

The Paradox of collecting in a digital age

Presented to the
Asian Library Conference on Collection Development
(ALCCD 2002 Penang)

Steve O'Connor
CEO, CAVAL Collaborative Solutions

ABSTRACT

This paper both acknowledges the strong desire by users and librarians to collect and deliver materials in digital form and, nonetheless, highlights the challenges and opportunities in this ambition.

There are challenges in the limited range of material 'born digital' at this time and in the future. There are huge opportunities for libraries to work collaboratively on digital collection development to achieve more than homogeneous digital collections across the globe. This paper will evidence examples of this happening.

These issues are especially important in the geographic areas of South East Asia including Australasia where the representation of digital content is not strong and the cultural and language interests of the region can be too easily lost. Technologies which allow libraries to create new futures will also be evidenced.

It is such a normal expectation today that, as librarians, we see libraries as being digital in focus, in both collection and in delivery. It has been, however, a quite recent development. Very few libraries had this image projected in the early years of the 1990s. Very few indeed. It is really only within the past 5 years that we have seen the bulk of libraries project this image to their clientele. So with a very recent development comes a great deal of uncharted territories and issues. These issues confront us all at many levels but there are particular issues which are still bound in the geographic nature of the environment we live and

The paradox of collecting in a digital age

operate within. It is my intention to explore some issues at both the global level and then at the geographic level but both within the collections area.

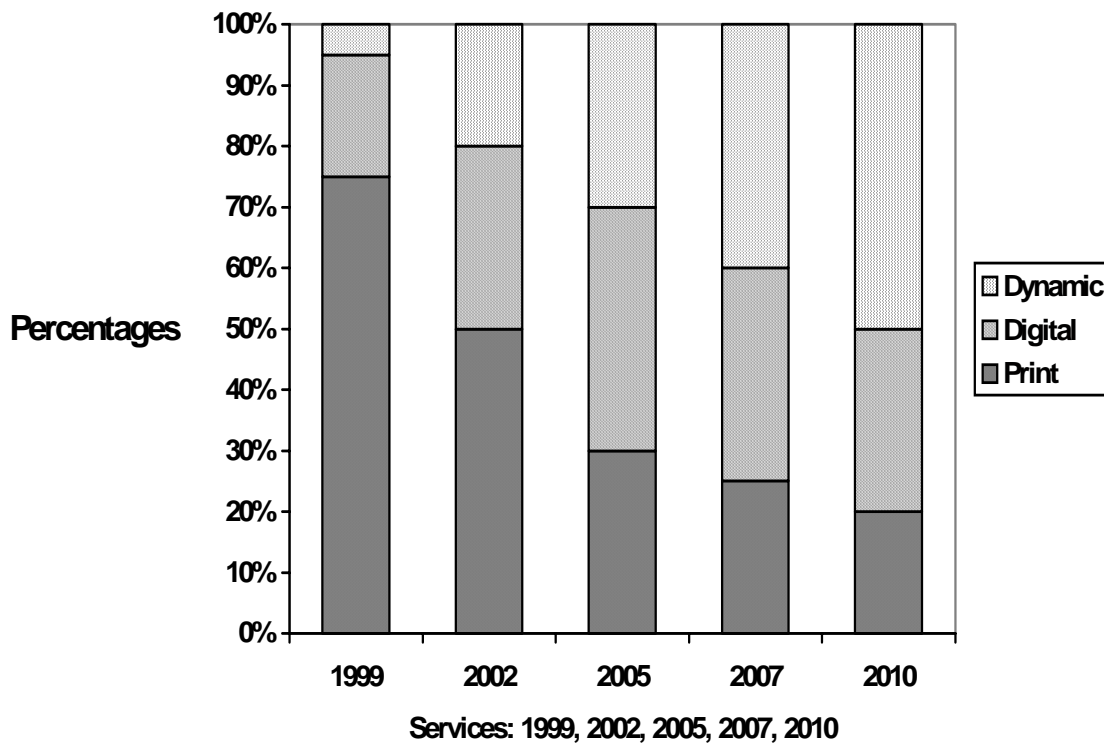
Rapidity of change

Any scholar of the future will tell you the future knows no boundaries and certainly no limits. Recently I worked with some colleagues¹ to determine the mix of media in which information is delivered now and into the future. We saw three main streams in which information is and will be delivered. This was not meant to include all technological conduits but only the generic lines. The three lines identified are print, digital and dynamic. Print has been the dominant form of information delivery in our lifetime. A huge repository of information and knowledge is stored in this form currently in libraries across the country. A number of informal analyses were used to arrive at the split of material arriving in academic libraries generally. It is clearly the case that most of the material arriving in the academic libraries at this time is still print. The forests are still contributing hugely to the presentation of ideas onto our library shelves. The estimate was that around 75% of materials are currently being delivered in print form while only 20% are in digital format and the remaining 5% are dynamic or web-based. Through this work I estimated that in 10 years time the delivery of material would be sharply different with 50% delivered via the web, 30% in digital form and 20% remaining in print.

My suspicion is not that the projections are wrong but that they will be arrived at much earlier time. This is based on my experience of the rates of change. Our capacity to predict the future is generally rather poor and under-estimated. Regardless of whether any of these predictions are correct or wildly wrong it is crucial to plan for a different pattern for the delivery of information both to and through and perhaps around our libraries. We cannot assume that libraries will remain the cornerstone for the access and delivery of information. The central role and power of libraries has been significantly eroded over the past 10 to 15 years. Perhaps that situation will be redressed but it most certainly will not be to that which was in existence before.

¹ Building the Victorian Academic Digital Library. A report for Caval Collaborative Solutions by Ian MacCallum, Brenda McConchie, Sherrey Quinn and Jan Whittaker. Melbourne, CAVAL, 1999.

The paradox of collecting in a digital age



Our Asian and Australasian region is in a unique position, largely shaped, as it has been by geographical isolation from the large publishing centres with relatively small publishing efforts of our own and generally small purchasing capacities. The impact which these factors have on collecting in a digital world need to be explored further as does our capacity to represent our own cultures in our 'born-digital' collections.

Australia and collection practices

The dual factors of distance and a small but geographically disparate population have inevitably shaped the policies and cooperative spirit amongst major libraries in Australia. Trends in collection development therefore have drawn on the British and American experiences but have not reflected either. The British experience has been one with a largely central-focus. This can be seen in the reliance on the British Library's large facility at Boston Spa. The American experience has seen a variety of responses within different states and also within the libraries with very different levels of resource to our own. There have been more examples within the American experience which have related to developments in Australia. The real differences have been

The paradox of collecting in a digital age

the high levels of top-sliced funding available in individual states and the availability of philanthropic funding to assist developments. Australian developments have been largely organic growing out of regional and special groupings. Examples of this are subject collecting programs with the NSW University Librarians, and the Metropolitan Public Librarians subject specialization scheme.

I know that there have been various projects within the South East Asian region with a strong governmental push toward a digital society. The report Social Equity and Prosperity: Thailand IT policy into the 21st century² highlights to size of the task ahead to provide the basic telephonic communication infrastructure across rural Thailand. There is a goal to produce educational courseware but otherwise the focus is solely on the technology itself. The excellent paper "Digital divide in Southeast Asia"³ highlights the eight Cs put forward by Madanmohan Rao⁴ for bridging the digital divide in the internet economy. The C for Content is especially important in the context of this paper. The remainder of the Consal Briefing paper deals essentially with the roll-out of the technology to our communities. The extension of the telecommunications system in Indonesia has been rapid of late but still lacking comprehensiveness and internet access for many.

There are good strategies, such as the National Library of Australia's Pandora Project for collecting that which appears on the net. There is, however, clearly an absence of appropriate strategies dealing with the complexity of encouraging the presence of content on the net from our cultures, communities and languages.

There are reasons for this: "local content...is often inaccessible for a variety of reasons—having been developed for a very local audience, for example, or on a system incompatible with current technology, or using standards rendering it difficult to integrate. It's probably true also that there are other factors...to do with things like the perceived benefit of wide dissemination of information, whether its cultural, political, financial or whatever."⁵

² www.nitc.go.th/it-2000

³ www.consar.org.sg

⁴ www.isoc.org/oti/articles/1000/rao.html

⁵ Private communication with Amelia McKenzie Email 24 May, 2002

Australian material 'born digital'

In recent analyses of collections in Victorian academic libraries, it has been determined that there are close on 250,000 serial titles held in these libraries. Of these some 47,000 are duplicated twice or more. But of these only 8% are paid subscriptions. Of these titles only half are available in electronic form as determined by a comparison with ULRICH's in electronic form. Nonetheless, these serials cost approximately half of the monies spent on serials in the state.

Serials in CAVAL Libraries	
Titles Held in 1 Library	198,090
Titles Held in 2 Libraries	24213
Titles Held in 3 Libraries	9736
Titles Held in 4 Libraries	5533
Titles Held in 5 Libraries	3356
Titles Held in 6 Libraries	2100
Titles Held in 7 Libraries	1261
Titles Held in 8 Libraries	583
Titles Held in 9 Libraries	145
Titles Held in 10 Libraries	29
Titles Held in 11 Libraries	7
TOTAL Duplicates	46963
GRAND TOTAL	245,053

Of these serial titles it is unclear how many specifically are of Asian or Australian origin. It would be surprising if the figure were high and even more surprising if the figure for 'born digital' were greater than 5% of all the serial literature which the academic communities are interested in. In the case of monographs it is even clearer. There are virtually no monographs from Australia being 'born digital' and I presume that this would be the same for the broader region. So what are the implications to be drawn from this?

Collection management in a digital age

As far as Victoria is concerned the implications for collection management are very simple.

The paradox of collecting in a digital age

Firstly, there are a great number of serials held across the state but that only a very small percentage are currently paid subscriptions. Of those which are paid subscriptions, half are duplicate across the libraries and only half of these are available in digital form. With half the total funds being spent on duplicate serial subscriptions it is clear that there is a homogeneous collection growing across the state. There is little diversity in collections.

Secondly, the amount of material being 'born digital' from European or American sources is only about 5% of published serial output. The percentage for monographs is very considerably less. It must be concluded, with libraries in Australia now expending 35 to 45% of their acquisition votes on digital materials, that the diversity in serial collections is very much less than what it may have been even 10 years ago. It is also seen that the domination by certain cultural regions is very strong.

Thirdly, there is a great amount of serial material held across the state of Victoria which are either closed or dead subscriptions. Many of these are duplicate as well. While it may be important to hold many of these titles for posterity it should be questioned as to whether it is necessary to hold more than one complete set in the state or even the country. There is a capital expense issue here as these collections occupy a lot of prime shelving space.

Fourthly, the poor purchasing capacity of the Australian currency is resulting in further massive cancellations of serials across the academic and state library domains. This will only further highlight the shrinking spread of serial titles and as more and more of the acquisitions budgets are consumed by voracious serials, there will be less available for monographs. Not only does this impact on the purchase of monographs purchased overseas but especially within this South East Asian region, including Australasia.

Fifthly, it has been calculated that in the early 1970's that over 70% of the world's English language published output was held somewhere in Australia but by the early 1990's this percentage had fallen to less than 50%. There have been strong programs to collect Asian published materials but these too I understand, from anecdotal advice, are suffering. CAVAL has a very strong program cataloguing multi-lingual materials across Australasia. Much of this material currently acquired is of a popular nature rather than from scholarly sources. The National Bibliographic Database KINETICA

The paradox of collecting in a digital age

records the following information about holdings across the Australia of materials in particular languages:

Language	Mono titles	Serial titles
Bengali	6,992	476
Burmese	3,473	216
Chinese	218,337	12,710
Hindi	17,886	674
Indonesian	73,791	11,379
Khmer	1,048	113
Korean	50,949	4,852
Malay	10,953	993
Sinhalese	251	44
Tamil	7,614	252
Thai	24,910	1,647
Vietnamese	41,056	1,018

In addition, a CJK data base is maintained by the National Library of Australia for materials listed in the vernacular. The following data applies to that database:

Language	Mono titles
Chinese	351,000
Japanese	848,040
Korean	83,960

Because of the different approaches of each database there may be some overlap but not significantly. Most of the CJK database materials would be housed in academic libraries.

This represent only a very small percentage of the print published output of these countries. The further impact in a digital environment must be even worse. According to the Financial Times⁶, "Chinese will top English as the most-used language on the web by 2007". This is according to forecasts by the World Intellectual Property Organization who also indicate that currently " a slim majority of the world's 460 million-plus Internet users are from English-speaking backgrounds, but by next year most internet users will have a mother tongue other than English, and by 2003 a third of users will be communicating in another language online."

⁶ Financial Times 7 December, 2001

The paradox of collecting in a digital age

This is building a picture of the collection management perspective, in a sense, as a precursor to action in the collection development arena. Whatever action result cannot be incremental on what has happened in recent years. A radical re-think is necessary for effective action. It will be necessary to create a new future and mode of action for libraries. This will require scenario planning for a future which is not connected with our present strategies and actions. It will be at least three years out from today and will be designed for a new future mode of collection development.

Collection Management

This paper has argued that in moving into a digital collection mode for our library systems there are a lot of issues to be taken into account of which some of the main issues are explored below.

Homogeneity

It has been demonstrated earlier in this paper that the collections being developed in our major libraries as 'born digital' collections are increasingly homogeneous and indeed that the range of material available as 'born digital' materials is still quite small and culturally narrow.

The main collection development issue is how to retain our own cultural identities in a digital environment. Our libraries over the years have actively collected materials in our own languages and cultures within our region. This will no doubt continue and, in a sense, is not the issue. The issue will be in the materials which will be widely available across the internet, particularly emanating from our libraries. If the economics of making a selected range of materials available in digital form stays as it is, then the publishing houses of Europe and the United States will dominate. It is interesting to note the CONSAL Topical Brief which is devoted to 'Library Genre—Science Fiction'⁷. The point is made that it is important to publish even as Robert Heinlein puts it " no matter how badly it is written [it] has a therapeutic value

⁷ <http://www.consal.org/resource/brief>

The paradox of collecting in a digital age

because all of it has as its primary postulate that the world does change". Unless we have the means of not only publishing on the net but also making the world of readership aware of it then we will end up with a very homogeneous culture globally. Our collections are littered with collections of small publishers who have had disproportionate influence on their respective cultures.

Cost of digital publishing⁸

It is apparent that the old boundaries or definitions of traditional publishing such as 'encyclopedia', 'journal' and so on are being aggregated simply as content. The costs of getting this content to the digital domain are very significant for the publisher while there are real savings at the library house-keeping end. The process of creating digital text is an extension of the traditional process where certainly much have been in electronic form before being prepared for the printed page output. The additional costs of creating 'born digital' text are compounded by the costs of SGML mark up (up tot 30% more or around US\$4.00 per page) as well as the costs of the server environments needed to maintain that on-line access. Distribution issues have now become technical access issues. The costs of maintaining IP addresses and resolving technical difficulties creates a whole new area of cost. Whether the traditional process is more expensive than the digital process remains largely undetermined. The costs at the publisher end and the library end also merit some debate.

A new issue in this publishing process is timeliness. It is reported in researches by Ingenta (UK) and Tenopir and King (US) that much of the reading of scientific literature occurs within the first six months after publication. The pressure and therefore costs to get literature onto the readers' screens can only further exacerbate costs. Establishing appropriate licences and licence conditions is another area of concern and especially if libraries, as they should , move into collaborative models by which to access these digital resources.⁹

⁸ I am indebted to John Cox Associates (www.johncoxassociates.com) for many of the ideas in this section. They have been communicated to me through conversations and electronic communications.

⁹ Again the licences developed by John Cox Associates (www.johncoxassociates.com) provides the best framework to developing workable legal documents in which both library and

Implications for our region

Despite living in what McLuhan described nearly forty years ago, as the “global village” we are still very much bound by our geography and languages within those boundaries. The electronic allows us to transcend geography and to create, for a time, what Meyrowitz¹⁰ called “no sense of place”.

The presence and transmission of quality writing and information from our own regions is strongly threatened in the digital arena by the dominance of the European and American publishing houses and their financial capacity to dominate. The percentage of print published materials from the region to Australia’s near north held in Australia’s libraries relative to the actual published output is not known but is likely to be quite small. Nonetheless there is some existing capability to learn from these materials. As our new generations come to rely on the internet for their information they display characteristics such as¹¹:

- * Constantly connected
- * Learn by doing
- * Comfortable with trial and error
- * If it is not on the web it does not exist
- * Multi-tasker
- * Uses collaborative approaches

As cultures in the more isolated parts of the globe we run the severe risk of having our cultures as expressed through our published works marginalised or even made extinct.

Publishing at the margins

The costs of developing the technologies to deal with the complexities of digital publishing, intellectual property rights, online access, marketing is very expensive. There is a new type of publishing house being developed to allow for small

publisher can achieve a proper balance. CAVAL has developed a licence for collaborative access to agreed digital resources through VADL (http://www.caval.edu.au/parties_and_research/VADL.html).

¹⁰ Joshue Meyrowitz *No sense of place* New York, OUP, 1990.

¹¹ Drake

The paradox of collecting in a digital age

print run publishing which also allows simultaneous digital publishing with full rights management systems and the commercial aspects to achieve access. The C-to-C(creator to consumer) system established by Common Ground is a very powerful example of a system designed to deal with conventional short print runs yet to deliver digital to the internet.¹² With their insistence on the use of standard approaches to data they can provide the detail of the newly published texts easily and seamlessly to distributors such as Amazon. Distribution and marketing are as central as the actual act of publishing but to also include Intellectual Property Rights Management and the commercial aspects to pay for a published copy or to access the electronic copy completes the package.

Conclusion

In a real sense there is no conclusion to this paper. It is a work in progress. Developments will occur, technology will take shape, the digital divide will be addressed and implemented but if what we have remaining is a very limited range of materials from our respective cultures then we will have failed. Part of our traditional task as librarians has been to collect materials representative of our cultures, whether substantial or ephemeral. In this digital environment this paper has been arguing that perhaps we have a role generating or stimulating the means to publishing in small numbers on the web. The strategies by which this could be achieved will be developed another day. Whether this occurs or not, the issue of the representation of each of our cultures on the web will remain the most pressing collection development issue over the next five years. A collaborative action agenda should be developed from the discussions emanating from this conference

¹² Common Ground technologies and development can be followed at www.C-2-Csystem.com and www.CommonGround.com.au

The paradox of collecting in a digital age