

To provide some context for my presentation – I will be talking STM and professional development content at undergraduate and post-graduate level.

Ebooks in this presentation are NOT hardware devices, but rather book content presented as web pages and PDFs or PDF-like book facsimiles.

The data I draw on is a mix of internal Elsevier data and publicly available research.

Why do we find it necessary to add an "e" to everything?

What does adding an "e" say about the new version of the noun?

E-mail is quicker, and more conversational than the regular 'mail' of previous centuries.

E-commerce ... it's quicker. Immediate, often.

E-journals? ... definitely quicker

And how about e-books?

There is no shortage of things with "e" stuck in the front.

They're not always better – is there anything more interactive than a real-live tutor?

This is probably a fair way to describe them:

They're the same-old thing

with "e" stuck in the front

And it's all a bit quicker.

"e" plus "book" -

Twenty per cent "e"

and eighty per cent "book"

Ebook = e + book

**What is an ebook?**

This is a story of starting a journey from the wrong place.

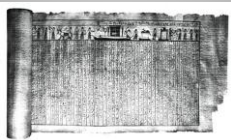


**Books are an industrial artifact**

Books evolved in an industrial world: they are an industrial artifact, designed for reading rather than using with qualities framed by historical process. The terms we use to describe their components: recto, verso, folio, stitching, and leading.

These are all words that have their roots in industrial process and the early practitioners of industrial book publishing

Had the first successful printing press made impressions upon papyrus scrolls rather than paper sheets and hadn't been largely developed in the European industrial world we would probably have completely different terms



**E-papyrus**

And we would have very different ebooks.

There would be no “page” feature, but scroll speeds

You might flip the scroll around to read annotations (or make them, in our brave new world of e-papyrus)

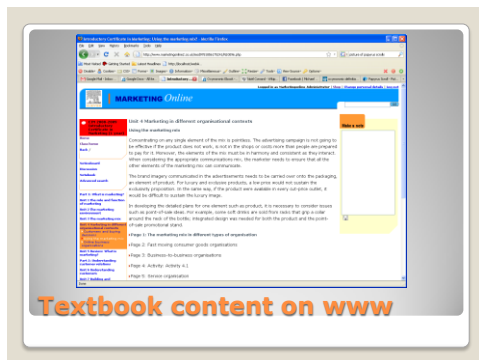
The book, as it has evolved, is designed for readers. It's the right size to be carried, and opened. Typefaces and type sizes work well with the human eye. The optical qualities of the paper are carefully gauged.

The book has evolved for reading and readers, and so, the ebook has been continued in this tradition

### **We have ten years experience of ebook usage**

By which I mean, data on how people read material that was previously only available on paper - and we know from this hard data that users of our online content have distinct behaviour patterns that underline the distinction between different product types that were previously only available in the normal book form factor.

For example, under-graduate level text book material...



Marketingonline is one of our most venerable sites of online content.

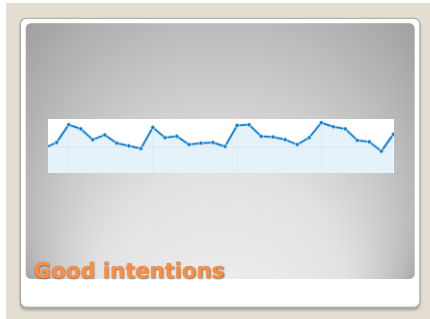
The content is freely available to users who buy the printed book. The books are written to an examination syllabus for professional marketing qualifications.

About 15% of book buyers make use of the site - it's the same content as the book but with some additional whitepaper material and revision notes.

These are textbooks - with chapters that are highly structured and linear in form. They have clear chapter objectives, and clear summaries

And the section headings in between are carefully mapped to both objectives and summaries.

The qualifications are, broadly speaking, at an undergraduate level and are highly directed by tutors, who typically have a weekly contact with the people using marketingonline.



Typically people use the site once a week. As a body, they start the week with good intentions: Monday and Tuesday lunchtime are the busiest periods.

About a quarter of pages viewed in this very **linearly presented content** are navigation pages, the rest are content pages - which are consecutively linked. As a set of pages accessed by users, content pages comfortably make up the bulk of the data

In a clear distinction from the regular book, content pages are divided up by content chunk rather than page length.

About 20% of users will scan these pages, and 80% spend over a minute engaging in the content in a more meaningful manner. We presume they are reading online, as this is too much time to scan, copy or print.

Searching and browsing are equally popular, and account for about a quarter of user behaviour.

The most popular chapters are those that are either introducing or summarizing a theme - those that are most suited to exam preparation, rather than the informative and in-depth chapters.

Users of this site typically spend quarter of an hour on the site and read a good chunk of a chapter – a quarter, a third.

As Daniel Pollack said this morning, we are not unsurprising to see that the traffic on an annual basis is heavily skewed to the weeks before examination.

### Turning to reference work usage at a higher academic level...

An Elsevier co-sponsored north American study from last year indicates that user behaviour with post-graduate, research and reference level material is very different to that exhibited by undergraduate level students using text-book type content.

Left to their own devices, these higher level users will not choose to start with ebooks - rather they will use **generic discovery applications**

Google, Wikipedia, Google Scholar figure highly - and less often university provided catalogues where adequate training provision has been provided

| Resource           | Usually start with... | Often start with... | Supplemental / never start with... |
|--------------------|-----------------------|---------------------|------------------------------------|
| Web search engines | 32%                   | 30%                 | 0%                                 |
| Wikipedia          | 14%                   | 23%                 | 0%                                 |
| Google scholar     | 12%                   | 15%                 | 0%                                 |
| Ebooks (web)       | 6%                    | 15%                 | 48%                                |
| Ebooks (library)   | 6%                    | 10%                 | 55%                                |

Starting points for online research

(Explain missing columns and multi-answer option, hence 100% not reached)

Users start searching with general terms and narrow down, **unless they are searching for a specific book**, in which case they will search on title and author pairings.

Remember that these are typically more sophisticated and self-steered learners than the marketingonline undergraduate using tutor guided material.

These researchers are looking for answers to help them with research, whether or not the answers come from publicly available web-pages learned articles, books or wikipedia.

**Such users do not usually expect to find the answers in books.**

When ebooks do show up in research results, some applications provide an immediately accessible - and detailed - table of contents

Users are able to scan this data quickly and make a judgment about whether a book is at a suitable level or not

Books **without extensive tables-of contents** are usually discarded without being accessed.

Users commented that - without page numbers –

they were unable to assign a length factor to the chapter,

and therefore assign appropriate resources to read it thoroughly,

however an extensive table-of-contents goes some way to providing the same information, the caveat being that it must be visible without having to open the ebook.

The next action is to either print or copy the appropriate section

Again - books that do not provide this facility

- or make it hard to use

- or exploit an unusual interface

are usually discarded, unless the restrictions are easy to get around

**The end goal of a researcher is acquisition over immediate consumption. This supports what Kate Price said this morning about user behaviour.**

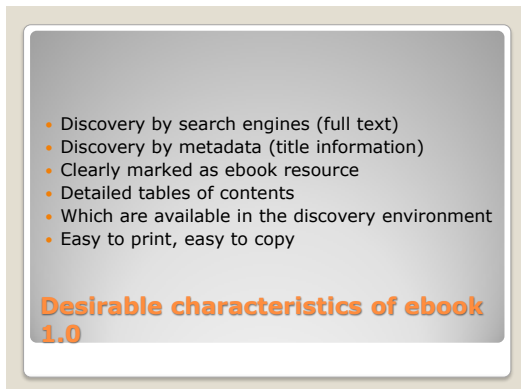


We see in these research results evidence that users are not benefiting from the real potential of online content.

As with other things with "e" added on them, the that the principle benefit is from the consequent speed

- of having ebooks indexed and searchable,
- printable
- and copiable.

**In a research context, therefore, ebooks that exhibit these characteristics**



- **of being discoverable by search engines**
- **being identifiable by metadata**
- **being clearly flagged as linking through to an ebook (rather than a web page or a journal)**
- **having clear and extensive tables of contents**
- **which are available to view within the discovery environment**  
(so you don't have to open them up to scan the content)
- **and are easily printed or copied**

**are highly valued by researchers.**

- "It's on screen, but very limiting ... This is enough to say I would go elsewhere. If you just had it as a PDF, would search and find in text" [Single page view]
- "I can now scroll down, from page to page and simply read." [Multi-page view]

**Multi-page PDF is preferred over single page view, user quotes**

Interestingly, when researchers are searching *with the intention of finding a book* which is, **in itself, a low-occurrence behaviour, they prefer to find a PDF** of the content - as they are comfortable with the navigational tools and functional behaviour in the PDF viewer.

Researchers in our study make specific reference to the importance of being able to cite material using the orthodox manner and praised the quality of PDF print-outs.

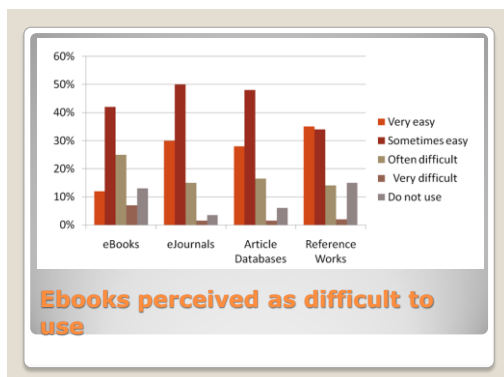
Notably – and interestingly - applications that returned single PDF or PDF type pages were significantly less valued.

### Why don't research students expect to use ebooks?

Constrained beneath the limitations of cultural legacy and poor functionality - whether deliberately restricted by digital rights management or poor usability design, ebooks suffer a bad reputation.

User expectation of ebook usability is poor – and so they do not want to find ebooks, they would rather find answers elsewhere.

And the key user task - the ability to scan and assess ebook content can be difficult.





"ebooks" are characterized as either "often difficult to use" or "very difficult to use" by nearly 40% of research-level users - a total unmatched by any other content type surveyed

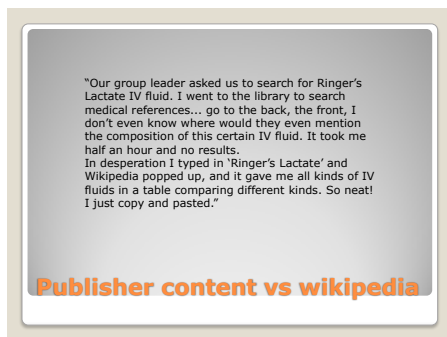
### **The elements of ebooks that are seen as being successful**

- particularly although not exclusively when referring to the hard physical ebook product
- are rudimentary and often epiphenomenal to the nature of "e"

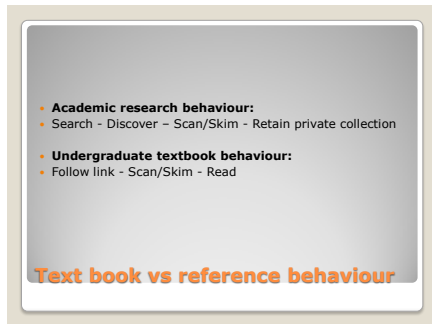
Searchable. Bookmarkable. Sometimes you can add notes, sometimes you can add highlighting.

However, our data shows that students prefer to make annotations and highlighting within their own context either by printing, or by copying-and-pasting on to their own computers.

And when a large proportion of the book is to be read, the student will often seek out a physical paper copy of the book to read rather than use the ebook that they have accessed



So although we – publishers - perceive these tools to be "must haves", the inability that students at higher academic levels have to mash-up the book with their variously sourced information in a generic environment will lead them to creating their own, private annotated library.



The typical post-graduate researcher use-case is, therefore:

**Search - Discover - Scan/Skim - Retain private collection**

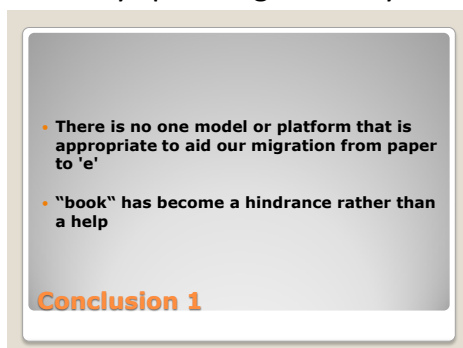
Whereas the under-graduate textbook use-case is more direct and “booky”:

**Follow reference - Scan/Skim – Read**

**By scan / skim I mean – whereabouts in this “book” am I? Is it at the right level? Is appropriate for my needs? These are orientation tasks.**

And indeed, the activity figures support this user-based research, with a 50/50 breakdown for researchers finding ebook resource and then using it,

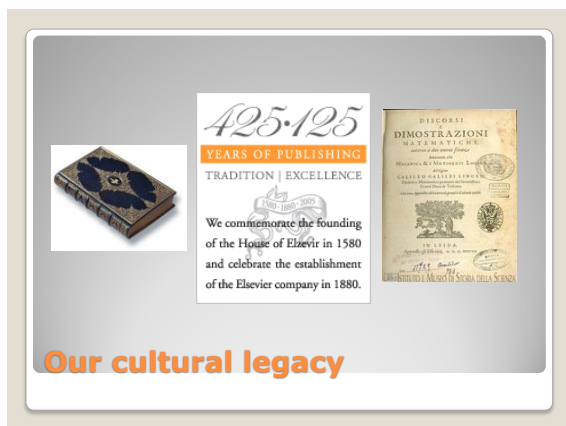
versus the session time for steered undergraduates finding and using the relevant portion of a textbook being closer to 25/75,  
their typical session is over 50% longer,  
and they spend significantly more time on each page.



**As a consequence, our first conclusion is that there is no one model or platform that is appropriate to aid our migration from paper to 'e' and that "book" - as a useful carrier term for the product that we create - has become a hindrance rather than a help**

To be specific: hanging onto a the notion of a book limits the scope of our product and damages the relationship with our users - in particular as they advance up the academic ladder

**Specifically, the more the characteristics of "books" that are embedded in "ebooks",** especially when put against decreasingly linear narrative, the greater we see an opportunity missed and the more we see usability – and the reputation of books - harmed



**But "book" is something that is very comfortable for us**

We are book publishers, we have published books for 100s of years

Our systems and our minds are built around books

Culturally, authors expect to write books and papers and chapters –

the book dictates - or has possibly evolved from - our units of comprehension and authority

Commercially, this is how we reward writers and pay them for their ability to communicate an expertise.

**And it is largely because books'r'us that the book analogy continues in the world of 'e'**  
**However, our principal users are turning away from the concept**

- “Yeah, it doesn’t work. It’s limiting. I can understand there’s publisher’s rights and so forth ... To me, I really wouldn’t come back to it after this. It would be enough to say ‘I can get this information elsewhere.’”
- Quote from user unable to copy and paste

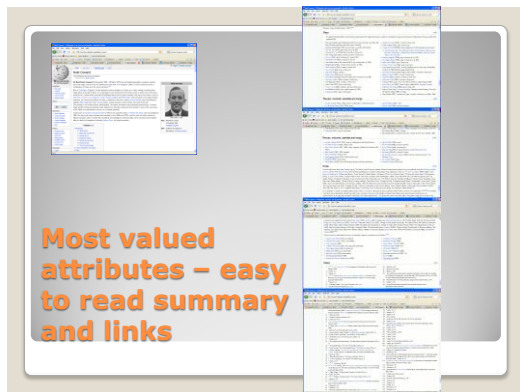
### Turned off by the execution of ebooks

When our users find ebooks, all our research indicates is that ebooks - if we're to carry on using the term – are an end-point - a termination rather than an open-end, a starting point.

They are not things that people particularly want to find if not given a link and told to go "there", they tend to be found serendipitously

We know that users like open-ends. We know this from their behaviour and from their statements - particularly in the case of wikipedia, where the principal value is seen as being in the **easy-to-scan summary** and the associated **links** to more authoritative information.

Here’s the Wikipedia entry for Noël Coward. The summary – pages of detail that are littered with links – and then another four, five pages of links.



Whereas our ebook is a dead-end - there is either no linking, or very limited, static linking - between books, from books to other content.

This use-case contrasts with pre-e use cases, where books were largely seen as starting points for research

And this characteristic - of limited linking - only seeks to reinforce the impression that books are a dead end.

Clearly cross-references - other than formal reference notation and internal cross-reference - have limited place in the printed book - and given the fluid financial model - they would be an expensive investment of uncertain return

And this should be coupled with our aim of having more agile published content - itself laying a direction away from proprietary systems, towards greater interoperability and blending, and right out of our comfort zone.

**But we have to make ebooks more attractive and potentially move away from this rather stifling model**

After all, if we sat down to improve the paper book, we'd hardly  
- deliberately make the pages difficult to turn  
- interfere with the paper and ink so photocopying was impossible  
- we wouldn't go out of our way to restrict people's ability to share them

**As things stand now, ebooks are not valued by their intended users**

## The facts are

- Students do not specifically search for ebooks
- Students have almost no intention of acquiring ebooks for themselves without a significant cost saving
- Students do not trust ebooks to enhance their learning
- Restrictions on copying and printing is a big turn-off
- The number of students accessing ebooks via a library is almost the same as those who get it "for free from the internet" (45% vs 43%)

**Ebooks are not highly valued and rarely bought**

- students do not search for ebooks (unless given a strong steer from their tutors)
- students have almost no intention of acquiring ebooks for themselves
- unless there is a significant cost saving – student say (they would) that 30-50% is a reasonable discount
- students do not trust ebooks to enhance their learning
- restrictions on copying and printing is a big turn-off

and finally,

- the number of students accessing ebooks via a library is almost the same as those who get it "for free from the internet" – 45% vs 43% (versus 4% who'd consider buying an ebook)

Although no-one knows what "get it for free" means...

- Current ebooks are the end of a search, a dead-end
- Ebook content must start to exhibit behaviour more like the rest of the web

**Conclusion 2**

**Our second conclusion is that ebooks in their current state act as an end point in a research session** - whereas previously books formed a key starting point in research

**and that ebooks - or ebook content - must start to exhibit behaviour more like the rest of the web.**

#### **In the Elsevier Labs research and development unit**

we have been working for several years on tools that may play a key role in the shape of our published content, and how it is discovered and used.

At the heart of nearly all our endeavours are text-mining techniques Using language parsing tools to uncover and expose statements and relationships that are buried within our textual content, we hope to benefit the user and scholarly activity by improving the ability to scan deep content for meaning and to contextualize learned material within the corpus.

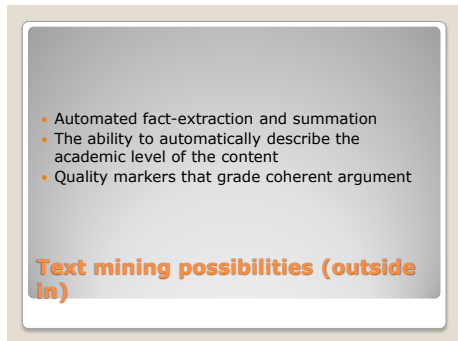
**Elsevier Labs has been investing in developing an expertise in both text-mining and the potential use of text-mined data.**

As I have said, our research indicates that the ability to scan e-content is of vital importance - and, in fact, is the key task for research material - and that material that does not allow users to scan through it is often discounted for that very reason.

Therefore it might be reasonable to explore the provision of either

- additional, manually authored preces,
- trust to the original authors' abilities, and pull out the first few paragraphs
- or to automatically generate sufficient data to improve the scanability of the deep content

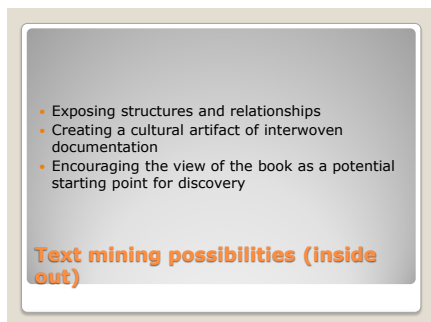
it is in the last area that text-mining promises to produce solid, user-centric results.



Research by Xerox and Elsevier Labs and other academic research institutes offers the prospect of producing

- automated fact-extraction and
  - the ability to automatically describe the academic level of the content
  - and even quality markers that grade coherent argument
- something that worries several academics I know

Digging within the deep content will enable greater discovery of our content too.



For example, we are working on an interoperability project that outlines a methodology to apply tags at paragraph level,

using a combination of expert-tagging, text-mining and user-based tagging.

The bulk of the work being done by text-mining applications, rather than experts or users.

The combined data will permit a descriptive envelope of semantic tags to surround health science material and will expose structures and relationships, not only within a book and within its chapters but between books, and extending beyond that, cross-relating to articles, methods and other web content.



The specific intention of this project is to enable a question-based discovery interface using tags that enable the questioner to frame the answers in the context of their own learning and academic versus practitioner needs

And there is richness potentially to be found in "internal cross-references"

There are several projects that seek to analyse and qualify the nature of formal references –

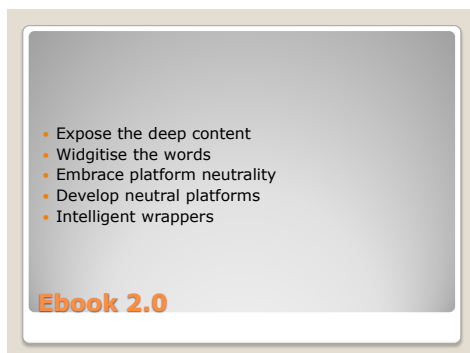
is this referenced material being supported, qualified or cited?

In turn, data which can permit richly interwoven documentation on a temporal / developmental dimension

We anticipate that enhancing links and exposing information about a book's content will increase its perceived value in an online environment.

It will encourage further exploration – and while enriching interior links from the book to the rest of the corpus might not change the view of the book as a potential starting point for discovery, it will certainly remove the functional dead-end that ebooks currently are for undergraduate and research users.

### **Before I finish, here's our ebook2.0**



with apologies for the 2.0 bit. There are no technological barriers left – **it's all about the paradigm.**

- **Expose the deep content, make it free-but-limited**, like Search Inside or Google - but less booky in its presentation. We must aim for ubiquity.

- **Add widgets to our ebook content**, enable our content to co-exist alongside webpages, papers, other publishers' material [youtube widget] - make it easy to cite and trackback. A single button to allow copying, citation and orientation – just like blog trackbacks, think of a youtube “share this book” feature

- **Develop platforms that can contain content and provide context from many places** - there's plenty of technology to support this [think about media players which can play 10, 15, 20 audio formats without the user knowing anything about it – let's get away from one format or another, and build for a plethora] – let's give students and researchers imaginative tools to build their private collections.

- **Intelligent wrappers - fact-extraction, auto-summation, level analysis**  
we should invest in enabling enhanced collections of material that improves the key student and researcher activities of acquisition, annotation and cross-reference