Conference report

Future Present: Changing your environment with E-Books?
An ALPSP Seminar held in London, 2 November 2007

Report by Fytton Rowland

Seminar introduction

Linda Bennett, Gold Leaf

This seminar 'set out to capture the e-book initiatives with staying power from the years of experiment and innovation'. Linda Bennett of Gold Leaf had brought together a knowledgeable panel of speakers, mostly from the publishing industry as one might expect at an Association of Learned and Professional Society Publishers (ALPSP) event, and she herself lauched the meeting. Her talk, though entitled 'Aggregator platforms and publisher-specific platforms', actually gave a potted history of e-books. She noted that early on it was asserted that academics wouldn't read books from screens, whereas now it seems as if students won't read from anything else. She then went on to identify some unhelpful analogies. One was the analogy with e-journals: journal articles are written by academics to achieve professional recognition, not to be read, and not to make income from, whereas books are written to be read and to make their authors some income. Another unhelpful analogy was that with the music industry: musicians can make a lot of money from live gigs, so they do not care about publishers so much; book authors need income from their publishers, and there is concern that electronic availability of books will reduce the overall sales of titles. She then pointed out that there are a variety of frogs that publishers now have to kiss: aggregators and intermediaries, conversion companies, search engine companies, digital asset management systems, even advertising agents. Librarians hold the key to success for e-book publishers. Librarians have to decide between publishers and aggregators; between single-title purchases, publishers' big deals, and subject-specific collections; between outright sale and licensing. They dislike the wide variety of access restrictions applied by different publishers; because of this complexity, the tendency is to play safe by applying the most restrictive rules to everybody's products, even though this is to the disadvantage of both users and the more permissive publishers. To succeed in the future books marketplace, publishers will need to reach all possible markets and exploit all formats; applying old business models to the new technology will not work.

Pricing models

Hannah Perrett, CUP

Neatly following on from the end of Linda's talk, Hannah Perrett of Cambridge University Press (CUP) spoke on pricing models. The needs are for choice, flexibility and simplicity – requirements that may conflict. The models available are: subscription; perpetual access; and subscribe-to-own. These may be based on single titles, or collections, which themselves may be pre-defined or chosen by the customers. Prices may be based on number of FTE students, and may allow for varying numbers of concurrent users, or only one user at a time for any one title. Bodies influencing publishers'

decisions include higher education funding agencies and library consortia. Market demand cannot be ignored either – for example, customers want perpetual access – and the models adopted by one's competitors are also relevant. It isn't just a matter of price, either: quality of service matters, as do customer preferences over platforms and aggregators. One also needs to distinguish between ordinary purchases and consortium purchasing. For CUP the major issues to be considered were: the threat of cannibalization of print sales, leading to caution, especially as regards textbooks; the fear of undervaluing content, especially of the back-list – they refuse to discount the back-list; the need for a level playing-field among aggregators; the problem of setting prices for rapidly changing markets; and the choice of delivery platform – do you make your own? In the future we can expect: approval plan integration by library suppliers; print on demand, including course packs; pricing for institutional purchase of e-textbooks; the integration of content with VLEs; and 'slicing and dicing', with a decision needed whether or not to sell units smaller than whole chapters. Factors in the choice of business model will be: ease of management for both the vendor and the purchaser; the funding model; the quality of content; and the level of support required by customers.

Decisions made and why

Louise Breinholt, Wiley-Blackwell

Louise Breinholt of Wiley-Blackwell described some of the decision-making processes at Wiley in the field of e-books. Wiley-InterScience has about 3,200 titles online on their own platform and from the New Year will add about 2,000 Blackwell titles. Overall about 7,400 Wiley titles are available through aggregators, though these are not all in the sciences. 'Wiley Plus' contains about 170 textbooks supported by other VLE material – this service is aimed at the USA higher-education market, with typically a set textbook for each course, but does not really fit the more diverse patterns of European HE. At the moment, e-books are responsible for only 6% of Wiley's STM revenues, but the year-on-year growth rate of e-book revenues is over 60%. Why should one go digital with a book? To get away from one person reading any given book at

any one time; to make the content discoverable; to save money for the publisher; and to save money for the library. For any given title, the decision is based on whether Wiley owns the digital rights in all countries, though there is an important issue with textbooks, as already mentioned by Hannah Perrett. For now, Wiley is putting all its books except textbooks into its in-house electronic depository. HTML goes in for everything, XML for some and PDFs for some books. If PDFs are present, each chapter has its own digital object identifier (DOI), there is an HTML summary of each chapter, and the references are tagged separately. Encyclopaedias are in HTML or XML; textbooks in Wiley Plus are in XML with some material in proprietary software; and intermediaries and aggregators are supplied with web-enabled PDFs. Wiley has had its own platform since 1996 for journals, so it was logical to add e-books to it, but since customers use aggregators, they are supplied too. They distinguish between 'intermediaries' such as NetLibrary, Ebrary, MyiLibrary and Dawsons, on the one hand, and 'aggregators' such as Knovel and Safari on the other. The latter companies integrate Wiley material into their own databases, so that it looks different and has extra functionality. Safari specializes in computer science and has strong market penetration in that field. Wiley started supplying e-books in 2002 with the 'collection' model but take-up was low; the market wanted individual subscriptions, and since 2005 this is what they have sold. Over 90% of their customers go for a one-time payment. However, they also offer pay-per-view (PPV) as an alternative and sell packs of tokens to libraries to pay for the PPVs. Who sells their products? The book sales force needed a lot of training in technical aspects of e-books but now sell licences. But they also sell through third parties and intermediaries. Wiley is also looking at the possibilities for hand-held devices and i-Phones, but not in the STM area – there is promise for other areas of their business, such as travel books and language phrasebooks. The key success factors are: get your digital files in order first; recognize that there are different solutions for different market sectors; motivate your staff, not just in a specialist department but throughout the organization; have a simple business model; and make the content discoverable.

The librarian's perspective

Hazel Woodward, Cranfield University and JISC

Hazel Woodward is the University Librarian of Cranfield University, and also Chair of JISC's E-books Working Group. She provided the sole non-supplier perspective. Despite all the hype about e-books, Hazel said, progress has actually been slow. Linda Bennett had done a survey for JISC, and Ebrary had done an international one, and a number of high-level conclusions can be reached from these survey results. E-textbooks and core reading-list materials are wanted by librarians. Flexibility in purchase models is wanted, as is an affordable pricing regime. There must be multiple concurrent access to e-books for users – otherwise, there is no point in buying e-books at all. Access should be easy. There is a problem with discovery – how do librarians find out what e-books are available? We need 'E-books in print' – JISC is in discussion with both Bowker and Nielsen's – and booksellers like Coutts and Dawson's are also improving their supply of information about e-books. Publishers must get metadata right: MARC records are essential and must be provided for all e-books including those in bundled collections; MARC records in OPACs drive usage. Other drivers are: recommendations by lecturers; word of mouth; marketing campaigns; and Google. E-books need to be integrated with other library resources – users don't care where the information comes from, they don't care who the publisher is, and they don't want to go all round the houses to get their information. We need to facilitate federated searching across all resources. There are too many different business models, they are too complex, and there is too much functionality that no-one ever uses. There are also too many purchasing options. Consortia have been slow to take on e-books because member libraries' needs are so varied, but Springer has a consortium offer. Bundled subject-based packages are available from Knovel and Safari; NetLibrary has 10,000 titles available for subscription. Single-publisher big deals for e-books are not liked in the UK, and Ebrary finds the preferences for outright purchase versus licensing come out almost exactly 50:50. The main impacts on libraries are greater complexity in discovery and acquisition, cataloguing issues, and marketing and promotion of e-books to users. What inhibits the use of e-books? A lack of critical mass of titles, which leads to a lack of awareness; lack of visibility on Google, everybody's source of first resort these days; lack of promotion of e-books by library and VLE staff; and difficulties users have in both finding and using e-books. Google's attempt to make a large amount of book material, obtained from libraries, available free of charge has led to legal battles in the USA, while their competitor Microsoft is making lots of out-of-copyright material available too, as Cliff Guren would describe later on.

JISC is just starting up a major project, the E-book Observatory. Publishers were asked to propose core reading-list materials, and JISC chose from those offerings a group of titles for which national site licences have been negotiated. These will be free to all HE institutions in the UK for two years; usage will be monitored by deep web log analysis undertaken by the CIBER group at UCL. There will also be questionnaire surveys at the beginning and end of the period, and all the data will help to inform decisions about future business models for e-textbooks and other high-use material.

The retail (non-library) market for e-books

Mark Camden, Ingram Digital

Mark Camden is from Ingram Digital, part of Ingram, the American book distributor that also owns Coutts and Lightning Source. He noted that small bookshops are in decline and even some medium-sized chains are disappearing – Borders has disposed of its British and Irish stores, for example. The general public is switching to Amazon in droves as its main source of trade books, and other inroads into the traditional book trade have been made by the supermarket chains, WH Smith, small cottage industry booksellers operating over the Internet, and publishers' direct online sales. Google is used for discovery. Bookshops provide content delivery, both in-store and online, and they carry the cost of delivery. Publishers need to decide what, if anything, they are going to do to access the retail marketplace with their e-books. Initially publishers were lukewarm about e-books, seeing them as marginal at best, if not an outright threat to their profitable print book business. Pricing is very challenging, and there is a strong temptation to 'cut out the middleman'; but the public wants to buy through booksellers, and publishers need a trusted

e-intermediary, which can provide asset management, aggregation and distribution services. The more alternatives you provide, the higher your total sales are likely to be. Any publisher needs to do something, now, in the e-books business – if you do nothing, you will be left behind. You need at least to experiment. Among the requirements are having some valued e-content to sell, thinking strategically, putting some metadata on to the search engines, finding a away to achieve online browsing (e.g. by having widgets to search inside books), and finding a digital asset management/ digital asset distribution (DAM/DAD) partner which will provide a repository for your digital material. A final question: with the launch of Amazon's Kindle product, will the hand-held e-book reader finally come good, after so many false dawns? It is now being advertised for reading professional and technical books, rather than for reading novels on the beach.

Question and answer session with the morning speakers

This produced more questions than answers; the key one: How do you price e-books? A usage basis, for example, could have a seriously inhibiting effect on decisions to publish in the first place. For HE libraries, a student number (FTE) basis is likely. It is important that pricing strategies are not over-complex. Textbooks are particularly difficult – there is no way that libraries can afford to replace all the revenue that publishers get from sales of print textbooks to students. And where is the dividing line between a textbook and a reference book, anyway? A good working definition is 'the difference between having a book and consulting a book'. One firm statement: piracy really is a major problem now that e-books are finally becoming big business and students are expecting all materials to be available on the Internet. There are large, well-organized textbook piracy sites already in operation.

Conversion: what are the options?

Mark Majurey, Taylor & Francis

Mark Majurey is from Taylor & Francis (T&F), whose e-book business is growing fast, with at

least a doubling (from 3% to 6% of total revenue) from 2006 to 2007. He identified three methods of production of e-books. Scanning the hard copy is needed for out-of-print deep back-list material where no electronic file exists. Google and Microsoft have gone down this road; it is cheap and dirty, and produces page images with very limited searchability. Secondly, one can convert existing print-ready PDFs to online format; this is expensive. Thirdly, and best, one produces XML first, building in the tagging from the beginning. This requires the company to reconfigure its entire workflow, and is thus a very expensive option. There are four considerations in making your decision on which of these methods to choose. First, know your market: what should you digitize? Remember the severe risk of piracy. Second, determine whether you have the digital rights and if not, whether you can get permissions. In about 10–15% of new T&F titles, there is material to which they do not have the digital rights – usually photographs. Third, implement an appropriate production workflow. Finally, choose a digital asset distributor (a DAD): these companies provide services in four areas – storage; search and return; segmentation (customized books and disaggregation); and selling. Among existing DADs are, in alphabetical order, Aptara, Code Mantra, CPI, MPS (part of Macmillan) and Value Chain. Unfortunately there is a lack of standardization between these, and e-retailers like Amazon demand a certification of DADs. T&F's decision was to minimize their risks by using one of these firms as a partner. They wanted in-house expertise but not in-house technology. They adopted a targeted approach, converting best-sellers and titles in certain specific subject areas first. They worked on the permissions problems with this smaller group of titles, and realized that the metadata is as important as the text itself – you have to decide between ONIX, MARC and NRD; you have to get into Global Books in Print; and you need to satisfy the libraries' metadata needs.

Issues for authors and publishers: copyright, royalties and DRMs

Joachim Engelland, Walter de Gruyter

Joachim Engelland works for Walter de Gruyter, the 250-year-old German publishing firm who first

published Humboldt, Kant and Nietzsche. They are building an integrated e-content platform which they hope will be able to offer new services to authors and thus make authors more likely to offer their work to de Gruyter. He identified ways in which e-books differ from e-journals – de Gruyter uses Atypon as their platform for journals. The differences are: books have a more complex structure; they are more likely to be pirated; book authors expect royalties; and book authors have a deeper emotional commitment to their work than authors of journals articles do. Taking royalties first, this issue is more complex for e-books, since individual chapter sales and customized books are now possible; it is no longer sufficient to record sales by ISBN, you will need to use DOIs. It would be more straightforward to switch from royalties to a one-off upfront fee, if this is acceptable to the author. The emotional issue is more difficult still: especially in the humanities, books are very important to their authors and a tangible print book is critical evidence of their academic worth. Print-on-demand technology means that no book is ever out of print, a positive point for authors. Usage and citation data are measurable, and if these are high, it is important to tell the authors so. CrossRef provides valuable linking to other e-books and to e-journals. The author's complimentary copies can be supplied as PDFs, which can be 'personalized'. The piracy issue is addressed by Atypon's functionality which prevents systematic downloads, and de Gruyter feels that nothing more is needed for their e-books – obtrusive DRM would annoy the users. So far as textbooks are concerned, the question of concurrent and total users is a matter for negotiation with libraries, and they intend to experiment. Finally, there is an issue in Germany regarding open access – the German research funding bodies sometimes either forbid or time-limit toll access to reports of work that they have funded. If this requirement should extend to an e-book, perhaps the free access could be technologically restricted to accesses that came via the author's own website or the author's institution's repository only. Future possibilities for de Gruyter's e-books include anti-plagiarism checking and quality labelling from (e.g.) CrossRef; combined promotion of books by the publisher, the publisher and the author's institution; and widgets to facilitate within-book searching.

Digital asset management

Ernst Kallus, LibreDigital

Ernst Kallus from LibreDigital defined digital asset management (DAM) as 'simple strategies for complex content'. The challenges inherent in digital publishing are: reaching the market; choice between technology options; format – PDF or XML?; production processes; the cost of entry; and DAM. An idealized Internet digital warehouse would ingest and store content for publishers, taking in a single format but outputting multiple format options; it would deliver and control content, allowing limited browsing by each individual publisher's rules. Delivery could be to individuals, to institutions, to aggregators (with the aggregator either uploading the content to their own site or pointing to the warehouse's site), to content assemblers for VLEs, or to print-on-demand sites. The DRM system provided by the warehouse would be flexible, so that different kinds of customer could have different levels of rights. The content structure is important and 'X is good' – XML, XHTML and XSL are all good. These formats would permit customized publishing with varied outputs for different customers, put together from granular content. PDFs could be produced on the fly to provide print on demand. There is a culture shock involved for publishers, though – they would have to abandon the centrality of the page as a unit of content – pages should be subservient to content, not the other way around.

The future lies with content

Cliff Guren, Microsoft

Cliff Guren from Microsoft continued from where Ernst Kallus had left off, agreeing on the dominance of content: customers want trusted, authoritative content, allied to live search. Microsoft's 'Live Search' product now contains 20 billion documents. These are divided into four groups – though experience suggests that most users ignore this division and search directly from the home page. One is Live Search Academic, containing journal articles. Live Search Books scans out-of-copyright material from a number of major libraries, plus in-copyright material that is offered by its copyright owner.

Over 150 publishers have now signed up, and others are in negotiation. The publisher determines the 'preview rights', and these are assigned to each individual book and can be changed at the publisher's request. They take ONIX feeds from publishers for the metadata, and have formed a partnership with Ingram Digital for supply of material that requires payment. Microsoft's message about piracy is this: pirate sites are easy to use. Legitimate sites need to be just as easy for the customer to use: don't make it difficult for your customers to obey the law.

Final question and answer session

Only a couple of major issues not so far discussed came up in the final discussion. One concerned Wikipedia: Why is it so trusted? The conclusion was that each entry is produced by people who share the same specialized interests as the reader. The final piece of advice from the panel to publishers was that they needed to become less faceless to their end customers – they needed to get involved with social networking sites like Facebook and MySpace to acquire a positive, non-punitive image with users, who would otherwise see nothing wrong with using pirate sites.

Summary of the main conclusions of the day

- Users want high-demand, textbook-type material to be available in electronic form, but publishers have so far been unable to devise a business model that protects their legitimate interests and those of authors while providing this kind of material.
- It is imperative that a solution to the textbook problem is found, as pirate textbook sites are already active, and expanding.
- While DRM software is essential, it must not be obtrusive or punitive – publishers need to project a positive image to their customers, and not make things difficult or complicated for them.
- Quality of metadata is critical, and libraries really want it in MARC format. Widgets to facilitate within-book searching are also important.
- So far as pricing models are concerned, there is a need to offer a variety of options to different customers, but this tends to lead to a lot of complexity, and …
- Everything connected with e-books is too diverse and complex at the moment. Don't add functionalities that no-one uses. Keep it simple, stupid!