

Casting Forward; collection development after mass digitization
Or
Doing One's Part; thinking globally, acting locally

A paper for the Fiesole Collection Development Retreat, March 18th 2004

Michael A. Keller
University Librarian
Director of Academic Information Resources
Publisher of HighWire Pres
Publisher of Stanford University Press

© 2004 Michael A. Keller

This talk is offered in homage to our dear, departed friend, Mario Casalini and his children, Barbara and Michele, who have so ably followed in his footsteps.

Before arriving in Florence late last week, several currents of conflict, competition, and collaboration occupied my thoughts about this talk. Among the items of conflict and competition are the flurries of activity involving the stm journals crisis: open access; “big deals”; mergers & acquisitions of for-profit companies; and the D.C. principles issued by 45 not-for-profit publishers, most of them clients of HighWire Press. My colleagues and I have been taking the next steps following the February 25 resolution of Stanford’s academic senate urging libraries on the one hand to insist on title-by-title selection and rejection of bundling or packaging of desired titles with undesired ones. That same resolution urged Stanford faculty not to place their articles nor provide their editorial or refereeing services with exploitive for-profit publishers.

Several of Stanford’s collaborative projects have been developing nicely in the past few weeks. These collaborations involve the Matthew Parker Library at Corpus Christi College, Cambridge, the World Trade Organization’s archive of the General Agreement on Tariff and Trade, and the Digital Library Federation’s Distributed Open Digital Library project. The Matthew Parker Library is a little known, even less used collection of over

500 English manuscripts dating from the 7th century to the 15th century documenting for Henry VIII a lineage for the Church of England separate from the Church of Rome. Stanford is collaborating with Corpus Christi to digitize the manuscripts and make them widely available in a new sort of virtual research and teaching space. A videoconference last week with colleagues at the World Trade Organization brought a few steps more progress toward releasing a digital archive of documents of the General Agreement on Tariff and Trade. And I have been working with colleagues in the Digital Library Federation, Deanna Marcum and David Seaman, to get past the conceptual and political stages of a project we now call the DODL, the distributed, open digital library, to the design development and prototype stages. DLF's DODL will make it possible for easy discovery and retrieval of digital objects across the boundaries of many institutions and will stimulate the creation of more digital collections for academic purposes.

These three projects illustrate some of the properties of what may be the next phase of the brave new world. The Matthew Parker Library project shows how many more people might make use of unique, precious, and really quite delicate items of huge cultural significance thanks to the creative application of information technology. The WTO/GATT project will demonstrate how careful digitization projects operating on recently developed archives might contribute to widespread understanding and debate on major questions like globalization of trade, commerce, and labor. And the DLF DODL project should make apparent how libraries might share unique collections of digital objects for mutual benefit and for the general improvement of scholarship and teaching.

Among these notions and projects that fit more or less neatly in their categories, there is one process that crosses a number of them, with implications that could change everything that librarians, publishers, and others in the information trade are presently doing. That one item is, of course, mass digitization. We are at the cusp of projects that will dramatically increase the amount of carefully selected, extensively validated, widely consulted, and often cited information available through the Internet. Only a few considerations hinder us from providing our readers, our students, and our faculty with vastly improved opportunities to search, to read, to cite, and to exploit the digitized contents of our libraries. And for publishers, those few considerations hinder the next level of exploitation of their publications, whether issued in the past decade or the past centuries. Significantly those hindrances are not mechanical. I am quite convinced that

we could digitize the 8 million volumes in Stanford's collections during the next 10 years. And I am quite sure that we will figure out how to maintain the resulting database of over one-and-a-half petabytes. While we are puzzling our way through the hindrances of bringing intellectual property rights holders into the project and organizing access in numerous old and new ways to the new, digital forms of our collections and others like it, we are building experience on such well functioning operations as that of HighWire Press. We – the publishers, HighWire Press, the Institute for Scientific Information, and a fourth party that must remain anonymous for now – are in the act of digitizing all the backsets of the journals associated with HighWire Press and providing much the same level of access, searching, and linking as with the front-set of those same journals.

All these currents, these pre-occupations and concerns, had suggested to me a theme for my contribution to this Fiesole Conference, a theme that might best be expressed as an open question: What should libraries be collecting for the long term once several mass digitization efforts are complete with the contents of those large research collections available in digital form, and after the conversion of scholarly communication to mainly network-based publication is largely complete? Stated another way, what should be the roles of libraries once the basic literature of virtually all fields is accessible in digital form to one and all readers everywhere?

Now, after several days in Florence, that open question seems to me to be especially apropos. Here in the cradle of the Renaissance, the birth of humanism, and the invigoration of rationality...here in the city of Galileo, Michelangelo, Vespucci, Brunelleschi, Monteverdi, Donatello, Ghirlandaio, Dante, Vasari, the Medici...the value of cultural custodianship performed by librarians and publishers alike, aided and abetted by booksellers and others in the information trade is especially apparent. In this city with so many libraries, archives, museums, and other cultural entities, there could hardly be more strong support for the rationales of collecting, for research and study, for preserving, and for making accessible through publishing of the results of creativity and scholarship.

So it is that theme of the eternal verity of institutional collecting that I want to discuss with you tonight, despite the contentiousness of some of the issues we face today. I want to remind all of us of the fundamental values and functions we are expected to perform, even in the face of recurring struggles for our time, our money, and our attention.

Matthew Arnold, in *Dover Beach*, expresses in poetic ways some of what I want to say about distraction from collecting for scholarship, collecting for better understanding among all of our tribes and nations, and collecting for the long term. Let me read that poem for you and then dwell on those underlying values that we might too often forget in dealing with the quotidian dull cares of our professional lives.

The sea is calm to-night.
The tide is full, the moon lies fair
Upon the straits; on the French coast the light
Gleams and is gone; the cliffs of England stand,
Glimmering and vast, out in the tranquil bay.
Come to the window, sweet is the night air!
Only, from the long line of spray
Where the sea meets the moon-blanch'd land,
Listen! you hear the grating roar
Of pebbles which the waves draw back, and fling,
At their return, up the high strand,
Begin, and cease, and then again begin,
With tremulous cadence slow, and bring
The eternal note of sadness in.

Sophocles long ago
Heard it on the Aegean, and it brought
Into his mind the turbid ebb and flow
Of human misery; we
Find also in the sound a thought,
Hearing it by this distant northern sea.

The Sea of Faith
Was once, too, at the full, and round earth's shore
Lay like the folds of a bright girdle furl'd.
But now I only hear
Its melancholy, long, withdrawing roar,

Retreating, to the breath
Of the night-wind, down the vast edges drear
And naked shingles of the world.

Ah, love, let us be true
To one another! for the world, which seems
To lie before us like a land of dreams,
So various, so beautiful, so new,
Hath really neither joy, nor love, nor light,

Nor certitude, nor peace, nor help for pain;
And we are here as on a darkling plain
Swept with confused alarms of struggle and flight,
Where ignorant armies clash by night.

Matthew Arnold sets the constant nattering and clattering of the world, “the grating roar of pebbles which the waves draw back and fling, at their return, up the high strand” and ignorant armies clashing in the night in opposition to the eternal attributes of the world and of life, the tides and the sea, the cliffs of England, faith and love. The parallels in our lives, I suppose, are the STM serials crisis, constant budget worries, and the fear of being overtaken by the Internet and, as a counterpoint, the too little celebrated roles we play in gathering in the records and expression of our civilizations for teaching and for study as well as making those cultural items accessible for our own and succeeding generations.

Here in Florence, looking out every morning on the Duomo, it has become more obvious to me that preoccupation with the journal literature of scientific, technical, and medical fields, 90% of which has a half-life of under 12 months, distracts us from literatures of more enduring qualities. And, ironically, as more of that STM journal literature is available on-line, we could be turning to an article economy, and just-in-time mentality toward it, rather than succumbing to the vapid blandishments of lots of access to spuriously important literature for the sake of seemingly improved title counts and measures of “hits” by unwitting surfers. That succumbing to those blandishments diverts vast sums of money otherwise available for publications of lasting value and for archives

or collections reflecting uniquely on otherwise forgotten histories of our institutions and environments is a tragic miscalculation. Even in the context of the stm disciplines themselves, such a miscalculation reduces access to better, more cited literature published by scholarly societies.

Maybe the murmuring of librarians and a few evangelists among scientists about open access will create more understanding of which publishers are better aligned with the research interests of the academy. Maybe the Stanford resolution, the DC Principles, and the refusal of Harvard and MIT to countenance more big deals will help ring down the curtain on a period of aggressive profit-taking by some.

While all this is happening, we should turn to what we might do in collecting for the long run.

Imagine a time when the stm journals crisis has been resolved and mass digitization of public domain and protected books and other copyrighted genres has been accomplished, with due observation of intellectual property rights. Imagine in that time that there are truly improved opportunities for the use of intellectual property formerly limited by the tyranny of ink-on-paper formats. Imagine that time when students and scholars at large and small universities and colleges around the world can gain access to the concatenated collections of several very large libraries and thousands of smaller, more specialized ones. In that time, what should we be collecting then and what will our readers have expected us to collect up to that time?

From my perspective, we should be collecting now and should then be collecting, at each of our institutions, the unique and special. If some major slice of the world's literature is available via the networks, we therefore should be collecting the records of our lives, our institutions, our hobbies, arts, and crafts that would otherwise disappear. We should look for the diaries, the drafts, the sketches, and versions of our artists, poets, and novelists. We should concern ourselves with the creative outputs of designers and architects, of those who are experimenting with the expressive possibilities of technologies and new media. We should worry about the newspaper archives and collections of photographs that might get pulped or discarded, because they no longer offer much to the quarterly reports of profits. We should concern ourselves with local history, with local

manufacturing, agriculture, trade, and transport. We should offer fire departments and women's clinics, local chamber orchestras and soccer leagues a place to deposit their records. We should let our government agencies know that their records are important, potentially vitally so, and that we expect them to either take care of those records or let us care for them.

I am not suggesting that we each collect everything in sight in a radius around our locations. I am suggesting that each library ought to consider what in its reach, whether in some nearby location or through its connections, is likely to be most revealing, most important, most worth considering through the prisms of history. And then, having set up a mechanism for such considerations to occur on a rolling, more or less continuous basis, I think we ought to acquire, preserve, and make accessible the items, collections and archives highest on our lists. Having acquired the stuff, then we need to provide at least a reasonable level of meta-data about it so that people everywhere can discover what we have and make use of these special collections.

I am suggesting, indeed, that we are too taken up with the details and mechanisms of the Internet as well as concerns of our guild, and too little taken up with our contributions to our society. We need to turn from the nagging and vexatious stm journals crisis, put it in its proper perspective, and re-focus upon acquisitions that have a longer-term possibility of return on investment. To the degree that the brave new world I think is emerging will actually exist, we need now to imagine what our contributions could be. There may be several different scenarios, of course, but clearly it is time for library leaders and library professionals to develop the scenarios, consider their implications, and devise the means to determine which scenarios are to become real. If there is agreement that there are significant social roles for libraries and librarians, not just professional or guild roles aspirations, then this sort of imagining is timely. And if our collective imagination then yields better collecting for long-term value, then we might realize in library, archival terms those brilliant opening lines of Samuel Taylor Coleridge's "Ballad of Kublai Khan":

In Xanadu did Kublai Khan
a stately pleasure-dome decree,
where Alph, the sacred river, ran
through caverns measureless to man

down to a sunless sea,
so twice five miles of fertile ground
with walls and towers were girdled round.
and there were gardens bright with sinuous rills,
where blossom'd many an incense-bearing tree.
And here were forests as ancient as the hills,
enfolding sunny spots of greenery.