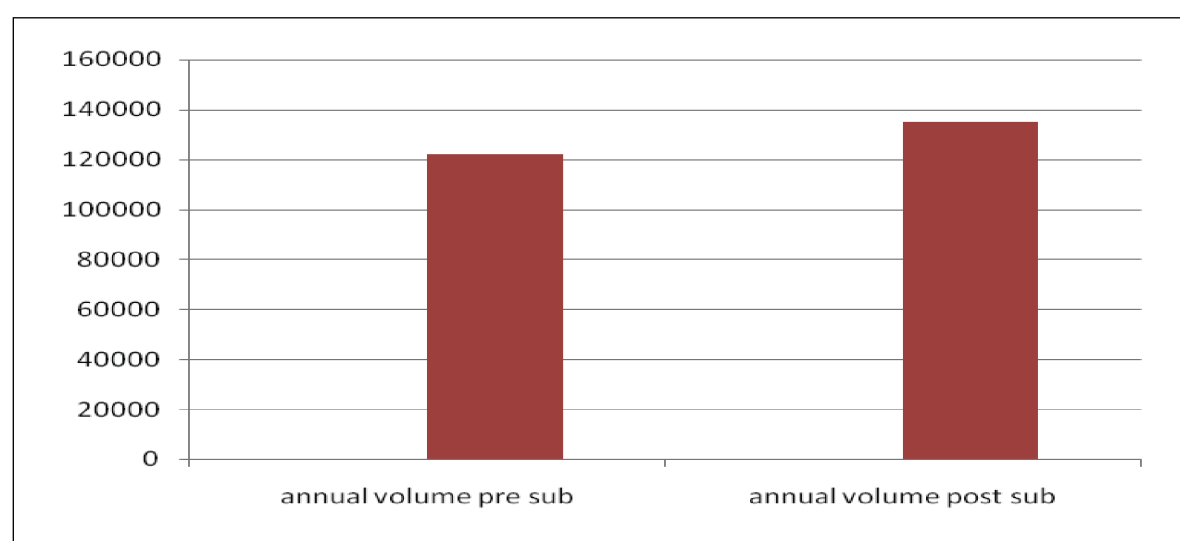


fig. 3.2 Annual sales for publishers already selling over 10 copies, for the year preceding and then following subscription to Nielsen BookData's Enhanced Service

Once more we see growth in volume sales of 11% year on year, with 44 out of the 65 publishers in this category seeing positive growth in volume sales. All of these positive growth figures should be taken in the context of a market which saw a decline of 2.7% in volume sales between 2010 and 2009, and by 6.3% between 2011 and 2010.

Summary

Although it is impossible to directly measure the impact that having accurate, appropriate and enhanced metadata has on sales, all of our analysis points to a clear relationship between the quality of product records and sales. We are not able to take account of other factors, such as increased marketing or improved distribution arrangements, but can feel confident in the assertion that improved metadata is part of a mix of factors which helps titles to reach their sales potential.

Our key findings have been:

- Titles that meet the BIC Basic standard see average sales 98% higher than those that don't meet the standard
- The addition of an image has a strong impact on average sales, of 268% in comparison to titles without an image
- Ensuring that all four key enhanced metadata elements are present on product records can help average sales rise by 55% in comparison to records where none of the elements are present
- Split into offline and online sales, offline sales see an increase of 35% for titles which have all enhanced metadata elements present, whereas online sales see a massive 178% increase
- For Fiction, Specialist Non-Fiction and Trade Non-Fiction the long description appears to be the most vital piece of enhanced metadata, whereas for Children's titles the most vital appears to be the short description
- Fiction is the genre most significantly affected by the completeness of both BIC Basic and enhanced metadata
- The difference in average sales between records which don't meet the BIC Basic standard, have no image and don't have enhanced metadata, and records which do meet BIC Basic and have all four enhanced metadata elements is on average over 2,600 units, which represents an increase of almost 700%
- New subscribers to the Nielsen BookData Enhanced Service can see annual volume sales increase by up to 28%.

In each way that we have analysed our data, we have seen a consistent positive relationship between the level of metadata supplied and sales, with this correlation being particularly strong for the online retail sector. It is logical that this is the case, given that consumers are reliant on the bibliographical data to locate the desired product, and would perhaps be dubious to complete their purchase online if there is insufficient information to confirm that the product is correct.

Given the indications that we have for online sales, there is a strong suggestion that supplying complete and enhanced metadata will be even more vital for e-books, where the bibliographic data is the consumer's only source of information. The implication is that as the book industry takes its next step into the digital age, metadata will not only remain an essential part of the industry but become increasingly important.

Appendix

- i Visit <http://www.bic.org.uk/17/BIC-Basic/> for a full list of BIC Basic requirements.
- ii From our top 100,000 selling titles we have extracted the titles which are identified as being in our outgoing data, and those titles which sold both online and in physical book stores, so that we can compare these two parts of the market. This gives us a data set of 95,344 records, from which our subsequent data sets are derived.
- iii In order to maintain confidentiality we are unable to show unit figures for comparisons between online and offline sales. We can however give the percentage changes experienced in the different channels, without jeopardising confidentiality.
- iv This represents 84,050 records from our initial data set.
- v 11,883 of our 84,050 records (just over 14%) had a table of contents.
- vi 67,312 of our 84,050 records (just over 80%) list an author and would therefore seem to be eligible for an author biography.
- vii Within this data set of records with no enhanced data elements, the children's titles have average sales of 2708 copies, showing that to a certain extent titles of this type are less dependent on enhanced metadata when looking at the overall market.
- viii In total, there were 201 new subscribers to our enhanced service in 2010, of which we were able to identify 154 by imprint within Nielsen BookScan data.



For further information contact:

David Walter, Research and Development Analyst, Nielsen BookScan
t: +44 (0)1483 712 353 or email: David.Walter@nielsen.com

Andre Breedt, Head of Publisher Account Management, Nielsen BookScan
T: +44 (0)1483 712 227 or email: Andre.Breedt@nielsen.com