Is the future index-linked?

Stuart James

This article surveys the current environment of indexing and prospects for indexers. It reviews ICT developments in the publishing of printed and electronic books, discusses the past, present and future of indexing for electronic and printed database, and the indexing of content creation projects. It discusses user requirements and some likely future scenarios and opportunities for indexers.

I shall look at indexing from a broad perspective — ultimately a personal one. My professional background is in librarianship, but heavily influenced by, and interested in, a whole range of indexing issues over many years. I see too many similarities between librarianship and indexing to describe the professions as parallel, for parallel means separate; they come too close, even intertwine, in too many respects for that. In my time with the Library Association's Cataloguing and Indexing Group I was always conscious that we should have been paying more attention than we did to the Indexing part. I shall discuss indexing in a fairly general sense here. I am aware of the distinction between open and closed indexing; one outcome of the new and developing environment seems likely to be a further blurring of an already indistinct distinction between the two. In fact, I will focus mainly on what indexers define as closed indexing.

Throughout my working life I have been a librarian in public and special libraries, and, for the last 20 years or so, in an academic library. I have been a part of much extraordinary change, and remain so. The same applies to my sixteen years' involvement in editing two journals (one of them a reviewing journal for reference books), also during a period of enormous change in which I participated from the inside, both in terms of my own working practices, and in reviewing the results of other people's efforts. I shall draw on all these experiences in various worlds to give you my thoughts on the present and likely future for indexing and indexers. The perspectives will vary of course, and I shall try not to forget the most important one: that of the reader or end-user. I shall throw in a few personal examples and observations, some of which are specifically Scottish. If one or two examples seem very specialist, I hope they will be instructive by illuminating some specific points I am making. For ease of constructing this paper, I am adopting a schematic overview of publishing in the present and in the immediate future. This comprises: printed books; electronic books; electronic databases; printed databases; and content creation. This is, of course, an oversimplification and the boundaries between the categories, even where they exist, are both very indistinct and constantly shifting. In addition to general comments and principles, I shall be looking at them from the specific point of view of requirements, implications, challenges and opportunities for indexing and indexers.

Printed books

Starting with the printed book and its index, I begin with the good news: the number of printed books published in the UK is still increasing — more than 125,000 new titles in 2002 (a 5 percent increase); and even better news: more than 50 percent of them in the academic/professional/STM fields which, almost by definition, are likely to require indexes. So there still seems to be plenty of opportunity for the traditional back-of-the-book index. But it is not quite that simple, as indexers will realize even better than I do. I shall discuss the effects of electronic book formats on the printed book market below, but more particularly, publishing is an increasingly competitive, cost-conscious business, driven by commercial and managerial, rather than traditional publishing and printing, practices. It has not by any means always been so. There is nothing necessarily wrong with this. Indeed, if such practices strengthen the market for publishing, they are as welcome as they are necessary. Publishers are adopting new methods for book publishing and production, which are gradually — or precipitately — squeezing out the human intervention between the author, submission and approval of copy, and the final printed product. This has led to the demise of the sub-editor and of at least some editorial staff. There are plenty of horror stories around about the consequences. Simply as a reader, I am distressed by some of what I find in books nowadays which I know results from a lack of firm — or any — editing. We all know the dangers inherent in that, and can all produce numerous examples of bad practice. The problem exists not just in this country; I am presently reading a beautifully produced, technically detailed and accurate French book on a specialist subject, where I am distressed to find, even with my imperfect grasp of the language, a number of typographic and stylistic errors which should have been edited out.

Indexers will be even more familiar than I am with specific effects of all this on commissioning and using indexes. Over the years I have read — and for that matter written — numerous adverse comments about indexes in Reference Reviews. To be fair, there have also been numerous positive comments, but I have to say that the adverse comments are certainly not getting any fewer. I have reason to suspect that in many cases this is due to increased reliance on poorly edited automatic indexing. The airline encyclopedia (Smith, 2002) for example, is a wonderful source book with a very thorough index, but that index has some idiosyncratic
entries and, most annoyingly, only references the title of an entry that might be as many as 70 columns of small typeface long.

Publishing was historically an industry beset by extraordinary inefficiency. What we have been seeing in recent years, and continue to see, are the effects of standard business practices being applied, sometimes for the first time, within a context of mergers and corporate agglomerations driven by accountants and business people, where traditional publishing and editing skills take a low priority. Some of this development was almost certainly inevitable, but not, in classic business terms, to the extent of adversely affecting the basic quality of the industry's prime product. We should hope to see some reaction against that, and a reinstatement of sound editing principles in the name of maintaining the quality of the product. But that means, with a few notable exceptions, still a naive and over-optimistic view. In mainstream publishing, business and accounting practices will continue to predominate with knock-on effects on the various parts of the process, such as indexing.

That is, and will increasingly continue to be, especially evident in the application of IT in all stages of publishing and book production. These techniques make technical and, above all, economic sense for publishers, and all their suppliers or sub-contractors simply have to follow their requirements whether they like it or not. If that means retraining, or learning entirely new techniques, then all I can say is welcome to the real world. XML is becoming – if it isn't already – an industry standard, with all that it implies. The options for indexers are either to adopt new techniques (XML could well be replaced by something new quite quickly) or lose business to those who are prepared to adopt them; they might be trained indexers or people with IT skills who think indexing is easy and are able to convince publishers it is, especially if they are offering a significantly cheaper product. A worst-case scenario – probably from everybody's point of view and not just the indexer's – is to lose out to fully automatic indexing. In making the necessary adaptation, indexers not only increase their chances of preserving their market, but also might optimistically gain enough confidence from their employers that they will listen to them if they spot something seriously wrong and point it out. Listen, but not necessarily act.

The printed book has its own enormous proven merits; I have lost count of the sometimes weird and wonderful book replacement devices that have crossed my desk then disappeared without trace over the last thirty years. The printed book will continue to exist – even flourish – for the foreseeable future, interestingly as much for economic as for other reasons. New IT-reared generations might show greater preference over time for IT-based formats, but the book not only refuses to die, but apparently goes from strength to strength. One very positive aspect of business practices taking over publishing is that the book as a generic as well as a specific product is being marketed effectively, thus holding its own in an increasingly competitive market place.

In a moment I shall consider how the traditional printed book affects its electronic versions, but let us not forget the reverse process where IT products affect the structure and format of printed books. There are numerous examples, perhaps the best being contemporary children's non-fiction, with page openings looking like computer screens, fact boxes like windows, contextual strips along the edges of pages, and of course plenty of colour illustration. The good news here is that indexes, however exiguous, remain core features and are actually promoted as an integral part of the book, aimed at helping children to acquire sound information-seeking habits. Where do those habits disappear to subsequently? I shall return to that later.

Electronic books

Electronic books are still in their infancy. Nobody understands the market for them and won't for several years at least, if the parallel and more advanced process of switching from printed to electronic journals is anything to go by. Electronic journals have been with publishers and libraries for quite a few years but the real revolution has hardly yet started. The market remains highly volatile with rapid and constant change, and there is a long way to go yet before any stability is discernible, if stability is ever likely to be possible.

Inevitably, electronic books start by being based on printed books, just as journals are still based on their printed versions, even if those printed versions no longer exist. However, they will quickly develop their own formats and approaches and will, as I have indicated, continue to exert an influence back on the printed book. There is an obvious historical parallel with the switch from a manuscript book to a printed book culture. Of course, the printed book is so firmly embedded in our culture and offers so many advantages that the switch is unlikely to be total, for some time yet at least, but I can't help wondering how many scribes said the same about the coming of the printing press. One of the advantages the book has is its adaptability – but everyone connected to producing it will have to be prepared to be as adaptable themselves.

As electronic books develop they will become increasingly dependent on their structure, employing their own version of adaptability with summaries, references, glossaries, hyperlink references and the rest, making use of the power and flexibility of the online environment. There will still be the coherence found in a single work – journal titles still exist in an environment where it was predicted that the individual paper would take over as the basic unit, something it has done so far only to a very limited extent – but the parts of that work will take on a greater significance, as will their relationships to other works, or to parts of other works, or to parts of the same work.

That leads to what are known as topic maps, or what I currently think of as mind mapping software, and cannot help but feel I have been calling faceted classification for the last 30 years or so. Many of us who have looked at, or worse still, tried to navigate some of the site maps on web sites will welcome any formalization or codification of the process of structuring a web site, or even a database or published work. On the other hand there are times when I am in the middle of a strange web site and I know exactly where I am: I am lost; I might even prefer to be lost. What effect this will have on indexes in the electronic environment, as I am presently discussing it, I find it difficult to assess at this stage. When I
come shortly to discuss databases, I shall be referring to
some that are so structured that they do not — apparently
need indexing. How these techniques will develop and their
practical effects on indexing I find it difficult to predict. I
have seen so many good ideas come, but as quickly go, and
sound logical techniques finally achieve only specialist appli-
cation where a more general use would have been appro-
priate. A case in point: for many years I have championed
the notion of advanced thesauri as aids to retrieval in codi-
ifying — but most definitely not altering or 'controlling' —
natural language, but their application in real terms has been
restricted and I cannot help wondering if the same will
apply to topic maps or mind mapping. Perhaps I am being
over-pessimistic; after all, thesauri and their principles do
sound logical techniques finally achieve only specialist appli-
cations. But in these cases, if they are subsequently
marketed electronically, the greater searching power of the
electronic version is usually cited as one of the major advan-
tages. For example, the latest revisions of the Oxford English
dictionary are only available as a complete product with
powerful and sophisticated search capabilities in an online
version. There will never again be a complete printed OED.

So publishers can see added value from an at least implicit
degree of indexing which they were not prepared to
commission for the printed version. The reason is, as ever,
economic: how much would be added to the cost of the
book and would the user be prepared to pay? The
power of the computer becomes starkly obvious in these
cases, but is that power being harnessed appropriately? Is
the 'indexing' actually free-text searching, or is some kind
of categorized access supplied, which of course uses classic
indexing techniques?

The range of databases, unless following a very strict defi-
nition, is enormous, whether in printed or electronic format,
or, increasingly, both. There are many examples of printed
reference books being issued in electronic formats too,
whether on-line, web-based, or CD-ROM, raising questions
about identity, format, proper use of the new technology and
many more. Some electronic versions follow a printed origi-
nal closely with little or no additional search facility or
indexing and suffer for it. The Worldmark yearbook (Bonk,
2001), for example, has a web version that simply repeats the
printed text and offers no real search capability since every
reference takes the user to the head of the entry rather than
the searched word highlighted in its context. Others take a
printed reference work and create a whole new product
from it, perhaps by enhancing or expanding the content, but
also by expanding the indexing and search capabilities
beyond anything possible in the printed version; again, the
OED is a classic example. Others create a new printed refer-
ence work out of an electronic version: an outstanding
example, complete with an excellent index, is the latest

Printed databases

Printed or manual databases existed long before the elec-
tronic, and continue to exist either in their own right or as
by-products of electronic services. Here I shall digress
briefly into reference books; these are, after all, obvious
examples of printed database construction and indexing.
Many, even most, reference books need and have indexing
to varying degrees of thoroughness; some have no index,
perhaps because they do not need it — A-Z format diction-
aries for example — or because their compilers or
publishers think they don't need it, such as Who's who publi-
cations. But in these cases, if they are subsequently
marketed electronically, the greater searching power of the
electronic version is usually cited as one of the major advan-
tages. For example, the latest revisions of the Oxford English
dictionary are only available as a complete product with
powerful and sophisticated search capabilities in an online
version. There will never again be a complete printed OED.

My own conclusion would be that planned indexing will
always be needed, but it will not always be provided if
publishers don't see the need, or see the need but reckon the
cost too high, or if end-users find, or think they find, what
they need without them. The same applies to libraries and
the internet. Obviously, indexers will have to adapt to this
new and developing environment, both in terms of innova-
tive practice — and I have no doubt that much has been
adopted already — and in terms of their own marketing and
promotional skills. If 'proper' indexing is better, indexers
have to be able to convince both publishers and users why
and how it is better, and do so in their (commercial, tech-
nical, and self-interested) terms. Is there a lesson here from
PRECIS? That was an intellectually rigorous indexing
system that provided an elegant index for a specific but
major purpose. But it failed ultimately, for institutional and
commercial reasons, to be superseded eventually by the
technically and intellectually inferior but commercially
dominant Library of Congress Subject Headings.

Electronic databases

I have comparatively little to say about electronic databases,
although I shall have more to say shortly both specifically
and by implication, when I compare them with printed
versions. Electronic databases have largely developed in
their own terms and driven frequently by the power of IT to
provide access — or new kinds of access — to data previously
inaccessible for a number of reasons, such as the impracti-
cability or cost of making it available in anything other than
electronic format. Indexing of electronic databases — or the
decision not to index them — is determined by the nature and
use of the data they contain. Many such databases are self-
indexing by their structure and/or content; some of course
do need indexing to be in any way accessible, while others
need, or would benefit from — there is a difference - thor-
ough indexing, but that need is not necessarily immediately
apparent to publisher or user.

James: Is the future index-linked?
James: Is the future index-linked?

dition of Peter Stearns' *The encyclopaedia of world history* (2001). These developments are the future. It is natural and commercially sound to enter a new technical market with a well-established brand name reference book or books, but if nothing more is offered than an electronic version of the book, no value has been added and, given the user friendliness of the book format, an impediment has been added by demanding the medium of a PC for access. Electronic reference books, even or especially when based on an existing perhaps classic title, must offer much more than just the same text and information online, otherwise the book still wins.

The opportunities for indexers to apply their principles and expertise must be obvious; the initial barriers will be political or commercial: IT people think nobody else can appreciate what their super technology can do, but missed opportunities or poorly constructed or indexed electronic databases give the lie to that notion, and as IT becomes increasingly part of everyday life, the mystique of the IT expert diminishes at least in this respect. And there have been enough disasters for publishers and IT people alike to realize that sometimes at least they need expert and traditional assistance derived from sound basic principles established within a range of different contexts over many years.

Content creation

The last category I am going to look at seems in some respects the most significant. This is the burgeoning field of content creation: all those digitization and online learning projects of all shapes and sizes — educational, commercial, heritage, opportunistic and the rest. Libraries and museums are prominent in this movement, but there are also major commercial players, some of them new to the more traditional publishing scene, such as the global communications companies or broadcasting bodies. Basically, the rationale ranges from making secluded resources in special collections more accessible — or in many cases accessible for the first time — to gaining additional revenue from television or other programmes already produced. Of course, the types and quality vary enormously, from straightforward reproduction of images from a collection with minimal captioning or indexing (or often enough no indexing) to fully researched, annotated and indexed scholarly projects. Good or bad as the products might be, their content might be of great historical or educational significance.

Formats vary from web sites to CD-ROMs or both, sometimes using the latest IT techniques to great effect, like the British Library’s *Turning the pages* software, complete with zoom and other facilities. As far as indexing is concerned, this remains varied except in the most advanced projects. In some cases, such as the Pont Maps of Scotland project, the indexing alone is a major scholarly task: while the availability of the material is paramount, access to it both graphically, by scanning and zooming the maps themselves, or by place name and other indexes adds a whole new dimension to access and offers a model for such projects. In others too, like the Whistler Letters Project at Glasgow University, thorough indexing is integral. A similar level of indexing was applied to the Statistical Accounts of Scotland database, taking an invaluable historical resource and making it even more accessible and useful to a vast range of potential users.

It seems to be in some commercial or part-commercial projects that indexing is at its weakest: a couple of specialist CD-ROMs featuring excellent and important source material from the Royal Aeronautical Society Library — but produced by a commercial company — although reproducing originals to satisfactory visual standards, are marred by haphazard arrangement, minimal single-line captions and no further indexing beyond a simple one-line-at-a-time list of those captions to scroll up or down.

One fascinating aspect to come out of all these developments that have now taken on a new momentum with so much digitization work being carried out, is new approaches to indexing graphical or visual materials. Funding for such projects is opportunistic, almost random. In Scotland we are just beginning to look at the national coordination of provision of such digital or digitized sources and evolve some kind of policy to ensure that appropriate or important material is identified for treatment, or duplication is avoided unless essential (there are more than 800 Robert Burns websites worldwide, of which a handful are excellent), and to try and ensure that funding is directed where it is most needed.

But in the present almost free-for-all scenario, opportunities for indexers are also going to be highly variable. Some projects seek, even rely on, indexing expertise; others tack it on if funds allow. Metadata and other standards have been devised and public funding agencies normally stipulate their use. And here is another technique into which indexers need to have some input, since metadata itself is intended to be a semi-automatic means of generating not only descriptive but also indexing data. Commercial producers will apply indexing when they see the need, or where they can see a profit either in the short term from more than commensurately increased sales, or in the longer term from a coordinated content development strategy. Experienced publishers of textbooks or reference books seem more likely to recognize the need for logical construction and proper indexing as they evolve into electronic publishing within the context of longer term commercial strategies.

There are so many overlaps and differences in approach and outcomes of all the formats I have outlined, but their similarities seem more marked than their differences. Indexing challenges are involved in them all, and these can be categorized as intellectual, technical or economic. Again there is much overlap between these categories and the approach to them is ultimately political or personal as well as economic: is there commercial added value from a good index? How good does the index really need to be in relation to use of the work? How do you convince the publisher or sponsor of that value? There are similar arguments here relating to the production of metadata that have many parallels with indexing questions.

Questions also arise about providers and users: what is the level of information skills at different stages or among different categories of users? These are problems common to indexers and librarians. Users, of course, think that Google or the like find it all for them, and so they do, and an awful lot besides which is of no relevance and simply clogs
up the user's capacity to evaluate, and often leaves them opting for the easiest or first hit which might not be the most appropriate. These problems are well recognized by us, but not by the general public, including students and others who should know, or have been taught, better. They are compounded by commercial self-interest where some who want to secure major funding claim technical solutions to these problems. Those of us with many years' experience in the field know these are at best wishful thinking, at worst fraudulent, and in either case totally unrealistic and over-simplified. But these people generally have more powerful voices or contacts than do we humble experts, and they present what seem plausible cases. Later they are just as good at presenting plausible excuses while we just have to sit and think 'told you so'.

The obvious solution is to teach information skills at all levels. Libraries are doing their best, from school librarians and public librarians responsible for the People's Network, to libraries in further and higher education. The means vary from online tutorial programmes to standing and telling them, but the results are patchy and we are fighting against a strong tide of assumptions that Bill Gates and the rest have done it all already with their search engines. Whatever else they lack, search engines do offer an apparent ease and certainty of results. It takes a major problem for a user even to realize that a problem might exist, and that can apply to publishers as much as to users.

I seem to have wandered over a range of topics, some more directly relevant to indexers than others no doubt. What kind of conclusions can I draw from this? The future for indexers should be bright, with plenty of opportunity for continuing book indexing and new and ever-developing and expanding areas of database or digital content indexing. But there are demands made by the market, and not a few risks. The environment — not to mention the goalposts — is constantly shifting, technically, intellectually and economically. All three are elements which should properly change and develop; the question here is how they develop and whether those developments are always appropriate. There are major changes to be addressed in your primary market, publishing. These changes are particularly, but by no means exclusively, evident in the developing electronic publishing market. Major commercial publishing firms still dominate the market in terms of volume, but they carry their own quality improvement.

There is also, especially in the digital content field, the rise of new publishers: libraries, museums and the like as content providers, alongside a host of small specialist, even individual, publishers: it is easy enough to digitize materials and put them up on the web, but is the work being done adequately, to accepted standards, and with appropriate, or any, indexing? In producing a printed book nowadays, the physical production is probably the easiest and, apart from the cost of paper, the cheapest part. As a result, almost any subject field will produce a number of specialist publishers with access to valuable material that they are well positioned to publish, but which lack almost any editorial skills or experience. This is shown by elegant and attractive publications with accurate (if sometimes poorly written or diffuse) text, but also with numerous errors of grammar, construction and the like. Any individual nowadays can – and frequently does – create their own website and in doing so becomes an electronic publisher. Some of these sites are very valuable, often in very specialist areas. Again, the very relevant question of how sites can be identified and evaluated to separate the wheat from the chaff is one I cannot pursue now.

Much of this is being driven by external policies: lifelong learning and social inclusion mean a need for more educational material of all kinds; similarly, the move to provide increased access to heritage materials has enormous implications, not least of them for indexing. Within that context also, Scotland has its own cultural policy, backed up by development of a Scottish cultural portal by the Scottish Library and Information Council, picking up the point I made earlier about national policies to identify Scottish cultural material for online access.

There are opportunities for major commercial or scholarly projects, perhaps indexing sets or sub-sets of materials, or major databases, if they are not already indexed or self-indexing. The nature has changed from the major printed projects of the past; they are more likely to be commercially viable in electronic form to an extent they never were in printed, but they are still as real in indexing terms. Again, the electronic environment generally gives the opportunity to do more or better indexing. As techniques such as topic maps or mind mapping develop they are paralleling knowledge management techniques which have been developed in a separate context, but thereby offering further developments and new avenues to those able and willing to grasp the opportunities they offer.

So the indexer has to adapt to, or better, embrace, all these changes in order to survive or — most optimistically — to lead these changes within their own specialist field. People, especially publishers and managers, like solutions to problems they have identified and to those they have not, if you can convince them that these problems are potentially very real. Those solutions have, of course, to be appropriate and feasible, technically and economically. The ultimate benefit must be for the purchaser in the most efficient and productive use of a product, and thereby for the publisher in the market success of an effective, efficient product, whatever form that takes from printed book to electronic database. All you have to do is keep coming up with solutions in a constantly changing, hugely volatile, technical and commercial environment, continually improving the quality of your own products at the same time as reducing their costs. It's as easy as that.

This article is based on a talk given at the Society of Indexers' annual conference, A Scots Quair, held at the University of Strathclyde, 27-29 June 2003.

References
Indexing the future of information

Glenda Browne

This article describes changes that have occurred recently in the work indexers do, and considers the implications of these changes for our future.

The Australian Society of Indexers has run five conferences over the past ten years. Two themes have emerged – the importance of partnerships, and the need to consider our role in the electronic publishing environment. Our first conference, run by the Victorian branch in Marysville in 1995, was titled ‘Partners in Publishing’, while the most recent, run by the ACT Region branch in 2001 in conjunction with the Canberra Society of Editors, was called ‘Partnerships in knowledge’.

The second conference, run by the ACT Region branch in wonderful old Ranelagh House in Robertson (NSW) in 1996, was titled ‘Electronic Age’, focusing on our role in aiding access to information in electronic documents on the web and on CD. The first NSW conference, held in Katoomba in 1997, was called ‘The FutureProof Indexer’, as we wondered whether we would be made redundant by computer search systems.

Fortunately we discovered that we were still needed – in fact, we were crucial for quality information access – and the Victorian branch felt confident enough to return to the erudite essence of indexing with ‘The August Indexer’ (as in ‘refined’ or ‘imposing’), held in Hobart in August 1999. Our current conference has the very positive title ‘Indexing the World of Information’ as we consider not only the continuance of our traditional roles, but expansion both to new opportunities and to work on information around the world.

I believe the future is already here, in that significant changes in information provision have occurred in the last few years. There have been great technological developments. As I remind my kids: ‘When I was your age, I had to walk to the TV to change the channel!’ On the other hand, some of the change has been overwhelming. As humorist Ashleigh Brilliant said: ‘There has been an alarming increase in the number of things I know nothing about’. For us these new concepts are mainly reflected in acronyms. Are you up to scratch with XML, XFML, XTM, taxonomies, ontologies, OWL, RSS, CSS, CMS, DTD, DRM, DC, Semantic Web, RDF, RTF, PDF, DOI, ROI, EAD, wikis, NIPs, and SEO?

The things I think are required of indexers and indexing now, and which will be more important in the future are:

- Single sourcing using XML
- Embedded indexing
- Teamwork/Distributed content indexing
- Innovative specialized indexing/Globalization
- User testing

And the most important thing in all the work we do will be networking with other indexers, and communicating with a range of professionals to ensure that the work we do fits into the environment we are working in.

Single sourcing/multipurposing

Single sourcing means creating a document once, in such a way that it can be output in a number of different ways as required. This often means creating the document in XML format and outputting it in different print and online formats. It can also mean outputting different levels of detail, for example, a briefer version for children. XML is more powerful than HTML for this purpose as it separates the content of a document from its display.

Notes

1. The British Library’s Turning the pages software may be viewed at www.bl.uk/collections/treasures/digitisation.html
2. The Pont map website may be found at www.nls.uk/pont
4. http://edina.ac.uk/statacc
5. The People’s Network website may be found at www.peoplesnetwork.gov.uk
6. The Scotland’s Culture portal can be found at www.scotlandsculture.org