DIGITAL ACCESS
TO
BOOK TRADE ARCHIVES
DIGITAL ACCESS

TO

BOOK TRADE ARCHIVES

Edited by

Berry Dongelmans
Ad Leerintveld
Adriaan van der Weel

2002
CONTENTS

Berry Dongelmans, Ad Leerintveld, Adriaan van der Weel
INTRODUCTION 7

David McKitterick
THE HISTORIAN’S VIEW 1

Daniel V. Pitti
ADVANCED TECHNOLOGY AND THE HISTORY OF THE BOOK 13

Meg Sweet
INTEGRATED ONLINE ACCESS AT THE PUBLIC RECORD OFFICE 41

Jutta Weber
MALVINE, LEAF AND KALLOPE: SOME CO-OPERATION MODELS 49

Simon Eliot
TWO CATALOGUES, THREE PROJECTS — AND A TENTATIVE PROPOSAL 69
ACKNOWLEDGEMENTS

Besides the authors of the papers included in this volume, the conference organisers wish to thank the following people for their generous help and valuable contributions: Dr. Lotte Hellinga of the Consortium of European Research Libraries; Professor Pieter Obbema, former curator of rare books, Leiden University Library; Hanno de Vries of the Dutch National Archives; Marieke van Delft and Marco de Niet of the Koninklijke Bibliotheek (the national library of The Netherlands), The Hague; Sarah Mahurter of Bookhad, London; Johan Vanhecke and Jan Stuyck of the Archief en Museum voor het Vlaamse Cultuurleven, Antwerp; Dr Kurt De Belder of the University Library, Universiteit van Amsterdam; Antoinette Kerkman and Marieke Dekkers, research assistants at the Leiden Centre for the Book. The following institutions have helped to fund both the conference and the publication of the present volume: Koninklijke Bibliotheek; Arts Faculty of Leiden University; Electronic Text Centre Leiden; Constantijn Huygens Instituut voor tekstedities en intellectuele geschiedenis, The Hague; De la Fontaine Verweij Stichting, Amsterdam; Scaliger research institute, Leiden University.

BERRY DONGELMANS (Centre for the Book, Leiden University)
AD LEERINTVELD (Koninklijke Bibliotheek, The Hague)
ADRIAAN VAN DER WEE (Centre for the Book, Leiden University)
INTRODUCTION

In the late Spring of 2001 some 45 delegates from Belgium, Great Britain, Germany, the United States and the Netherlands convened in The Hague, The Netherlands, for a two-day conference on Digital Access to Book Trade Archives. This book collects some of the papers presented on that occasion. The conference, organised jointly by Leiden University’s Centre for the Book and the Koninklijke Bibliotheek (the national library of the Netherlands), aimed to explore ways in which international access to book trade archives scattered around the globe could be improved. The participants represented the three most closely interested parties: the keepers (librarians and archivists) of archival materials relating to the book trade; researchers (especially book historians); and the people who are going to be vital in providing the technical and intellectual infrastructure.

That there were many people eager to collaborate in such an undertaking, and to share their expertise and resources, was a foregone conclusion. But the actual impetus for the conference came from the technological advances made in recent years, which have brought ideas voiced before as ideals into the realm of possibility. Distributed union access to primary resources wherever they may reside would have sounded near impossible only a few years ago. Though it certainly remains a major challenge today, this is no longer for technical or intellectual reasons. Rather the challenge today is one of organisation and politics; what are required most are vision and patience.

With the infrastructure provided by the networked computer, the world lies at our feet in a technical sense. And if we have learned anything over the last decade it has been that any scheme that is to take advantage of this powerful network will need to use well-documented, open, application and platform independent standards. Both technically and intellectually, the development of the Encoded Archival Description (EAD) guidelines, which were formally enshrined in version 1.0 in 1999, represents a major achievement in harnessing the power of the Internet for providing distributed access to archival resources. Based on the Standard Generalised Markup Language (SGML, an ISO standard since 1986) and on General International Standard Archival Description or ISAD(G), EAD combines the benefits of both. Not suprisingly, it has rapidly and widely found support in the archival
world. EAD offers precisely what is needed to realise the ideal of distributed union access to primary resources wherever they may reside.

In The Netherlands, the conference has brought together a lively EAD working group of archivists, librarians and book historians, who communicate through live meetings and the EAD-NL discussion list. In June 2002 Daniel Pitti gave an EAD workshop at the University of Amsterdam which was attended to capacity. This was concluded by a mini symposium presenting the EAD experiences of the National Archives in The Hague and the International Institute for Social History in Amsterdam. In the meantime, Henny van Schie, senior archivist at the National Archives, has been appointed a member of the international EAD working group. This offers the welcome opportunity to represent the interests of the Dutch working group in the future development of the EAD standard.

But the greatest stimulus of the conference has been to cause the institution of an International Book Trade Archives Working Group. This group, which is in now in statu nascendi, has as its aims to:

• Found an international EAD consortium for book trade archives;
• Create an inventory of significant book trade archives;
• Create an inventory of the existing finding aids in those archives;
• Stimulate the conversion of existing finding aids into EAD or similar finding aids;
• Find a leading institution to host and publish the contributed finding aids in order to create union access to EAD (or similar) finding aids;
• Stimulate the creation of EAD or similar finding aids where none exist;
• Stimulate the creation of digital surrogates and/or marked-up transcriptions, to be attached to item-level descriptions in EAD or similar finding aids.

The editors hope that this book may make its contribution, however modest, to these aims, however ambitious.

BERRY DONGELMANS (Centre for the Book, Leiden University)
AD LEERINTVELD (Koninklijke Bibliotheek, The Hague)
ADRIAAN VAN DER WEEL (Centre for the Book, Leiden University)
Leiden / The Hague, 2002
THE HISTORIAN’S VIEW

DAVID MCKITTERICK
Wren Librarian
Trinity College Cambridge, UK

Lord Acton, founder of the Cambridge modern history, thought at the end of the nineteenth century that the world lived in what he called a “documentary age”: a time when, with the archives of the nation states open to view, and many of them published, it would be possible to unravel the major questions posed by historical events and movements. He knew that a final assessment would never be possible, but that this would be more nearly achievable given the new archival base.¹

Acton’s belief is debatable for several reasons, but I wish here to focus on just one issue that he raises. The last hundred years have indeed seen an appreciation and understanding of archives on a scale that Acton himself could never have envisaged. But they have also proved to be a period fraught both with challenges and with unlooked-for difficulties. Archives have become victims of their own success. They have been mined with vigour by historians; and we now expect much more from them.

But what are these archives? For the historian of the book they begin with the books themselves. When in 1644 John Milton wrote “as good almost kill a good Man as kill a good Book”, he was thinking of the words of authors, as set down in print. We need to go further than that, and we may do so bibliographically. Books, as (again) Milton wrote, in the same paragraph, “contain a potencie of life in them to be as active as that soule was whose progeny they are”.² It has become a hackneyed passage, but it is a useful one: books are vessels, whose freight is the words of the author and the images of the artist. In order to understand one part it is necessary to consider the whole—the physical book as well as the words and images that it contains. Moreover, since books change over the course of time both in themselves (becoming worn, betraying signs of readers in their annotations and other marks, being rebound, etc.) and in their contexts, so they both,
in a manner, speak to us and are themselves a part of their own archive. National laws of copyright deposit in our major libraries are designed to maintain some of that archive; but it has always to be remembered that each copy of each book has the potential to be a living document, more or less open to interpretation according to what we can discover of its provenance and other history.

In acknowledging the special preoccupations of the conference for which this paper was originally written, it may be useful sometimes to recall the kindred interests of archive and printed publication. Their separation on grounds of use and for administrative convenience can sometimes be a distraction. Nevertheless, in common parlance, to speak of book trade archives is to speak not of printed books, but of the records that lie behind their manufacture, their publication and their sale. I turn now to these, while stressing the importance always of remembering the printed archive, its own bibliographical, social and physical strategies, and its integral position in the whole.

The last three decades or so have seen a striking change in direction in the ways that the history of the book trade are pursued. Whereas previously it was entirely acceptable to write on the subject from secondary sources, we now expect to to see archives exploited as never before. In particular, but by no means exclusively, we have seen the rise of the publisher’s archive in historical consciousness. The importance of these archives was stressed at a meeting in Oxford in 1985, attended by an international group of scholars and notable for the first steps towards the creation of a location register specifically of book trade archives. That conference was told of archives in Britain, France, Germany, Poland, The Netherlands, Switzerland, Italy, Australia, Canada and the United States. Despite its geographical range, it has proved a modest enough list. In turn it has led to such finding aids as the list of financial records of British publishers by Alexis Weedon, the lists of American publishers’ archives by Nan Albinski and the Book Industry Study Group, and the Canadian publishers’ records database.3

Indirectly, by drawing attention to the unusual nature of book trade archives, the Oxford conference also helped to encourage the concentration of many archives in particular centres—including, for example, Reading University Library, the Business Archives Centre at Glasgow, the
typographical collections at Cambridge University Library; the collections in the United States at Chapel Hill, the Lilly Library, Columbia University and the Humanities Research Center at Austin; and the Institut Mémoires de l’Edition Contemporaine (IMEC) at Paris. In fact, as inspection of the National Union Catalog and the (British) National Register of Archives quickly reveals, while there are indeed concentrations of this kind, most archives are widely dispersed and responsibilities are shared by many kinds of libraries and archive repositories, national, university and private. In Britain at least, this has serious and fundamental implications for their long-term funding, management, security and access.

Large-scale publication of archives seems to have peaked, partly for copyright reasons, partly because those owners who are willing to share their archives in this way have now mostly come forward, partly because some archives are (frankly) in too much disarray to be publishable without major cataloguing and sorting efforts, and partly because library budgets are under increased stress. Publication has generally been in microform, with accompanying printed guides—these guides being sometimes major reference works in their own right, as in the case of the Bowyers’ printing ledgers of eighteenth-century London.

The position is incomplete. Obviously, not all even of the major archives have been published. We still await a long-promised annotated edition of the ledgers of the Strahans, the other great document of printers in eighteenth-century London. The records of Cambridge University Press have been published on microfilm down to c.1900, but not those of Oxford University Press. In Belgium, the Plantin-Moretus archives are of fundamental importance both at a local level and internationally. Among archives that have been published, there are still major gaps in coverage. The microfilms of the Macmillan records, for example, do not include the records of edition sizes, kept with the firm’s modern working papers and still in the keeping of the company itself.

Here we touch on the relationship between public and private ownership and access. This is a complicated issue, as has been found in other kinds of businesses. Questions of copyright, commercial confidentiality, the need for rapid access in a business environment, and sheer bulk all play their part. With the rapid changes in ownership of many of the older companies in the last few years, and the growth of large multi-national corporations
whose business may lie only partly in book publishing, hierarchies of
decision-making have been changed. New policies can be introduced
to suit larger conglomerates, and new anxieties about the unknown
have emerged. The move from company to publicly accessible library or
archive building may not always be one-way. Other major independent
companies (Faber is perhaps the most celebrated English example) have
so far preferred to keep their own archives under their own control both
physically and legally.

The reasons for pursuing a strategy of independence are clear enough.
However, it should be noted that, in essence, many of the questions are the
same as for other collections of business or family papers, where (in Britain)
the Historical Manuscripts Commission, as a non-acquisitive advisory body,
has been often able to help ensure the long-term survival and good care
of major archives. For archives that have been deposited in university or
national libraries, and where ownership has been retained by the depositor,
there always remain questions of status and ultimate responsibility. Here,
some flexibility is needed in working with owners.

Above all, perhaps, is the question of location. Many archives have been
dispersed beyond recovery. The papers of the great nineteenth-century
literary publisher Richard Bentley are now divided between Illinois,
California, New York, Oxford and London, and more remain loose in
the world at large. After his death, the papers of Friedrich Nicolai (d.
1811) were plundered for correspondence with Lessing and with Moses
Mendelssohn. The Lessing papers are now in Wolfenbüttel, but the bulk
of the Nicolai archive remained together, to be partially destroyed in
the Second World War: in that the Lessing and Mendelssohn papers had
already been removed, dispersal can have unlooked-for advantages as well as
disadvantages. The archives of the Chiswick Press, one of the best printers
in nineteenth-century London, are now in the British Library; but the
personal and some professional papers of Charles Whittingham, who ran
it for many years, are in Vermont, USA, and yet further papers are in New
York Public Library. The archives of the pioneering late nineteenth-century
firm of Chatto and Windus are mostly in Reading, but there are caches
elsewhere, including Invercargill, at the very southern end of New Zealand’s
South Island. More recently, the archives of Macmillan have been divided
between the British Library and Reading University Library, though (as we
have already seen) some of the crucial nineteenth-century records remain at Basingstoke. Other archives have become so dispersed that they now have no coherent value as archives. The professional and family papers of the Nichols family, printers and publishers of the *Gentleman’s Magazine*, dating between the eighteenth and twentieth centuries, are now to be found in at least sixty locations round the world.\[^{11}\]

The movements of authors between publishers, and the endless mergers, de-mergers, separations and sales among different companies that now form a regular part of publishing, make the historian’s and the archivist’s tasks more difficult. Different companies have different policies with respect to their historical record. This is not a new problem, though it is much larger than it was. I take one small, and simple, example. The Nonesuch Press, founded by Sir Francis Meynell in 1923, entered into an agreement in 1927 with Bennett Cerf, of Random House (then a much smaller company than today’s conglomerate) for the sale of its books in America. Then, in 1936, the Nonesuch Press itself was sold to George Macy, founder of the Limited Editions Club. In 1952, Macy gave the Nonesuch Press back to its founder, who subsequently entered into a working agreement with Max Reinhardt of the Bodley Head for finance and distribution. The last Nonesuch Press book appeared in 1992, published privately rather than through the Bodley Head. Most of the Nonesuch Press archives were destroyed in the Second World War, but the pre-war residue, and those for the period between 1952 and the 1970s, are at Cambridge. Even if we leave aside what we might wish to recover from the archives of the relevant printers, we thus face archives in Cambridge, Columbia (Random House), Texas (George Macy companies) and Reading (Bodley Head).\[^{12}\]

Here there is an obvious part not just for computer-based searches, but for the virtual recreation of archives, their recovery from their myriad locations round the world: their revaluation not just as disparate collections of autographs but as coherent reassembled archives. This has been partly achieved in some collections of microfilms, but the opportunities and possibilities for computer-based collaborative ventures are in some ways much greater.

This naturally raises the question of relationships between archives. Book manufacture and use lead to widely differing kinds of archives—some personal, some financial, some technological. They do not fit easily into pre-
ordained patterns. They include paper-making, type-founding, ink-making, printing, authorship, editing, publishing (newspapers and magazines as well as books), wholesale and retail bookselling, libraries, censorship and all kinds of legal records, local and national. Some of the most useful information on the eighteenth-century British book trade is in the bankruptcy records filed in the Public Record Office in London. Some of the best information about the publication of the eleventh edition of the *Encyclopaedia Britannica* by Cambridge University Press in 1910–11 is in court records in New York and New Jersey. Detail of a kind apparently unsurpassed for the sixteenth- and early seventeenth-century English book trade is in the witness statements, taken down verbatim, preserved in the local court records in Cambridge.\(^\text{13}\)

For this, developments such as the British EAD (Encoded Archival Description), ISAD\((c)\) descriptions and other collaborative facilities can only be welcomed. Confusing though the many acronyms can be, they herald what has been called a move away from data exchange to a more collaborative approach.\(^\text{14}\) In many ways they stem from the proposal put forward in 1998 by a report to the National Conference on Archives\(^\text{15}\) for a national archives network: progress has now well advanced on several fronts. The benefits are clear, and have the potential to affect everyone in need of access to historical papers of all kinds. The effects are likely to be little short of revolutionary. However, it is as well to be realistic, and to recognise that it will be many years, if ever, before time can be funded to catalogue and calendar the records on a national scale to the depth ideally required for the kind of searches involved in this or in any other kind of historical investigation.

While (provided funding can be found) it is reasonable to expect that images will increasingly become a part of descriptions, linked to the words and figures that we now rely on, it is well to advance with due caution as to the long-term costs, problems of space, copyright issues, the maintenance of metadata, and the fragile nature of the captured images themselves. It is by no means a straightforward menu.\(^\text{16}\)

I turn now to overseas issues. Much of the publishing trade is an international one. It has been so since the mid-fifteenth century, and it was true of manuscripts as well. But it is remarkably difficult for historians of the book
to establish the scale of that international trade. Sometimes we are lucky, and (as with the records of several institutions in North America) uncover records of imports from Europe. Much can be discovered from the retail price lists of booksellers in Europe from the sixteenth century onwards, and we must therefore look forward eagerly to the publication of the work of Chris Coppens in this field. But while these may give us some indication of geographical spread (the appearance of books published in Italy in catalogues from the book trade in the Low Countries or England, for example), they cannot give us much idea of quantity or retail value.

For example, it would be a useful exercise to map the extent to which the revised (2nd) edition of Newton’s *Principia* penetrated the continental market. It was published at Cambridge in 1713, and we know that copies were bought by booksellers in London who specialised in the export trade. But where did they go? How many more copies were needed? How fast were sales? Another, cheaper and probably unauthorised, edition of the *Principia* was published at Amsterdam in 1714. What is true of a famous book such as Newton’s *Principia* is as perplexing a question as for the many other books that were printed in England and exported (mostly via the Low Countries) in the late seventeenth and early eighteenth centuries, a time when British publishing was at last finding its international feet. One obvious way is to search through the catalogues of private libraries. But it is a tedious job, and one day, perhaps, we will see an index by author etc. to the great series of book auction catalogues from the Dutch Republic in microfiche facsimile that was instigated by the late Bert van Selm and which has been carried on by J.A. Gruys and H.W. de Kooker. The foreign sales records of the Dutch publishing and bookselling firm of Luchtmans in the eighteenth century suggest something of the nature of the problem: an international one, on at least a European scale.

In the early eighteenth century, the export trade from Britain was increasingly important. By the late nineteenth century it was very valuable, and by the mid-twentieth century exports were critical to the survival of many publishers. Between the mid-nineteenth and the mid-twentieth centuries several publishers launched special colonial editions of their books so as to exploit a well-defined overseas market at the lowest possible cost. By 1939, exports of these and other books accounted for 30 per cent of all book publishing in Britain, and (because they permitted larger print runs)
directly affected the price of books in the home country. By the mid-1970s, over 40 per cent was exported. In 1999, Cambridge University Press world sales included exports at 78% by value. Figures today are more complicated, and more difficult to extrapolate in a meaningful way than they have been in the past: the advent of large multi-national conglomerates, the increasing practice of local printing for local markets, the rise of international library agencies and (not least) the success of dot.com companies mean that it is difficult to keep track of where books go once they leave a publisher’s warehouse. But it is clear that many publishers, whether in The Netherlands or in Britain, export most of their work.

As historians, we would like to know more about this international trade. So far, histories of the book (France, the United States, Britain) have rightly looked first to domestic questions. But if (as is clearly the case) the domestic industry depends on international activity, then we should be looking towards more international questions. Historians of reading are usually aware of the international dimensions; historians of publishing and printing have so far been comparatively slow to recognise them.

Overseas sales are the most important aspect; but they are not the only ones. Translation rights, film rights, electronic rights, book club rights, newspaper serialisation rights: all may have an international dimension, and all (regardless of their customer) impinge on the profitability of a title, the reputation of an author, and the well-being of a firm. Production managers of many publishers expect routinely to compare printing and paper prices across the world. Authors are not published just in one country, nor are they necessarily published in their home country. This, too, has been true since the fifteenth century, but it has become a major issue in the history of the book in the last century, to an extent that profoundly affects the economics as well as the rationale for the book trade. The phenomenon requires much more research, though by no means all publishers have translation files so informative as those of Julius Springer Verlag in Berlin between 1912 and 1939. Cambridge University Press may be unusual (it is not unique) in that only about 40% of its current authors are based in the United Kingdom, but this further measure of internationalism presents yet another stark reminder of the need to think across boundaries.

And yet it can be extraordinarily difficult to discover the history of some of these features. In principle, if not in practice, it is (usually) a
comparatively easy matter to discover what is published in translation. We have access to the archives of some of the big book clubs, and Janice Radway has written an entertaining as well as invaluable account of the American Book-of-the-Month Club. But each of these various rights involves negotiations that often led nowhere. We need to know not just what was published, where, by whom, in how many copies at what price, etc., but also the process by which that decision was made. In other words, how does competition work in the book trade?

Many of the archives that we would now like to consult have not survived. It is hazardous to make a generalisation; but it seems that, on the whole, records of production have survived better than have records of sales. In Britain, despite its reputation as a nation of shopkeepers, relatively little has been kept of bookselling, whether of sales to wholesalers or of retail sales to the public. The archive of the booksellers, newsagents and stationers W.H. Smith, stretching from the mid-nineteenth century to the present, is exceptional both in its scope (as that of the firm that for many years had a virtual monopoly of sales on railway stations, as well as a ubiquitous high street presence) and in its nature. Most booksellers do not keep records for more than the few years required for tax purposes.

More seriously in some ways, book trade history can be a study in microcosm. It is one that is reflected in the past emphases on what has been kept. For we have only partial records of most of what we study. The internal memoranda and everyday details of negotiation (now, of course, often telephone calls or emails), the office copies of outgoing correspondence (especially from printers), the details of personnel employed, the decisions about what not to do, rather than just those about what was to be done: all these and more tend not to be kept as parts of archives. To the historian, seeking both to put flesh on bibliographical bones, and also to bring those bones alive by the presence of individuals faced with choices not just theoretical, but practical and everyday, these and other gaps present conundrums that no amount of computerised cross-referencing will answer.

But then, most companies cannot justify the space required—be it warehouse space or (now) computer space. The patterns of preservation in book trade archives of all kinds have been remarkable for their sharp
divergence in emphasis: between on the one hand the necessary legal and financial records (contracts with authors, supporting evidence for tax returns etc.) and on the other hand the emphasis on authorship. There is a long and useful tradition in the historical pride that publishers take in their authors, a pride recorded in volume after commemorative volume of anthologies of letters drawn from old files. Some autograph letters also have a high financial value; and the inevitable trend towards the saving of the better-known authors at the expense of the more everyday is not one that a historian can welcome. Asset-stripping of this kind is a short-term gain at the expense of a better perspective.

In other words, here, too, are questions about selection. Arguably, one sign of a good archivist is the ability to throw away, to be selective and yet comprehensive. Similar issues face decision-making about what should be kept of originals, and what should be selected for special attention in an electronic environment of reformatting and more detailed and better cross-referenced guidance. Computer-based cataloguing presents us with the opportunities to relate different parts of archives, and different archives, even in different countries, together as one: the casus belli that we call, for convenience, the book. This is, of course, potentially expensive. In an environment where public and private funding is struggling to keep up with demand, and where the growth of archives outpaces our ability as a society to care for them, many ambitious projects must seem like daydreams. But the history of the book in its many manifestations, the history of its influence, the history of printing and publishing (again in the widest sense), the history of its use as the bedrock of so much of our society, offer a cause that might be taken up on a collaborative European basis.

NOTES

1 Lord Acton, A Lecture on the Study of History, London, 1905. The lecture was delivered at Cambridge in 1895.
All concentrate on the nineteenth or twentieth centuries. For the Canadian background, but with wider implications, see Laura M. Coles, *Archival Gold: Managing and Preserving Publishers’ Records*, Vancouver, B.C., 1989.


5 <http://www.hmc.gov.uk/nra/nra2.htm> Searches are possible by subject, including bookbinding, bookselling, printing, publishing etc.

6 Keith Maslen and John Lancaster (eds), *The Bowyer Ledgers: The printing Accounts of William Bowyer Father and Son Reproduced on Microfiche with a Checklist of Bowyer Printing 1699-1777: With a Commentary, Indexes and Appendices*, London, Bibliographical Society, 1991. The editors’ meticulous and extensive work make this one of the most informative documents to have been published so far on the eighteenth-century printing trade.


10 Pamela E. Selwyn, *Everyday Life in the German Book Trade; Friedrich Nicolai as Bookseller and Publisher in the Age of Enlightenment, 1750-1810*, University Park, PA, 2000.

11 Julian Pooley, “The Papers of the Nichols Family and Business: New Discoveries and the Work of the Nichols Archive Project”, *The Library* 7th ser. 2 (2001), pp. 10-52. This lists 51 institutional libraries and ten private libraries. Between 1828 and 1951 there were at least eleven sales of significant material at Sotheby’s.


14 *Society of Archivists Newsletter* 141 (April 2001), p.3.

15 <http://nca.org.uk>.


21 Sir Stanley Unwin, *The truth about publishing*, 8th ed., revised by Philip Unwin (1976), pp. 139-40. The first edition of Unwin’s classic book was published in 1926; in its successive revisions much can be discovered about the changing nature of British book publishing in the middle fifty years of the twentieth century.


ADVANCED TECHNOLOGY
AND THE HISTORY OF THE BOOK

DANIEL V. PITTI

Project Director, Institute for Advanced Technology in the Humanities
University of Virginia, USA

I. INTRODUCTION

One of the most often stated and least justified claims of apologists for the “digital revolution” has been that the Internet has succeeded in replacing an obsolete print culture. Such claims have met with justified skepticism among students of the history of the book, but, unfortunately, justifiable skepticism has all too frequently given way to a less discriminating hostility. Despite exaggerated claims and special pleading from those who hype the Internet, scholars and students of the history of the book should recognize that advanced network and computing technologies have the potential to significantly advance their studies by facilitating collaboration, improving access to essential resources, and providing new methods of publication. While the technology should be approached cautiously, much of what is currently available is sufficiently mature, stable, and broadly supported to merit its use by archives, librarians, and scholars.

For disciplines, such as the history of the book, that are inherently international, network and computing technology can not only facilitate communication among scholars distributed around the world, but also provide universal, union access to distributed resources essential to the study of the book. In some cases, selective access to digital representations of primary source materials can—for some but by no means all research purposes—function as adequate surrogates for the originals themselves. For other research objectives, they can facilitate analysis and research that would be difficult, or perhaps practically impossible, when using the original materials. The purpose of this paper is to lay before you the progress along these lines made by a number of intellectual disciplines closely related to the
history of the book. Although I do not want to positively assert that any of what follows is a road map for what ought to be done by the “community” of book historians, I hope that you will see that these approaches to advanced technology offer rigorous and potentially viable means of pursuing fruitful computer assisted research in the history of the book.

II. TECHNOLOGY: EVALUATION

Many librarians, archivists, curators, book historians, and other book professionals have been reluctant to embrace advanced technology. In part this is a response to the exaggerated claims and naïve predictions of technology enthusiasts, and in part it is an understandable and prudent response to emerging and unproven technologies.

For well over thirty years, “visionaries” have been predicting the death of cataloguing, books, libraries, and publishers, and in some Postmodern inspirations, authors and readers as well. It is easy to dismiss the enthusiasts, if not ignore them. Many of them, based on their affiliations, are obviously inspired more by social and commercial self-interest than they are by the technology. Many of their predictions have simply failed to materialize, or attempts have resulted in humiliating disasters. They frequently are forced to humbly retreat when reality turns out to be more complex than they initially thought, or ask, yet again, for a lot more money and a little more time. The complex nature and role of the book and the institutions supporting and depending on it have proven not to be easily reducible.

In spite of the enthusiasts among them, many thoughtful technologists have learned from their technical (and political) failures, and have applied the lessons to developing technologies that enable users to define and solve their own problems in ways that are appropriate and responsible. In this regard, they have made particularly important advances in developing publicly owned standards that ameliorate the dependency of data on proprietary hardware and software (and technologists). These standards enable users to take control over their data, representing and exploiting it to serve their own interests and objectives. Rather than technologists determining the future, inventing and imposing on others solutions to problems imagined by them, users now increasingly have it within their power to master the technology and employ it to serve their own objectives. It is now the
The responsibility of users to recognize and take advantage of the emerging and existing opportunities.

The experience of librarians and technologists in applying technology to cataloguing provides a useful example from which we can learn both good and bad methodologies. Early collaborations between technologists and librarians were fraught with misunderstanding and miscommunication. The technologists initially woefully underestimated the complexity of books, publishing, and cataloguing. Many naively viewed libraries as large warehouses, and catalogues as inventories of the items stored in them. Completely overlooked or ignored was that most warehouses contain a large number of a small variety of items, while libraries contain a large number of mostly unique items. Further, the unique items are frequently “ill behaved”, defying easy categorization and description, and exist in extremely complex interrelations with one another. All of this and more make even the most basic catalogue far more complex than a simple inventory. Cataloguers were also naïve about the technology. They frequently had little or no experience with it, and tended either to accept the view of the technologists without question, or to reject the technology out of hand. There was little or no mutual understanding or shared terminology.

Despite all of this, the technologists and cataloguers managed to invent something extremely useful, machine-readable cataloguing, or MARC, as we have come to know it. But they also made some serious mistakes, some of which led to damage that was expensive and, in some cases, impossible to repair. A noteworthy and instructive example is in the area of authority control and key word access.

When keyword searching became possible, many technologists predicted that it rendered authority control obsolete. Based on this prediction, many library administrators instructed their cataloguers to stop doing authority control. While the librarians and the users of catalogues quickly determined that keyword searching was a powerful and useful tool, enabling retrieval impossible with printed cards, they also determined that it was no substitute for experienced professionals making difficult judgments and distinctions, and recording them in machine-readable and therefore exploitable form. For example, computers are still incapable of recognizing that all of the following names refer to the Muslim philosopher al-Ghazzali (1058-1111): Ghazzali, Gazali, Abu Hamid Muhammad ibn Muhammad ibn Ahmad

On the other hand, the computer can do some wonderful and some not too wonderful things with these distinctions, once they are recorded in computer-readable form. Using keyword access, computers can easily direct users from unused to used headings. Discovery that might take a great deal of time and persistence in a card catalogue, and might not be possible at all, is made efficient. But these same distinctions, if not carefully used, can also lead to embarrassing and perhaps disastrous consequences.

One particular mishap in computer assisted authority control has become notorious. In the 1980s, OCLC, the large, international bibliographic utility located in Dublin, Ohio, decided that it needed to “clean up its catalogue” after many of its clients complained repeatedly that it had a serious authority control problem. The programmers at OCLC decided to write a program that would match headings used in catalogue records against unused variants found in Library of Congress authority records, and where they found matches, substitute the heading in the authority record for the heading in the bibliographic record. On the surface, this seemed like a perfectly reasonable thing to do. Unfortunately, it had many unforeseen consequences. One has become well known: the program changed all of the headings for Madonna, the popular singer, into “Mary, Blessed Virgin, Saint”. Many librarians specializing in authority control took this as clear evidence that computers could never do authority control, as any reasonably informed human being would not make such a stupid mistake.

This evaluation, however, was not entirely sound. Many of the librarians saw only the “collateral damage”, and failed to recognize that a good portion of the program worked quite well, and accomplished the desired goal. The programmers involved, having experienced success tempered by embarrassing humiliation, analysed their failures, and began to approach
the problem of identifying when the same name did and did not apply to the same entity more carefully. They improved the algorithms to recognize when a match was “safe”, for example a personal name qualified by one or more life dates, and when it was not, for example, a personal name with only two components and without qualification. A careful, fair analysis of the results demonstrated that they could do a lot programmatically, but that there would always be a remainder that only a trained, intelligent, professional person could sort out. They switch from trying to make the computer do everything, to trying to do as much as could be done accurately and reliably while leaving the remainder to the professionals with suggestions and information to help them in their problem solving. Technology, carefully and deliberatively applied, could perform many of the most routine and tedious chores, while isolating the most challenging tasks for librarians. The result is that cataloguers now have more time to spend on the problems and challenges that resist reduction to computer algorithms.

The development and application of MARC has taught technologists and librarians a great deal. Much though not all of the early overconfidence in technology, on one hand, and overly skeptical assessment of it, on the other hand, have been displaced. In their place are clear, realistic collaborative assessments of what technology can and cannot do, applications that exploit computation while respecting, facilitating, and exploiting the irreducible contributions of professional cataloguers, and more careful, reversible experiments that test and extend the current technological limits.

MARC has been unquestionably successful. The emergence of online, networked catalogues has realized for the first time in history the long held dream of the “universal catalogue”. Since the late 1980s, there have been major advancements in computer and network technologies and our power to represent and exploit information to serve a wide variety of goals and interests. As we encounter these emerging technologies, we need to do so with a methodology informed by our past experience.

Historians of the book can draw on the experience of the librarians in adapting technology to serve their professional objectives. Historians need to acquaint themselves with the technology, to understand and evaluate what it can and cannot do, and to determine appropriate and responsible uses of it. They also need to collaborate and work with technologists. This collaboration necessarily requires assisting technologist in understanding and
respecting the complexities of their discipline. Shared terminology needs to be negotiated. Technologists alone should not determine the applications. The historians should proceed deliberatively, based on informed and careful consideration of what is and is not possible now, and what is and is not likely to be possible in the near future. They should welcome what is clearly useful, adapting it to serve their professional goals and interests, and defer use of technologies whose stability, utility, and independence are not demonstrated.

III. TECHNOLOGY: OVERVIEW

Since the early development and use of MARC, there have been major technological developments that are having an extensive, pervasive impact on all of the major institutions of modern society and culture. The Internet lies at the centre of these developments. It interconnects and thus makes other technologies far more powerful than they would be in isolation. At the same time and as a direct result of its empowering of other technologies, it constitutes the most important economic force driving much of the investment and development of computer and communication technology. Database technologies have matured significantly, enabling complex, large-scale representation and manipulation of certain classes of information. Since the early 1980s, the emergence and development of markup technologies has for the first time made it possible to accurately represent texts of arbitrary length and complexity based on user supplied specifications. A variety of technologies for the digitisation of existing media and original digital creation of analogues of existing media has also emerged. Finally, the emergence, of affordable, increasingly powerful personal computers and increasingly accessible software has put all of these technologies into almost all major institutions, and into the homes of many private citizens as well.

INTERNET
The Internet stands out as the most transforming of the computing related technologies to emerge in the last twenty years because it interconnects computers and the people using them wherever they are in the world, and any time of the day or night. While most of us first became aware of the Internet in the early 1990s, the research and early prototypes that led to it
began in the 1960s. As the network began to be realized in the 1970s and 1980s, its potential for facilitating communication between computers and through them people became increasingly clear. New standards and software emerged to take advantage of the potential. Telnet emerged as a way for a user on one computer to connect to another computer. Researchers developed File Transfer Protocol (FTP) to enable moving files back and forth between computers. Other researchers developed electronic mail (email) to enable sending and receiving messages. Related to FTP, researchers developed client-server technology, to support distributed processing of information. In conjunction with email, scientists developed listserv technology to support group discussion and communication. Still later, researchers invented Hypertext Markup Language (HTML) to enable online viewing of texts with “links” that led to yet other texts. To take full advantage of hypertext, researchers developed browsers that enabled not only interrelating texts, but also interrelating texts and other digital media. The emergence of browsers and hypermedia coupled with the Internet in the early 1990s spawned the current Internet “phenomenon”.

The Internet has had an effect on most if not all of us, professionally or personally, or both. We are more than likely to use email on a daily bases to communicate with individuals within our own institutions, and with colleagues in others. We are also likely to subscribe to one or more listservs that focus on one or more of our professional responsibilities or intellectual interests. We can easily send messages and files. We are also likely to use browsers and various indexing utilities such as Google to find, retrieve, and read or print articles and essays of interest to us, or to access the Oxford English Dictionary, the Encyclopedia Britannica, or the bibliographic catalogues in varies research libraries. While the technology is far from flawless—proprietary files have been a persistent problem—it has progressed significantly and continuously over its short history, to the point where many of us cannot remember at time when it was not there. We mostly do not realize how much we have come to depend on it, except in moments when we do not have access when we expect to. For many scholars, especially humanists and social scientists, who had no access to the early networks and frequently had little or no opportunity to collaborate and communicate easily and regularly, it has improved immeasurably our communication with colleagues, and our access to information.
DATABASE TECHNOLOGY

Database technology is designed to store, manipulate, and access large volumes of highly regularized data. Modern database technology began in the early 1960s with efforts to develop techniques to conceptualize, structure, and manipulate data independent of the specific hardware used. The most prevalent types of database are hierarchical, network, relational, and object oriented, with relational databases being the most prevalent. While object oriented databases have not been widely implemented, the technology has contributed conceptual and functional models that have influenced the most recent relational database implementations. Relational databases with functionality inspired by object orient databases are called object-relational databases. Affordable yet sophisticated relational database software frequently comes packaged with personal computers.

The widespread availability of database technology enables individual scholars to compile and manipulate large amounts of data to support research. In addition to individual projects, the technology supports large collaborative projects, enabling scholars and researchers to cooperatively build shared sets of various kinds of data: bibliographic, census and demographic, statistical, genealogy, and others. In addition, database technology provides the data infrastructure for sophisticated Geographic Information Systems, and computer graphics. Database technology also provides the infrastructure for MARC-based online catalogues and maintenance systems, as well as access, description, and control systems in archives and museums.

Database technology is most useful for representing and exploiting specific kinds of data. The kind of information found in forms and questionnaires fits well in databases. For example, personnel records and job application are perfect candidates, as are records describing publishers and book traders. In general, database suitable documents have the following characteristics: each document has the same set of data elements; the order of the data elements in any given record is not important; and the data elements in any record have few or no hierarchical relations with one another. Documents that not are not generally suitable for databases have the following characteristics: the documents differ from one another in the number, kinds, or sequence of components; the order of the document components is important; and components have many, frequently unbounded hierarchical relations with one another. Texts, such as those found in books and journals belong to this
type, and markup technology rather than database technology has emerged as the optimum way to represent and exploit them.

**MARKUP TECHNOLOGIES**

All information in computers is encoded to facilitate processing it. In the early history of computing, the codes associated with textual information were typically procedural codes. Procedural codes specify certain operations or procedures that are to be applied to the information. Word processing programs, the most common text application, associate codes with text to facilitate printing it. The various codes represent different styles that are to be applied to information. For example, the title of an article might have codes that will facilitate centering it on the page, and printing it in a large, bold font. Most procedural encoding is proprietary and devoted to one output, the most common of which is print.

In the late 1970s and early 1980s, an alternative to procedural encoding emerged. Instead of embedding procedural codes in texts, descriptive or declarative codes were embedded. Descriptive encoding of text specifies what the text and its components are rather than the procedures to be applied to it. Descriptive encoding is a process of naming, as opposed to procedural encoding, which is a process of associating verbs or actions with text and text components. The descriptive or declarative approach has the major advantage of supporting multiple procedures, even procedures not anticipated at the time of encoding. For example, declarative markup might state that a given string of text is a title. This string can then be printed using one set of procedures, displayed on a computer screen using another, and indexed using yet another.

Standard Generalized Markup Language (SGML), first codified in 1986 by the International Standards Organization, is a descriptive method of representing or encoding textual information in computers. While SGML is both standard and generalized, it does not provide an off the shelf markup language that one can simply take home and apply to a letter, novel, article, catalogue record, or finding aid. Instead it is a markup language meta-standard, or in simpler words, a standard for constructing markup languages. SGML provides conventions for naming the logical components of documents, and a syntax and meta-language for defining and expressing the logical structure and relations among the components. SGML is a set of
formal rules for defining specific markup languages for individual kinds of
documents. Using these formal rules, members of a community sharing a
particular type of document can work together to create a markup language
specific to their shared document type.

The specific markup languages expressing these analytic models and
written in compliance with formal SGML requirements are called Document
Type Definitions, or DTDs. For example, the Association of American
Publishers has developed four DTDs: for books, journals, journal articles,
and mathematical formulae. After thorough revision, this standard has been
released as an ANSI/NISO/ISO standard, 12083. A consortium of software
developers and producers has developed a DTD for computer manuals called
DocBook. The Text Encoding Initiative (TEI) has developed a complex suite
of DTDs for the representation of literary and linguistic materials. Archivists
have developed a DTD for archival description or finding aids called Encoded
Archival Description (EAD). There are even several DTDs for representing
various varieties of MARC. A large number of government, education and
research, business, industry, and other institutions and professions are
currently developing DTDs for shared document types. DTDs shared and
followed by a community can themselves be standards. ANSI/NISO/ISO
12083, DocBook, TEI, and EAD are all standard DTDs.

HyperText Markup Language (HTML) is an SGML DTD that has enjoyed
enormous success as the encoding standard underpinning the World Wide
Web. As a specific application of SGML, the HTML DTD limits itself to simple
procedural encoding dedicated to online display and hypermedia linking.
Constraining the set of tags has made it easy to build applications that make
life relatively easy for authors and Web publishers. The ease of use has been
a major factor in the Web’s remarkable success.

The developers of HTML, the World Wide Web Consortium (W3C),
recognized that HTML, as useful and popular as it has been, would not support
complex, community-based use of shared information on the Internet. Because HTML implements a small, closed set of procedurally oriented tags,
it is incapable of supporting sophisticated searching, navigation, display, and
communication. Evidence of HTML’s limited ability to support intelligent
searching and document discovery, let alone complex display, navigation
and other processing, is not difficult to find. Many of us have used Web
search engines to look for both known items and items on a particular
topic. More often than not, we are overwhelmed by voluminous results, with many and perhaps most of them being irrelevant. Our patience frequently is exhausted looking for an item or two that satisfies our need. The small, closed tag set has thus come at a price: HTML has extremely limited functionality.

The w3c recognized in SGML’s declarative approach and extensibility the means to overcome the limits of HTML, but they also noted that SGML presents its own set of problems. It is very complex for software developers, and as a result, software products for exploiting the richness of the descriptive encoding have been limited in number and almost always expensive. In 1996, the World Wide Web Consortium (w3c) founded the eXtensible Markup Language (XML) Working Group to address this problem. The Working Group, in a short period of time, wrote a specification for a simplified subset of SGML named XML. They simply eliminated the features of SGML that were problematic for programmers. XML is simplified or normalized SGML.

XML encoding of text provides a means of representing textual semantics and structure, but it does not in itself provide support for the procedures that are likely to be applied to texts. Presenting or displaying text on a computer screen and printing it on paper are two obvious procedures. The w3c has developed eXtensible Stylesheet Language for standardizing both of these procedures, as well as other transformations of text. XML Linking Language (XLink) is standardizing hypertext and hypermedia behaviour. In addition to supporting the kinds of links familiar currently on the Web, XLink will enable linking not only to other documents and digital media, but also into them, even when the (author of the) referencing document does not control the referenced document. XLink will also support annotation of texts and objects, again regardless of owner. XML Query (XQuery) will support the standardization of searching texts (and databases as well). Together with XML, these standards and other related, supporting standards represent a relatively complete, standard approach to textual information.

While the origins of SGML lie in the processing of texts, XML is also being used as the basis for encoding and communicating many kinds of data. A large number of the current XML initiatives involve data that is created and maintained in databases, but is communicated among databases and
published on the Internet using XML. Many of these initiatives involve commercial databases and business transactions. Still others involve what has come to be called, for better or worse, metadata.¹² Noteworthy also is Scalable Vector Graphics (SVG), an emerging standard developed by the W3C “for describing two-dimensional graphics in XML”. A companion standard, also under development by the W3C, is Synchronized Multimedia Integration Language (SMIL), which supports, as the name suggests, integrated presentation of multiple media.¹³ SVG and SMIL appear likely to have an impact in the presentation of geographic information, though yet another effort is devoted to creating XML-based representation of the geographic information itself. This effort, the Geography Markup Language (GML), is organized and led by the Open GIS Consortium.¹⁴ The Web3D Consortium is developing an XML-based standard for three-dimensional graphics, Extensible 3-D (X3D). It is based on the SGML-based ISO standard Virtual Reality Markup Language (VRML).¹⁵ X3D will provide a standard encoding of data supporting an extremely wide variety of three-dimensional objects. Examples are rooms, buildings, automobiles, and, of course, books. Like SVG, X3D also promises to provide support for the representation of geographic topographical information, as well as Computer Aided Design (CAD), used extensively in architectural and engineering design. All of these efforts are beginning what is likely to be an extended period of standardizing machine-readable data for various media. Some, such as XML itself, are well along in development, with an increasingly wide range of products of increasing quality available. Others are in various stages, from just underway to nearing approval by the W3C and other authoritative bodies.

Both markup technologies, as represented in XML and related standards, and database technologies are particularly significant because they enable users to represent semantically and structurally rich intellectual understandings of text and other data in machine-readable form that can be exploited using a wide array of existing procedures as well as procedures yet to be devised. The wide array of current XML and database initiatives demonstrates the importance of this descriptive and representational power. In the humanities research community, semantically rich machine-readable expressions have already significantly enhanced intellectual access to cultural objects, through MARC and other cataloguing standards, and are beginning to enhance analysis and interpretation of them as well.
Markup and database technologies are also significant because they ameliorate the dependency of data on hardware and software. Migrating information out of an obsolete standard into a new, better standard is an important procedure that will sooner or later become necessary. This is typically overlooked in the conception and design of projects and programs. Fortunately, standard, descriptive encodings inherently provide more support for this procedure than do proprietary, procedural encodings. When you know what the information and its parts are, migrating from one standard to another is a matter of semantic mapping rather than procedural mapping. XML even has its own standard for accomplishing this transformation, XSL.

While markup and databases technologies enable semantic encodings, they do not in themselves define the semantics. Specific cultural heritage disciplines and communities sharing intellectual and professional objectives must be responsible for the analysis, specification, expression, and application of the semantics. Libraries have accomplished the most in this area, though archives and museums are also actively engaged in standards development. Primarily through the Text Encoding Initiative, humanists have also begun to define standards, though a great deal of work remains to be done, and much remains to be done in collaboration between these communities. Developing shared semantics and structures represents the most important challenge facing the cultural heritage disciplines and communities in the near future.

DIGITIZATION TECHNOLOGIES
Pictorial, audio, and audio-video media are also increasingly standardized. For pictorial material there is one major de facto industry standard: Tagged Interchange File Format (TIFF).\textsuperscript{16} While we should prefer open, public standards to de facto industry standards, when a proprietary encoding achieves widespread industry support, its fate is no longer tied to one company, and one or two computer programs, and thus the dependency and risk involved in its use is ameliorated. TIFF as well as de facto standards such as PostScript and Portable Document Format (PDF) fall into this category. Because an image stored in TIFF preserves all of the information captured, it is generally considered acceptable as an archival encoding. TIFF file sizes, though, tend to be quite large, making them unsuitable for Internet transmission given
current bandwidths. Thus smaller files using techniques for compressing the files are generally used for Internet publishing. The most popular of these compression formats is the ISO JPEG standard. The current version of JPEG is not appropriate for use in archiving and preservation because the techniques it employs lead to information loss (lossy compression). Currently ISO is developing a successor to JPEG called ISO 2000 that will support compression without loss of information (lossless compression), and thus will be suitable as an archival format.\textsuperscript{17}

ISO also has developed standard encodings for audio and audio-visual materials, though these are primarily compression standards and thus are designed for audio and audio-visual files used on the Internet. Audio (and thus by implication also audio-visual) capture that is sufficient for long-term archiving and preservation is problematic because sound is continuous, and digital capture, by definition, is discontinuous. Analog is converted to digital through sampling, and until sampling rates achieve an acceptable threshold there is significant loss of information. Nevertheless, research and development continues in this area with the expectation that capturing audio data in digital form using high sampling rates will make digital preservation of audio and audio-visual material feasible in the near future. Despite these limitations, used cautiously, many of these \textit{de facto} and public standards are sufficient for many purposes.

Taken together these technologies and standards offer book professionals many possible opportunities to apply advanced technology in their research and teaching. Many of the technologies are well established and proven, and based on solid, open, public standards. Others are also well established and understood, but not yet based on standards. Nevertheless, serious standards efforts are underway for many of these as well. Certainly not all problems have been solved, and there are risks, especially for naïve users. Many of the conveniences that are introduced by the advanced technology also introduce uncertainties that threaten to upset the balance of control and interests upon which the current “order of the books” rests.\textsuperscript{18} With due thought and caution, though, much of the technology is sufficiently mature, stable, and broadly supported to merit its use.
IV. HISTORY OF THE BOOK: COLLABORATION AND COMMUNITY

One of the great opportunities presented by advanced technology is that of facilitating collaboration. The sciences and the social sciences, even before the advent of the Internet and related technologies, frequently employed collaborative projects to achieve shared research objectives. The scale and complexity of many projects motivated these collaborations. Complex, labour-intensive projects required group effort to be successful. Without collaborating, the research was simply impossible. Collaboration also had a major secondary benefit as well. Designing and carrying out complex projects required intensive exchange of ideas and negotiation that led to many intellectual advances and breakthroughs. Many of these would have been difficult or more slowly realized, if at all, by scholars working alone.

Outside of a few dictionary and encyclopedia projects, humanists rarely have collaborated in this manner. The ongoing building of bibliographic catalogues stands out as a major exception. Most often humanists toil in solitude, communicating now and again with one or two trusted colleagues. Until now, they have lacked by and large both the means and the motivation to engage in collaborative research activities.

Technology provides both the opportunity and the motivation for historians of the book to engage in large-scale, complex projects. The history of the book is international in nature. Book historians are distributed throughout the world, as are the tools and resources employed in their research. The book trade itself is international, with many of the significant figures and firms operating across borders. The international nature of the book trade accounts for some of the distribution of resources, though collectors have also contributed. The distribution of people and resources has been a major, time-consuming, and frequently prohibitive obstacle to both scholarly communication and research. Advanced technology provides the means to overcome this obstacle, though only if the historians of the book are prepared to collaborate with one another, and with librarians, archivists, publishers, and others, in complex and intellectually intensive collaboration.

Existing and emerging technologies present several opportunities to historians of the book. It makes possible providing universal, union intellectual access to resources in the form of specialized bibliographic catalogues and
archival description systems. It also makes it possible to provide selective access to digital representations of bibliographic and archival resources that can function as adequate surrogates for the original for some, though by no means all, research purposes. In some cases, these digital representations may also facilitate analysis and research that would be difficult, or perhaps practically impossible, when using original materials. The technology also enables building analytic and pedagogical tools that can be shared. Finally, it offers the opportunity to create new forms of publication and pedagogy employing these resources.

As the foregoing has illustrated, achieving such things in a manner that will assure their usefulness over time requires the disciplined efforts of a community. An essential factor in establishing a collaborative community or consortium is having one or two lead institutions that are willing to provide hardware, software, and technical expertise to host, maintain, and publish resources, and to facilitate communication among participants. While some of these functions can be distributed, such as distributing responsibility for communication to one institution, and resource maintenance and delivery to another, distributing resources is problematic technically, especially if the resources constitute a wide variety, and at the same time have many interrelating links. For now the existing technology does not easily support such distribution. There is a major economic advantage to centralizing some of the more complex operations, as it relieves participating individuals and institutions of having to invest time and money in mastering complex supporting technology. Distribution of creation and maintenance activities, however, is absolutely essential, as the expertise needed to gather resources and the resources themselves are distributed.

The major challenges in building a history of the book consortium on the Internet are not intellectual and technical, as difficult as these are, but political. Politics, in the more attractive sense of the term, is community building. Communities first and foremost must articulate and share common interests and goals. Agreement can be difficult to achieve, as it requires negotiation and sacrifice. Individuals will only voluntarily participate in a community if it enables them to more effectively pursue individual interests, and to achieve goals that are difficult and perhaps impossible to achieve working alone.
V. HISTORY OF THE BOOK: PROJECT SUGGESTIONS

The necessary first step in building a consortium is developing a shared vision of objectives. In what follows, I will suggest some possible objectives. Coming from an outsider with no detailed knowledge of the nature and methods of the discipline, they are all offered merely as suggestions, and not as recommendations. They are intended to promote discussion, criticism, and counterproposals.

The suggestions begin with establishing the communication necessary for collaboration. Immediately following these are proposals to improve intellectual access to resources. Following are proposals for providing structural and digital image representations of resources. Intellectual access is intentionally presented before digitizing, as it is especially important, and it provides the foundation necessary for building digital resource collections. A final section groups together reference materials, critical secondary resources, and geographic information. While most and perhaps all of these proposals are ambitious, perhaps immodestly so, building a consortium should begin with modest, realistic projects that are relatively easy to accomplish. Early, modest successes will establish the trust and expertise necessary to accomplish more ambitious objectives.

LISTSERVS

Generally, to carry out collaborative projects such as I have been proposing requires the establishment of one or more listservs to facilitate communication. How many listservs, and devoted to what purpose, depends upon the number, size, and complexity of the projects undertaken, and thus the degree to which specialization is necessary. At a minimum, an emerging consortium needs at least one list to organize and discuss the activities of the consortium itself. Over time, the need may arise for specialized discussion groups devoted to activities such as administration and governance, technical infrastructure, and intellectual and technical standards. In addition to administrative communication, listservs can serve to facilitate scholarly communication. There are already several discussion lists devoted to the study of books and related technologies:

- ExLibris: Rare books and special collections
- Book_Arts-L: All book arts
Rather than establishing competing lists, a consortium might instead choose to concentrate on lists that complement them.

ACCESS TO HISTORICAL EVIDENCE
Since books themselves obviously constitute a major source of evidence for book historical research, improving intellectual access to especially significant books and book evidence would greatly expedite and improve research. Existing online bibliographic catalogues have already improved access, though they have a number of disadvantages. They are distributed, and thus require serial searching. A researcher must have a good idea of where a particular book is likely to be located before beginning a search. Catalogue interfaces also vary, adding to the complexity of the challenge. Both OCLC and RLG, though, have come a long way in solving this problem. More difficult, though, than the distribution of catalogues is that most MARC cataloguing lacks forms of access that would be useful to students of the book. Existing practices at the Bancroft Library at the University of California, Berkeley may serve as a useful example of improving access for book historians. Using more detailed MARC records than are typical, the Bancroft Library is providing specialized access to its collections. Through the online catalogue, the following indices are available:

- Chronological: inverted geographic access to place of publication, subarranged by date;
- Typographical: access to printer or publisher;
- Binders: access to bookbinders;
- Association: access to former owners (provenance);
- Genre/Form: access to printing and publishing evidence; binding, genre, and paper terms; provenance evidence; and typographical evidence.
Using the Bancroft Library’s and similar efforts as a starting point, a consortium could develop content and encoding standards for specialized headings. The consortium could work actively to encourage use of these standards, and lobby the large bibliographic utilities and vendors of MARC catalogues to provide specialized searching based on them. A more ambitious project might bypass the utilities and vendors by creating its own union, international catalogue incorporating the specialized headings. Such a catalogue might be based on MARC, or alternatively on XML, using, for example the UNIMARC XML DTD developed by the BiblioML project in France. This would make it possible to use XML indexing and publishing software instead of a MARC system. Mapping the various dialects of MARC and non-MARC records into a UNIMARC DTD would be quite complex, though feasible, as demonstrated by the Manuscript and Letters Via International Networks (MALVINE) project, discussed below. Any or all of these initiatives would improve access to book evidence.

Encoded Archival Description (EAD) is an emerging international standard for encoding detailed archival descriptions of fonds or collections. EAD provides a standard representation of descriptions of the records of corporate bodies, and the papers of individuals and families. While EAD has many rationales, perhaps the most compelling is that standard archival description supports the long-cherished dream of providing both professional and public researchers universal, union access to primary resources. Currently there are dozens of institutions throughout the world using EAD, with the number growing rapidly. Many of the EAD implementations are consortia with several repositories participating. These are generally organized geographically, such as the Online Archive of California, or by discipline or subject, such as the Physics History Finding Aids project. The MALVINE project is organized geographically (European Union) and by genre (letters and manuscripts). The Research Libraries Group (RLG), an international consortium of research archives, libraries, and museums, is currently providing union access to finding aids from throughout the world through its Archival Resources service.

EAD makes it possible to greatly improve access to the records of individuals, families, and firms that have made significant contributions to the history of the book. Like the history of physics and many other disciplines, the history of the book transcends national borders, with many
of the significant individuals, families, and firms active in more than one country. The records documenting the activities are distributed in many countries, and within countries, in more than one repository. A worthy project might focus on identifying the significant *fonds* and collections distributed in European repositories, and centralize descriptive access to them using *ead*. This would necessarily involve a wide variety of activities. Identifying what collections are processed and described, and evaluating the quality of existing descriptions would be a first step. Developing a strategy for converting existing descriptions into *ead* would follow. Creating the archival description system would require converting print finding aids into machine-readable form, and mapping and writing conversion scripts for finding aids in word processing and database formats. *ead* implementers already have extensive experience in working with vendors that convert paper finding aids, and vast in-house experience converting word-processed and database finding aids. Organizing and seeking funds for processing unprocessed collections would follow conversion of existing finding aids. The existing *ead* consortia all have a lead repository or institution hosting and publishing the contributed finding aids. Providing international, union access to significant archival evidence would greatly facilitate access to significant archival evidence, and complement the access to book evidence discussed above.

The Manuscripts and Letters via Integrated Networks in Europe (*MALVINE*) project provides an excellent European example of how such an initiative might be organized and implemented. *MALVINE* is funded by the European Union. There are fifteen participating archives, libraries, museums, and documentation centres, located in nine European countries. *MALVINE* is coordinated by the Staatsbibliothek zu Berlin, though the Humanities Information Technologies Research Programme at the University of Bergen coordinates and provides technical support for conversion and publishing. Other responsibilities are distributed among the other participants. Key components of the success of *MALVINE* have been the collegiality of the participants and the highly capable technical expertise brought to the effort. *MALVINE* represents an excellent model for European collaboration.

To complement access to bibliographic and archival evidence, a consortium might also explore options for describing and providing access to tools and apparatus used in the production of books. The International
Council of Museums’ Conceptual Reference Model, and other museum initiatives and standards efforts should be explored in this regard. Some museums are also experimenting with EAD to provide access to museum artifacts.

**REPRESENTATION AND ANALYSIS OF EVIDENCE**

In addition to enabling enhanced access to book and archival evidence, existing and emerging technologies also enable creating machine-readable representations of the evidence itself. Such representations will be suitable for some research purposes, and may improve existing analytic methods and inspire new methods.

Machine-readable representation of evidence can be done in three ways. First, XML can be used to create descriptive, structural representations of objects, for example, the physical features of books. Second, imaging technology can be used to capture graphical information. Two-dimensional imaging can be used for page images, manuscript and print archival resources, and similar flat resources. Three-dimensional imaging can be used for books, tools and apparatus used in the production of books, and similar resources.

Using the TEI DTD, Terry Catapano and Syd Bauman provide an example of a machine-readable structural representation of the physical features of a book. The TEI, by design, is optimised for encoding the intellectual structure of a book: chapters, paragraphs, poems, lines of poems, and so on. It is sufficiently flexible, however, that Bauman and Catapano were able to devise a prototype description of the physical structure of a book. They state that there are certain advantages to such a representation:

The re-arrangement of the pages of text as imposed for printing may make apparent places where the text was affected by typographical exigencies. It is also useful in electronic bibliographic analysis

- for identifying which compositor set which forme, in order to distinguish their individual spelling and punctuation habits;
- to track identifiable pieces of type to determine the order of printing; and
- to discover the course of proofreading and correction.
Their example is intended to demonstrate, in a very preliminary way, the feasibility of using XML to represent the physical structure of a book. It is not a fully developed system. Using this demonstration as a starting place, a project might attempt to design and develop a comprehensive XML DTD for representing the physical features of books and attempt a variety of computer-assisted analysis to determine its utility. If successful, such a DTD would facilitate building a shared collection of representations of individual, exemplary books that could be made available to the scholarly community for analysis, discussion, and teaching. Such a DTD optionally might be developed in collaboration with the TEI Consortium.

While TEI provides a comprehensive scheme for text description and representation that facilitates literary, editorial, historical, and linguistic analysis, historians of the book are likely to find much in it that they do not want or need, and not fund many elements that they do. Medievalists, for example, have found that TEI lacks sufficient detail for both description and representation of medieval manuscripts. Projects such as the Manuscript Access through Standards for Electronic Records (MASTER) and Electronic Access to Medieval Manuscripts (EAMMS) are working with the TEI Consortium to extend TEI to optimise its use in the study of medieval manuscripts. Initiatives similar to these might be appropriate for the history of the book community.

Page imaging of books also offers significant opportunities for collaboration. There is an ongoing and well-justified controversy concerning the use of page imaging in preservation. Preservation is an extremely complex issue, as all migration of information from one medium to another involves loss of information, namely the medium from which the information is transferred or migrated. When the medium itself is the primary evidence, such as it is in the case of books significant in book history, media transfer is an unacceptable preservation method. The only acceptable preservation method for significant books is preservation of the book itself. Page imaging, though, can be useful in facilitating access and analysis of evidence for some purposes, for example, the study of typography, and for bringing together distributed materials for comparison. The William Blake Archive provides an excellent example of the use of page imaging. A worthwhile area of collaboration would be in experimenting with and establishing best practices in image capture and quality to facilitate the utility of imaging in support
of research and analysis. Subsequent to establishing image quality standards, collaborative building and sharing collections of page images would seem a worthwhile undertaking.

Widely demonstrated and accepted standards and technique for high quality three-dimensional imaging have yet to emerge, though research, prototyping, and development are well underway. Large-scale use of this technology would be premature at this time, but exploring and experimenting with the technology to identify useful applications would position the community to take full advantage of the technology as it becomes standardized and widely used. Uncle Tom’s Cabin & American Culture at the Institute for Advanced Technology in the Humanities provides simple but suggestive examples of three-dimensional book imaging using QuickTime®. An experimental project, the Brazil Rendering System, provides some striking examples of three-dimensional renderings exported to standard two-dimensional imaging formats. In addition to books, three-dimensional technology would also be useful in representing tools and apparatus used in book production.

OTHER PROJECTS
While improving access to primary resources should be the first priority of the history of the book community, there are also a number of other potentially useful project opportunities. Reference materials and authoritative secondary literature can aid in the use of the primary resources, as can geographic information.

Securing digital rights to significant secondary works and reference materials, and publishing them on the Internet is certainly worth considering. Materials that fall into this category are landmark histories of the book and dictionaries and glossaries, in essence, the kinds of materials that any historian of the book has within arm’s reach in his or her office, and which are consulted regularly. Reference materials of this kind are especially useful online, as the technology facilitates quick access, and the information sought is typically brief and thus easily readable on a screen.

Making important geographic information available can be extremely useful for reference, research, and teaching. Geographic Information System (GIS) provide sophisticated access to highly accurate spatial information. Geographic Information System also can be linked with social science data
systems and other relevant datasets. For example, GIS linked to datasets might provide dynamic graphical presentations of the location and movement of people, firms, trade, materials, technology, and ideas over time. Even static page images of maps can be extremely useful. Two projects at the University of Virginia demonstrate the use of map information in conjunction with primary resources. The Valley of the Shadow Project uses animated maps to trace the movements of soldiers from Pennsylvania and Virginia during the American Civil War. While still under development, the Salem Witch Trials is linking geographic and temporal information with a database documenting families, individuals, institutions, significant events, and documentary sources in a variety of formats. A prototype map displays in space (location of homes and households) and over time (from 29 February to 31 March, 1692) accusers, accused, and accusations. When completed, users will be able to visually follow the spatial-temporal unfolding of the events that gripped Salem in the seventeenth century, and access detailed descriptive information on individuals, institutions, events, and related documents by clicking on icons on the map.

The proposals offered here are intended to initiate discussion and to invite criticism. Some of them, upon close scrutiny, may turn out to be impractical, simply not useful, or in fact already underway or completed. The list is also by no means comprehensive. Certainly worthy of discussion are projects devoted to developing and sharing pedagogical materials, to establishing one or more online, peer-reviewed journals, and to providing access to relevant social science datasets. No doubt there are many other promising candidates as well. Individual scholars, librarians, collectors, and students in the field are quite likely already undertaking important projects that would benefit from collaboration. Many scholars in the field, undoubtedly, could reel off a fairly long list of such projects, probably associated with the names of imaginative colleagues. Ultimately, the community itself needs to identify its most important needs, and determine whether and how collaborative use of technology can address them.

VI. CONCLUSION

The technological landscape has change considerably in the last twenty years. At the close of the 1970s, computing technology was generally not
available for most humanists. The equipment was expensive, and it required engineering and programming expertise that few humanists had the time or the inclination to master. The Internet was available to only a few, and was primitive when compared to today. Most computing was devoted to “number crunching”, for which most humanists had little or no use. Database applications were still quite crude by current standards. Markup and related text technologies and imaging, audio, and audio-visual technologies were only on the horizon. Standards were virtually nonexistent. All of this has changed significantly.

Particularly important among the technologies and standards that have emerged are those associated with databases and text markup. These two technologies, for the first time, make it possible for humanists to rigorously articulate structures that reflect their intellectual interests, to instantiate them in machine-readable form, and use the instantiations in computations that exploit the structures. For many years now, librarians have demonstrated the power of database technology for representing and exploiting descriptive cataloguing. Archivists, librarians, and humanities scholars have also successfully applied markup technologies for accurately describing and representing both the physical characteristics and intellectual content of cultural objects and collections of objects. While there are a number of existing and emerging standards already in place that would benefit the history of the book community, many others remain to be identified and developed. Database and markup technologies, in concert with the other technologies described here, present the field with the opportunity to determine its own future and the appropriate role for technology in it.

While technology presents great opportunity, it also presents very real dangers. Book historians cannot simply rely on technologists for guidance, especially when the technologists have conflicting interests. The community needs to develop its own technology experts. Because the technology presents such a wide-open space for the imagination, there is great risk that time and money will be invested in activities that ultimately are all form, and no content. This danger, however, should not stifle experimentation. The technology represents, in many respects, terra incognita, and to determine what does and does not work, and what is and is not useful, will require exploration. Experiments, especially those involving external funds, need to be carefully designed, with clear hypotheses and methods for evaluating
results. The most serious danger, though, is that fear of the dangers leads to doing nothing as a community. Inaction will have two probable outcomes. Individuals will engage in isolated projects that will make building a community at some later date much more difficult, as the individuals will be very reluctant to sacrifice their work. The other probable outcome is that technologists and other outsiders will determine the technological future of the history of the book community. Carpe diem!

NOTES

1 Unique is used here, of course, to mean unique in the context of the individual library, not unique in the bibliographic universe.

2 The list of variant names for al-Ghazzali and Shakespeare are both taken from the Library of Congress Name Authority File. There is a certain additional complexity that is not apparent in the two examples. Some of the entries represent cataloguer transliteration of non-Roman alphabet text (which may themselves be transliterations from Roman-alphabet texts), and others are transliterations in the published texts.

3 We find ourselves in much the same position as the protagonist in Goethe’s Faust. After Faust conjures up Mephistopheles, he asks him who he is. Mephistopheles answers that he is “part of a power that alone works evil, but engenders good”.

4 With a bit of creativity, the technology can be used to work with data for which it is not optimised. Some data is not easily classified, which is to say, it has characteristics of more than one type. For such data, it is sometimes best to base the selection of what technology to employ on the most important functional objectives, perhaps sacrificing less important objectives in the process.

5 The description of types of data that follows is derived in part from Steven J. DeRose, “Navigation, Access, and Control Using Structured Information”, American Archivist, vol. 60, no. 3 (Summer 1997).


7 For more information on TEI, see <http://www.tei-c.org/>.

8 For more information on EAD, see <http://www.loc.gov/ead/ead.html> and <http://jefferson.village.virginia.edu/ead/>.

9 For a list of current DTD initiatives, see <http://www.oasis-open.org/cover/xml.html> under “applications”.

10 “The World Wide Web Consortium (W3C) develops interoperable technologies (specifications, guidelines, software, and tools) to lead the Web to its full potential as a forum for information, commerce, communication, and collective understanding.” For more information, see <http://www.w3c.org/>.

11 For more information on XSL, XLink, and XQuery, see <http://www.w3.org/Style/xsl/>, <http://www.w3.org/XML/Linking>, and <http://www.w3.org/XML/Query>.

12 Metadata is first and foremost what librarians call descriptive cataloguing, used for providing intellectual access to and description and control of bibliographic entities. The definition of metadata has been extended to cover a wide variety of other kinds of control,
such as rights management, age appropriate filtering, and communicating and controlling the ordering and interrelation of complex-compound digital objects. This latter category is frequently called “structural metadata”. An example of structural metadata is the information need “to bind together” into a “book” large numbers of digital page images.

13 For more information on SVG and SMIL, see <http://www.w3.org/Graphics/SVG/Overview.htm> and <http://www.w3.org/AudioVideo/>.

14 For more information on SMIL, see <http://www.opengis.net/gml/01-029/GML2.html>. For more information on the Open GIS Consortium, see <http://www.opengis.org/>.

15 For more information on the Web3D Consortium, VRML and X3D, see <http://www.web3d.org/>.

16 For an excellent introduction to archive and library imaging, see Anne R. Kinney and Oya Y. Rieger, Moving Theory into Practice, Menlo Park, RLG, 2000.

17 For more information on ISO20000, see <http://eto.vub.ac.be/~chchrist/recpad00_paper.pdf>.

18 The expression “order of books” is taken from Roger Chartier’s The Order of Books, Stanford, Stanford University Press, 1994.

19 For a detailed description of the Bancroft Library’s specialized indices, see <http://www.lib.berkeley.edu/BANC/specialfiles.html>.

20 For more information on BiblioML, see <http://www.culture.fr/biblioml/>.


22 The International Council on Archives’ General International Standard Archival Description (ISAD(G)) defines a fonds as “a complex body of materials, frequently in more than one form or medium, sharing a common provenance”. For more information on ISAD(G), see <http://www.ica.org/ISAD(G)E-pub.pdf>.

23 More information on the Online Archive of California can be found at <http://www.oac.cdlib.org/>. The Physics History Finding Aids is organized and sponsored by the American Institute of Physics, with additional funding coming from the National Endowment for the Humanities. Information about the consortium and currently available finding aids can be found at <http://www.aip.org/history/ead/index.html>.

24 Malvine is a European initiative to provide access to “disparate holdings of modern manuscripts and letters, kept and catalogued in European libraries, archives, documentation centres and museums”. For more information on Malvine, see <http://www.malvine.org>.

25 More information on Archival Resources can be found at <http://www.tulane.edu/~miller/ArchivesResources.html>. Archival Resources currently provides access to approximately 20,000 findings, and is currently growing at a rate of 1,000 each month. The number of finding aids and contributing repositories expected to grow steadily and considerably for the foreseeable future, as more repositories begin to adopt EAD, with many of them choosing also to contribute them to this growing international databases.

For more information on the Museums and the Online Archive of California, see <http://www.bampfa.berkeley.edu/moac/>.


Ibid.


The William Blake Archive can be found at <http://www.blakearchive.org/>. While all of the plates are available for comparison when more than one copy exists, the following link will provide access to one example. Near the bottom of the screen is a “compare” button: <http://www.blakearchive.org/cgi-bin/nph-dweb/Blake/Illuminated-Book/MHH/mhh.f/@Generic__BookTextView/685;cv=java;pt=450>.

*Uncle Tom's Cabin* & American Culture can be accessed at <http://www.iath.virginia.edu/utc/index2f.html>, and the QuickTime® movie of an edition of *Uncle Tom's Cabin* be accessed at <http://www.iath.virginia.edu/utc/uncletom/editions/edhp.html>. Access the three-dimensional images of “the books on the shelf” by clicking on the spines, and click and hold down the left mouse button and drag to rotate and open the book.

The Brazil Rendering System site can be accessed at <http://www.splutterfish.com/sf/index.php>. In the “Gallery” there are a number of interesting examples, such as the following: <http://www.blur.com/blurbeta/brazilgallery/img_Lunarwolf_pump_test_7.jpg>.


The Public Record Office, the National archives of the United Kingdom, and in particular of England and Wales, acts as the nation’s memory. We select, preserve and make available to those interested in them the records of the state. The records, ever accruing, cover c. 160km of space and equal almost eight and a half million orderable units.

By 2005, in accordance with British government policy, we shall be able to deliver all our core functions electronically. Most initial access will be made through our web site (see http://www.pro.gov.uk): at the heart of the web site lies the catalogue to the records.

From the outset of our AD (Archives Direct) 2001 programme, designed to provide electronic access to the records, we were convinced of the need to build one unified, standards compliant, seamless, multi-level, web-based catalogue to the records. This would replace the plethora of existing finding aids (many dependent on others and none, apart from the printed information leaflets and irregularly produced microfiche editions of the high level Guide, available remotely) with a means of one stop resource discovery from anywhere in the world. Rather than hunting the index of the high level Guide for “publishing” and coming up with one series entry from the Department of Trade’s Publishing, Tourism, Films and Distribution Division 1974-1985 and working out from the Guide’s departmental administrative histories that references to the book trade, publishing and books might feature amongst the files of the 26 series from the Stationery Office or the 384 series from the Board of Trade etc., etc., or maybe trawling through the titles of information leaflets for a lead, people would be confident that a search within one system would yield comprehensive and reliable data.
Principle was, inevitably, simpler than practice. Agreement was rapidly
reached that ISAD(g) (the General International Standard for Archival
description, at that point still in its 1st, 1994, edition), supplemented
by a couple of extra data elements (such as scale) to cover non-textual
material, was to be our archival descriptive standard and that ISAAR(CPF)
(the International Standard for Archival Authority Records for Corporate,
Personal and Family names, 1996) was to be the basis for authority records.
The authority records for people, families, corporate bodies and families
would be constructed according to our own national rules and the UNESCO
Thesaurus (1995) would be the basis for subject indexing. We recognised
seven levels of description for the catalogue. The complications arose with
putting all the data together in the one system.

First of all c. 350,000 pages worth of individual series, or class, lists had
to be retroconverted into electronic form. Although the retroconversion
itself (in the form of rekeying) was undertaken by an external contractor all
the preparatory analysis, photocopying, marking up with coloured pens,
post retroconversion checking and loading into a simple holding database
was done inhouse. It is a salutary lesson to learn how much that is easily
interpreted by the human intelligence (indentations/capital letters/spacing
etc. to indicate different levels; \textit{ditto} or \textit{””} to indicate repetition of what has
gone before) is nonsense as far as a machine is concerned. All of this was hard
work and required meticulous organisation but was not of itself complicated.
Complications were caused by our only serious year 2000 compliance scare.
There was no guarantee that the hardware of our document ordering
system (\textsc{ris}) was year 2000 compliant. \textsc{doris} (or “daughter of \textsc{ris}”) had to
be built very quickly, using the simple series lists holding database as its
document inventory. Constraints were immediately put upon the eventual
design of the full multi-level catalogue: it had to be constructed around
the existing database; it had to interoperate with an already constructed
document ordering system.

Meanwhile the Guide level data, comprising full series level descriptions
cross-referenced to administrative histories of creating government
departments and agencies, had to be made ready for inclusion in the unified
catalogue. For some years the Guide had been maintained in an Oracle
database but its interfaces were never considered sufficiently user friendly
for use by others than \textsc{pro} staff. It was printed annually and made available
in hard copy on site, with occasional microfiche editions sold too. We decided to map the Guide data to Encoded Archival Description (EAD) (see the EAD official web site at http://www.loc.gov/ead/) and to edit the EAD files themselves into the required catalogue form. EAD was selected as a data structure standard for this work because it was designed for online archival finding aids, and because it was ISAD(G) compliant, was independent of proprietary software, made for data exchange and XML compliant. This all went very well, the only issue being a certain degree of difficulty in manipulating large EAD files.

An original expectation had been that EAD would feature largely in the final design of the unified catalogue. In fact its role is limited, partly because of the constraints already noted (fitting around an existing database; interoperating with other systems; problems in manipulating large EAD files) and partly because of worries about performance in transmitting large EAD files over the Web. Furthermore, new accessions of records to existing or new series are constantly being made as well as a steady stream of catalogue improvement work. Text files were not seen as the best vehicle for frequent or, on occasion, global editorial changes. Neither were off the shelf proprietary database packages a viable option: at the time (c. 3 years ago) none claimed to be able to deal with the sheer bulk of records held by the PRO. Instead our contractors (Quidnunc), working with staff focus groups, came up with a hybrid solution: a relational database capable of EAD import and export, and with an XML authoring tool embedded to allow all the formatting advantages of EAD editing for the large text fields (e.g., scope and content, administrative history). Verity search engine is used.

The new catalogue, known as PROCAT, has been available at Kew since October 2000 and, in its 1st release version, on the Web since March 2001 (http://www.pro.gov.uk, catalogues, PROCAT). It allows for everything from an absolutely basic keyword search, through varying degrees of refinement and sorting to complex field searches and very sophisticated Boolean searches. The catalogue can be browsed from any point and documents ordered for viewing. There are ready prepared popular searches as well as facilities to save personal searches and bookmark catalogue entries.

PROCAT is already something of a knowledge management tool (e.g., hits are sorted meaningfully to an enquirer before being presented on screen) but further developments are planned (most immediately moving from
Figure 1: PROCAT basic keyword search.

Figure 2: Results for PROCAT basic keyword search.
the rather static information leaflets that presently form the basis of the guided search option to much more interactive ones, guiding the enquirer through various possibilities.

Huge step though providing one seamless integrated catalogue to one’s own holdings is, it is still only the first step in terms of e-access. The next step, and actually already part of PROCAT, is to provide links to other known holdings of related material. A further route to improving access for end users is via common searching across a network, whether national, international, subject based or whatever, or one point entry, through a portal. A couple of examples will suffice.

Within the UK there is a growing National Archives Network, with strands for Scotland, England and higher/further education. A further strand for Wales is planned. The English strand, Access to Archives (or A2A) is housed at the PRO. Our intention for the 1st phase, up to March 2002, is to produce a virtual archives catalogue of the equivalent of 400,000 pages for England, principally through the retroconversion of paper based finding aids. They are all to be available from the one site (see http://www.a2a.pro.gov.uk for the beginnings) as well as locally. Common standards have been a prerequisite for the Network to ensure interoperability: common searching is the logical next step. In the meantime users need to search the A2A site.

Figure 3: An A2A search for “publishing”.
specifically to pick up certain data that may be of interest (see Figure 3). (NB: Because A2A is not primarily an editing system and because the constraints surrounding PROCAT’s build did not apply, the A2A system is an XML document management system with EAD absolutely central.)

Culture Online is a new British government initiative to bring together culture (as represented by museums, galleries, libraries and archives) and learning (see http://www.cultureonline.gov.uk for the vision statement). Technology, content and access are to be harnessed to facilitate searching, accessing, retrieving and working with cultural materials. There is to be an emphasis on quality and interaction: the PRO plans to participate actively.

Simultaneously there is increasing demand for direct access to digital images of the documents themselves. At the PRO the key project to meet this demand at present is the digitisation of the 1901 census returns for England and Wales, census returns being our most popular records. A pilot scheme, encompassing the 1891 census returns for Norfolk is available now (see http://www.census.pro.gov.uk), with the 1901 returns available from January 2001.

Fast on the heels of this very distinct project comes PRO Online, an umbrella project to fulfil the PRO vision of making records themselves available electronically by means of a high quality and socially inclusive service. Access will be widened and the original documents better preserved. By 2005 the PRO e-business strategy states that:

The PRO will have available in electronic format as much as possible of the original sources consulted by over 80% of our users. These include core genealogical sources – the censuses to 1901, military service and probate material.

The delivery system is to be completed in two stages, with completion of the 1st by September 2001. This phase will permit Internet access to especially selected records (wills dating from 1850–1858) which will be scanned form microfilm for the purpose of electronic delivery, together with other images digitally captured in response to specific user requests and the existing Image Library (though in the case of the Image Library the images will not be of a quality appropriate for commercial use:
those requiring commercial quality images will be referred to the Image Library itself).

For this service there will be a charging mechanism. The images themselves will be searchable via metadata created during transcription (effectively providing an 8th level of description for PROCAT) and will be linked to PROCAT. There will be a PROCAT interface but access will also be possible directly from the PRO Web site. The scanning is to industry standard TIFF file format, with conversion into PDF files for delivery.

The second phase will more fully integrate PROCAT and PRO Online by widening PROCAT’s searching capabilities to include PRO Online seamlessly. “Born digital” records, those transferred to the PRO as electronic records will be incorporated into the system.

PRO Online will be backed up by the office’s new e-preservation system which will hold the authoritative versions of transferred electronic records and those digital images created for access and/or preservation purposes.

The Public Record Office holds a wealth of material of potential interest to those researching publishing and the book trade. Our strategy for providing access to all requiring it has been to build one online catalogue, to provide links from it to other sites, to work with others in the development of networks, common searching and portals, and to provide the framework for digitising the records themselves.
MALVINE, LEAF AND KALLIOPE:
SOME CO-OPERATION MODELS

JUTTA WEBER
Staatsbibliothek zu Berlin
Germany

Manuscripts are one possible manifestation of the idea of a work. Thus dealing with better access to data about modern manuscripts is at the same time dealing with better access to data about other manifestations of the idea of a work, e.g., printed books. In this sense the projects mentioned below are models which can be useful when thinking about the establishment of new co-operation structures in other cultural heritage sectors.

The project MALVINE (Manuscripts And Letters Via Integrated Networks in Europe) has demonstrated that co-operation between most heterogeneously working institutions (libraries, archives, museums) is possible when there is an agreement upon reaching a final goal in smaller steps. MALVINE started by establishing a common European metadata format for the electronic search and retrieval of existing data files describing modern manuscripts and letters, both on the collection level and on the item level. The MALVINE search engine is able to retrieve data provided by the participating institutions adapted to the MALVINE metadata format via EAD/XML and accessible via 239.50. The MALVINE Maintenance Agency in the Staatsbibliothek zu Berlin will co-operate with National Nodes in all European countries in order to secure effective work in a common European responsibility in future.

The next step consists in the integration of authority file information into the MALVINE search and retrieval process. The project LEAF (Linking and Exploring Authority Files) will build a model for establishing a common European Authority File as a work in progress, based on actual user requests.

One example of a national node is Kalliope which in Germany is an open distributed union catalogue for modern manuscripts and letters based upon the use of cataloguing rules and national authority data.
The desideratum remains that still only a very low percentage of relevant documents is catalogued and accessible on the Internet. Digitised images of the originals are rare, national or international co-operation is much needed in order to avoid ineffectiveness and excessive costs. By initiative of the national nodes new co-operation structures between libraries, archives and universities should be established in Europe to find common strategies for enhancing and enlarging the actual data offer.

GENERAL BASICS
The strategic idea of the three projects malvine, leaf and Kalliope is a step-by-step plan:

1. To provide overall access to existing item level description catalogues about modern manuscript holdings via a search engine (malvine);
2. To enhance the results provided by the search engine malvine by the integration of authority data (leaf);
3. To secure maintenance and integration of more and until now unknown holdings by establishing national nodes and a maintenance agency (Kalliope);
4. To complete the data offer by providing step-by-step the digital images, at the beginning of the most important and then of every requested document (to be done on a local level in co-operation, to be funded on a national or international level).

The basic preconditions are as follows.

1. The idea of a work (literary work, scientific work) is subject to different manifestations during its way from the head of the inventor or author to the public. There might exist notices, hand written draft versions, fair copies, setting copies, typescripts, computer files, galleys, proofs, before the idea finds its way to the public in the form of a printed book. Since many changes of the idea of the work can take place during this way, all these stages and changes can be of high interest for those who are occupied with scholarly work about it. Institutions like libraries, archives or museums are collecting the very different manifestations of the work. Our cultural heritage consists in these different manifestations.
2. Only a very small part of all human ideas (fortunately?) ends up being published. Most of them are enclosed in a person’s head for ever, but some of them happily reach the status of a hand written text. In this status they may find their way to a restricted public (which may, in fact, prepare a printed edition of these texts). Only a very small part of this restricted public knows by itself how and where to find these texts or has the time to search for it personally in different places all over the world.

3. Enhancing access to these hand written manifestations of a work thus means that advertising is needed. Advertising to a librarian means cataloguing. Only when catalogued or explicitly described the work can be made known to a greater public. The more concrete and transparent a catalogue is and the more open it is to the public the more people will be able to use it.

4. Every work is part of a wider context of works dating from very different times, dealing with very different subjects and carried out at very different places of the world. Describing a work also means showing the work within this framework.

5. Describing a work means giving some relevant information about certain core elements of its manifestations: Author, title, date of creation, place of creation, subject, relations to other works, etc.

6. Describing a work also means using certain standard terms, e.g., the name of the author, the title, the place, the subject: the retrieval of all elements of the description should be as comfortable as possible. Comfortable in this context means that the user can be sure that in searching for a work one name, one title, one subject keyword will retrieve every existing manifestation of a work.

7. The elaboration of a core set of description elements has been done several years ago, the Dublin Core (DC) and the EAD (Encoded Archival Description) initiatives fostered the idea of a commonly acceptable data structure model. Certain specialisation and qualification nevertheless will be needed to add some special information about very special manifestations of a work.

8. The use of standard terms is on the point to leave its “homelands”, the libraries, where the use of authority information (person name authority files, subject headings etc.) since long time was experienced.
Every enhancement of the access to this authority information thus is one of the desiderata of co-operative work beyond the limits of libraries and beyond national frontiers.

9. The presentation of the work itself to the public, showing the original in an exhibition or the digital copy of it on the Internet may underline the necessity of doing advertisement for these works which otherwise are likely to be forgotten forever.

10. To achieve these goals international co-operation is needed, but it must be supported by national activities, co-ordinated by “national nodes”, access points responsible for the integration of as many institutions as possible into a network of cultural heritage information.

MALVINE (HTTP://WWW.MALVINE.ORG):
A CORE SET OF DESCRIPTION ELEMENTS

One of the most fascinating aspects of Europe’s cultural heritage is the fact that every country owns a considerable number of collections of modern manuscripts and letters, written by the most famous or less well known people, but all together building the backbone of our culture. As cultural life has never been isolated within one nation or region, all kinds of national and international relationships between these collections exist. Europe’s cultural history is defined by those relations, whose documentations are kept in as many places as there are institutions with archival functions.

The main goal of the project MALVINE is to open new and enhanced access to these disparate holdings of modern manuscripts and letters, kept and catalogued in European libraries, archives, documentation centres and museums. MALVINE is the name for a search engine which is the basis for building a network of institutions in Europe providing information about modern manuscripts and letters on an item level description and which is independent of heterogeneous technical solutions. The data available under the MALVINE web site are accessible from all over the world as if being a homogenous unified data base. A multilingual user interface, the agreement to a core element set of description terms enhance the access to modern manuscripts and letters in Europe considerably and offer a totally new service in Europe (cf. Appendix 1).
It is well known that institutions owning manuscripts apply a considerable variety of local cataloguing rules. This not only applies to different European countries but also to different institutions in one country. Nevertheless, the different cataloguing traditions in Europe are a fact which is not easily overcome. MALVINE shows that this is even not necessary and proves that different European traditions may well suit together under the conditions of a well organised overall view of them. The MALVINE solution in fact demonstrates, that big changes within the cataloguing traditions of the institutions can be avoided. MALVINE shows that unification on a certain level is necessary but that grown traditions and specific interests need not be given up in order to reach a broader public from all over the world.

The technical basis for the MALVINE concept is mainly based on already existing standards and s&r-components. The primary goal consists in achieving interoperability of different OPACS containing different types of data. This means that two levels of interoperability have been considered: technical interoperability and semantic interoperability. The basic problem consists in doing a “multi server search” at the same time, initiated by a Z39.50 based client system. An appropriate solution has been achieved by introducing a “retrieval manager” between the Z39.50 client- and server system. This software component is located at a “search server”. To achieve “semantic interoperability” the Retrieval Manager at the search server site has access to a “knowledge base”. The “knowledge base” includes the semantics of different OPACS as well as terms taken out of the MALVINE thesaurus.

The Retrieval Manager provides “slots” for the integration of different OPACS described by different Z39.50 application profiles. On the semantic level the MALVINE project has investigated the applicability of the Dublin Core (DC) Metadata Set. The development of a specific (Z39.50) MALVINE profile which allows the integration of different types of collections (bibliographic and object types) has been developed on the basis of DC. The consortium decided to promote the establishment and the future maintenance of this specific MALVINE profile. A “standard” OPAC system is available to be set up using a “MALVINE” toolbox to enable the participants of the MALVINE consortium who do not have a Z39.50 server system to connect to the network. This toolbox includes the generation of an OPAC connected to the MALVINE Z39.50 target system which provides a gateway
to SQL-databases. To “fill” this OPAC with data, a data entry and a data exchange system, based on EAD/XML are included in the “MALVINE-tool box”. Both modules have access to the OPAC data base either via an API or an ODBC connection.

The major service objectives of MALVINE are:

- Harmonised access to heterogeneous databases and European consensus in the basic information concerning literary archives, modern manuscripts and letters;
- Search & retrieval possibility on a collection and document level in different languages;
- Use of a unified Metadata profile, the MALVINE metadata profile;
- Use of a harmonised terminology in Europe;
- EAD/XML based support of data migration;
- Integration of users’ feedback on catalogue records.

After the end of the project the consortium has established a business plan for the exploitation of the project results and for gaining new partners to become members of the MALVINE community. The Co-operation model is based on a structure (see Diagram 1) which will include different levels of responsibility:

A MALVINE Maintenance Agency (MMA) located at the Staatsbibliothek zu Berlin will act as a Centre of Excellence for other European National Nodes which themselves will organise the participation in MALVINE of all sorts of institutions in their country. They will be responsible for the establishing and maintaining of national co-operation models and for the maintenance of the co-operation with the MMA in Berlin.

The MMA in Berlin is responsible for the maintenance of the MALVINE server and the intellectual support which will for example be necessary when new data providers will join the MALVINE community.

A search engine can only be as good as the data which are provided. Also in MALVINE the agreement upon standards is one of the indispensable
preconditions for a perfect co-operation. The work of the MALVINE Maintenance Agency will therefore put considerable effort in the development of the preconditions for the enhancement of data quality. The crucial point will be the use of authority data.

**LEAF (HTTP://WWW.LEAF-EU.ORG):**

**THE USE OF STANDARD TERMS**

At present authority information, of which name authority files are a very important part, is used in various ways in European libraries, archives and museums. Name authority records may be available as part of a national name authority file (e.g., in libraries) or only on a local level (e.g., in archives). Thus better access to name authority file information will close a huge gap of information in this sector. It is widely accepted that the national and international sharing of authority information is a suitable means for reducing the costs of cataloguing work in libraries and archives as well as the costs for biographical research work undertaken in scientific projects. At the same time the importance of allowing the preservation of national, local or rule-based differences in authority information is generally appreciated.

In other words the above means:
1. There is no common authority data usage in libraries, archives and museums
2. There is no commonly accepted model for authority data in Europe
3. There is no integration of users and user needs in the building of authority data
4. There are no possibilities for end users to utilise authority data.

**LEAF** started on March 1st 2001 and is carried out under the 5th framework in the IST program and will within the next three years work with the intention to enable institutions and end users to make better and extensive use of existing authority data. **LEAF** will develop a model architecture for a distributed search system harvesting existing name authority information (persons and corporate bodies) aiming at automatically establishing a user needs based common name authority file in a specific sector highly relevant to the cultural heritage of Europe. The project results will be implemented into the fully functional, international online Search and Retrieval **MALVINE** service network and aims at extending this into a global multilingual and multimedia information service about persons and corporate bodies based on user needs. The model architecture is intended to be applicable to other kinds of cultural/scientific objects and data, ensuring through the use of authority file information that the representation of the objects in question is one of high quality. The **LEAF** demonstrator will thus provide a valuable example of how dynamic user interaction with the cultural/scientific content can considerably enhance the user experience.

The following scenario might clarify the different benefits of **LEAF**:

- a user submits a biographical query to **LEAF** via a browser;
- **LEAF** queries the **OPACS** of the organisations;
- The organisations send back the results to **LEAF**;
- **LEAF** saves the positive results (name authority records *plus* information about where this data record originated from) in the common name authority file; this common name authority file will dynamically grow with every new user query;
- the user can transfer all or parts of the common name authority file to the users’ name authority records, this being part of the User
Work Space; because the system will be multilingual, the user may choose a preferred language of the authority file information; the user may make use of the conversion tool and receive the data in the desired format;
• the user may transfer (download) the data contained in the user’s name authority records to his or her local hardware platform (perfectly outside of LEAF);
• the user may offer additions, corrections etc. to LEAF; this data will be checked by an Intellectual control agency (this being part of LEAF), to be accessed via the Interface for control agency;
• the Intellectual control agency will forward to “LEAF Organisations” (i.e. the LEAF consortium): corrections and additions to existing data records; the “LEAF Organisations” may in turn either update the data also in the common name authority file or get the record deleted in the common name authority file to have it created again by another user query;
• the Intellectual control agency will forward to the common name authority file information about further locations of material related to the relevant person/corporate body (e.g., a small organisation without electronic data may add information about its manuscript holdings to an existing name record);
• the Maintenance Suite communicates both with the OPAC systems and the Search Server Administration.

In other words: The user her- or himself will not only benefit from the new LEAF service in a one-way fashion but will also be able to contribute to the existing information and thus enhance its quality. Every user’s query will automatically create a name record at the LEAF site that comprises the information of various available authority records about the same person/corporate body and additionally offers the information which institutions own relevant material and/or information related to that particular person/corporate body. The implicit information this newly created name record contains is that the particular person/corporate body was searched for. It therefore shows that this newly created name record is in fact of actual relevance. Thus a common name authority file will be built on the basis of what expert and public users are really asking for.
This means that the authority file information will not be “finalised” at a certain point in time but that it will be a work in progress. Further information about persons/corporate bodies will also be made available (biographical reference works etc.).

As information about persons or corporate bodies is not restricted to text information, a new range of additional multimedia information will enhance the data offer and make it more attractive. Digitised photographs, films, music, bibliographies or the results of research projects are only some examples of possible additions.

**LEAF** proposes the concept of user interaction as the basis of incorporating authority file information in a virtual database. By performing search queries the users will automatically and gradually create a specific data repository of enhanced authority file information that will:

- directly represent the users’ needs in the area of biographical/manuscript information; and,
- provide assurance of the data through the provision of individual work spaces.

**LEAF** will generate an adaptive model that can be used in other related domains. This model will utilise novel ways of:

- dynamically creating information;
- manipulating and managing the information by using and/or considering standards like XML in the area of user profile management and user work space management, WML for the purpose of remote access to the information via a WAP proxy etc.;
- creating virtual cultural/scientific landscapes by offering an access point to related data that can use the same authority file information as the centre for new search and retrieval scenarios;
- enabling both public users (end users) and expert users (staff in libraries, archives, museums, documentation centres) for the first time to harvest the same authority file information;
- managing the information by implementing a compatibility suite that will automatically process many user interactions aimed at integrating new data to the system.
The LEAF model will be tested through a demonstrator system. The model architecture is intended to be applicable to other kinds of cultural/scientific objects and data, ensuring through the use of authority file information that the representation of the objects in question is of high quality. The LEAF demonstrator will thus show how dynamic user interaction with the cultural/scientific content can enhance the user experience.

LEAF will contribute to a better dissemination and use of standards in Europe. The distinct shared common responsibility for the project’s outcome, the “Common name authority file” will foster the European idea in a very practical context.

KALLIOPE (HTTP://KALLIOPE.STAATSBIBLIOTHEK-BERLIN.DE):
THE NATIONAL NODE

The goal of the project Kalliope is carried out by the Staatsbibliothek zu Berlin and which is funded by the Deutsche Forschungsgemeinschaft and started in April 2001. The intention of Kalliope is the establishment of an open information system about modern manuscript holdings in Germany. Kalliope thus will play the role of a national node acting together with the German partner institutions as data provider in the European network established in MALVINE and LEAF. The design of the system structure allows for communicating with every institution keeping relevant material in the sector and for providing to every partner institution access to standards and to authority information. It also builds the bridge between the Kalliope consortium and the MALVINE/LEAF community.

Diagram 2 shows the co-operation structure, in which institutions can chose between:

- The use of a Kalliope client (free of charge);
- The exchange of data captured in their own system via MAB2;
- The communication between their own system and Kalliope via Z39.50;
- The use of a simple template.

With these four offers Kalliope will be able to communicate with every sort of institution keeping relevant material: smaller and medium
sized institutions which at the time don’t use an own cataloguing system are interested in using a Kalliope client; big institutions equipped with special cataloguing systems will be able to send and receive their data via the German exchange format MAB2; institutions working with powerful cataloguing systems will be able to communicate via the Z39.50 protocol on the basis of the MALVINE protocol which has been mentioned above. And very small institutions without any cataloguing facilities and which own only very few documents will use a template just to send the short information that relevant material is accessible at that place. The MALVINE protocol will also be used for the communication between Kalliope itself and MALVINE.

The German cataloguing tradition in the sector of modern manuscripts aims at giving information on the collection level and additionally on the item level of each document (when possible and meaningful) and is based mainly on two facts:

1. Cataloguing rules;
2. Authority data.

Cataloguing rules are the necessary precondition to give a structured form to the description of the documents. Authority data are the necessary
precondition to base the description of documents on an agreed terminology. Both are indispensable in a communication network which is built up with the goal to show specific material in a coherent framework. Both can be used at the moment only on a national level. In Kalliope this works very well. A very simple mechanism connects name authority records to corporate body authority records and gives thus information where documents related to a certain person are kept. This relationship is an 1 (person or institution) to n (institutions) relationship and is the basic structure of the model called “diana” (Deutscher Index zu Autographen und Nachlässen). The possibility of just adding (via an electronic template) the name of an institution (or a private person) which keeps relevant material of or concerning a person or institution is the most simple way of enhancing the information service of Kalliope.

Here the circle closes: The search engine malvine will be as good as an agreement to common data structures is considered to be obligatory or at least desirable. The more institutions will agree to give access to their data in malvine the more terms of comparison will be provided and the more examples of “how to do” will be given. This is in fact one of the expected outcomes of malvine: To give as much examples as possible how Europe’s modern manuscripts and letters are described and how they can be found on terms of this kind of description and thus to encourage new participants to do it the same way. Every participating country or institution will profit from this European overview in the long term.

The leaf project aims at establishing a European authority file. This means that authority data will have in the long term a good chance that their use will be more common and less difficult than it is today.
Kalliope demonstrates that preconditions for constructive and well organised participation in the realisation of a European or international co-operation model national preconditions have to be established which are fit to include more than the biggest and best known institutions into this strategic goal. Only when all relevant institutions—also the smallest ones—are able to participate in the realisation of a European or even world wide initiative the desired impact of the “information society” will have a chance to really being achieved. This also means that every kind of institution—museum, library, archive, documentation centre, scientific institution—must have the chance to provide their information in a suitable way.

The desideratum remains that still only a very low percentage of relevant documents is catalogued and accessible on the Internet. Here much engagement and even more funding is needed which at the moment seems to have to be organised by national nodes and national funding organisations. But also in this context international co-operation could stimulate financial support in a very positive way. The same strategic deliberations are fitted for another aspect: Digitised images of the originals are rare, local initiatives are restricted by lacking funding possibilities. National or international co-operation is much needed in order to avoid ineffectiveness and excessive costs.

In what concerns costs: By initiative of the national nodes new co-operation structures between libraries, archives and universities should be established in Europe to find common strategies for enhancing and enlarging the actual data offer. Co-operation is good as far as there is enough money available in order to act as rational as desired. To avoid an untimely end of the realisation of a good idea there have to be provided structural preconditions which are fitted for both securing a long term financial and legal basis. We can achieve this in Europe and in the world only when trying to see things in a bigger context and to act in a more global way than we are accustomed to today.

NOTES

*Projects and initiatives*

DCMI (Dublin Core Metadata Initiative) <http://www.dublincore.org/>.

ead (Encoding Archival Description): <http://lcweb.loc.gov/ead/>.


FRANAR (Functional Requirements And Numbering of Authority Records, Working Group on Authority Control created under the auspices of the Division of Bibliographic Control and the IFLA UBCIM Programme): <http://www.ala.org/alcts/alcts_news/v10n1/gateway.html>.


Kalliope (Offenes Verbundsystem für Nachlässe und Autographen in Deutschland) <http://kalliope.staatsbibliothek-berlin.de>.


oai (The Open Archives Initiative): <http://www.openarchives.org/>.

UKOLN (The UK Office for Library and Information Networking): <http://www.ukoln.ac.uk/>.

Publications


APPENDIX

Some samples of the search and retrieval facilities in MALVINE (see following pages).
The idea of MALVINE is to build a network of these institutions in Europe, a network which is independent of heterogeneous technical solutions and which is accessible from all over the world via Web based technology as if being a homogenous unified database.

A multilingual user interface will be provided, using an agreed common terminology.

MALVINE and LEAF. Gateways to Europe’s Cultural Heritage.

On the 4th and 5th of December 2000 the Berlin State Library will host the International Conference “MALVINE and LEAF: Gateways to Europe’s Cultural Heritage”. This event will be a platform to present and discuss the results of the EU-funded project MALVINE and the intentions of the planned EU project LEAF (Linking and Exploring Authority Files).

Figure 1: The MALVINE start page

Figure 2a: The simple search facility
Figure 2b: The simple search result set

Figure 2c: The short title list
Figure 3: The extended search facility

Figure 4a: The biographical search facility
Figure 4b: The short list of results of the biographical search
TWO CATALOGUES, THREE PROJECTS —
AND A TENTATIVE PROPOSAL

SIMON ELIOT
Centre for Writing, Publishing and Printing History
Reading University, UK

I TWO CATALOGUES

1.1 BRITISH BOOK TRADE ARCHIVES LOCATION REGISTER

One of the last jobs I did as founding editor of the *Book Trade History Group Newsletter* was to give over virtually the whole of issue 12 (January 1991) to a list of archives relevant to the British book trade 1830-1939, which was compiled by Dr Alexis Weedon. This represented the first stage in mapping out the archival resources available to volumes 6 (1830-1914) and 7 (1915-2000) of the *Cambridge History of the Book in Britain*.

The consequences of this publication were wider and more diverse than we had anticipated. Firstly, it raised awareness of the importance of the less regarded parts of archives, most notably, production and financial information (which in the past had been discarded in favour of in- and out-letters). Secondly, it directed more higher and research degree students towards the subject. Thirdly, it stimulated a similar project in the USA (the Albinski list). Fourthly, and most importantly, it sensitised librarians, archivists and others to these very particular sorts of materials. In consequence, the second edition of the Location Register, published as a HOBODS (*History of the Book — On Demand Series*) in 1996 with Mike Bott of the University of Reading as co-editor, was more than twice the length of the first edition of 1991.

A third edition is now being worked on. It, too, will be larger than its predecessor though not, on this occasion, double the size. This time its publication medium will be the Web. This will allow us to update the Location Register on a regular and frequent basis. More importantly, we
shall also be able to convert the Location Register into a form of Book History archive portal that will incorporate—a hot link to the appropriate catalogue or register. Thus an entry for the nineteenth- and twentieth-century publishing firm of Chatto & Windus will include a link to the University of Reading Library’s catalogue and to its archive site where a partial catalogue of the archive exists in machine-readable form. Where no catalogue exists it might still be worth including a link to, say, the UK’s National Register of Archives through which a user will at least be able to browse all the information that has been gathered on the archive. All this will not be done at one go: the Location Register will first be placed on the Centre for Writing, Publishing and Printing History’s Web site as an augmented third edition with a few obvious links included. Additional links will then be introduced in phases.

All that having been said, we shall maintain the hoboids tradition of on-demand publishing by printing copies of the register from the Web whenever an un-networked book historian requires a copy. It is important that we do not forget that some very able and enthusiastic book historians, certainly of the older generation (many of whom are amateurs in the best sense of that word) are still not, and probably never will be, au fait with the Web. We should make sure that our high-tech enthusiasms don’t sweep them away unregarded.

Of course, although a Book Trade Location Register is a good in itself and provides extremely useful initial information, as soon as it becomes more a portal and less an end in itself—which is inevitable once it is mounted on the Web—then it raises expectations about what it can do, and those expectations are likely to be frustrated unless the catalogues and finding lists it directs users to can actually deliver the goods.

What do we mean by “deliver the goods” in this context? Ideally, the Location Register should be able to send a scholar to a catalogue that would enable him or her to identify a particular document, say a letter to an author or the production record of a particular title, even perhaps then be able to call up a digitised image of that document.

What follows is a number of examples of existing or planned projects in the digitisation of book history materials that illustrate the possibilities and problems of delivering such goods. Each example I give will, I hope, suggest one particular promise and one particular problem raised by the
application of IT techniques to book history archives. The promises I believe are obvious, so at this point I will adopt the devil’s advocate role and simply list a few of the problems in the order in which my examples raise them. Some are common to almost all projects of this sort, some are specific to book history archives.

The first is the problem of managing the revolution of rising expectations in the users of new IT resources. The second is the possible copyright problem associated with making previously unpublished manuscripts available on the Web. The third is the need to retain the original manuscripts and artefacts and ultimate access to them. The fourth is the problem of instituting practical, consistent and uniform procedures and forms of project control within and between digitisation projects. The fifth is the problem of dealing and negotiating with owners of commercially significant archives not in the public domain.

1.2 THE PUBLISHERS’ ARCHIVES IN READING UNIVERSITY LIBRARY

The following descriptions are based on information provided to me by Mike Bott, archivist at Reading.

1.2.1 Item Cataloguing

*Elkin Mathews*

The Elkin Mathews collection, a rather small one by Reading standards, was catalogued by the then archivist, Jim Edwards, in the 1970s and 1980s. The letters which make up most of the collection are calendared.

*Macmillan*

The first tranche of Macmillan letters was dealt with rather differently. It was known that the two hundred or so file-boxes were a series of chronological / alphabetical sequences, but most of the boxes had lost their spines and there was no space for the cataloguers to sort the boxes back into their proper sequence. They were simply taken off the shelves one by one and for each letter a catalogue card was made recording the name of the correspondent, date and location code. Letters from the same person can thus be scattered through any number of boxes. The only good thing that can be said for this
is that letters from individuals or corporate bodies are retrievable. There is currently no subject index to the letters. Nor has any authority for names been used. Apart from anything else it would have been a nightmare trying to trace many of the individuals whose works may or may not have been published! Checking was perfunctory and it is possible that the same person might have been recorded under different headings, e.g., Smith, J.; Smith, John. Mis-read handwritten signatures generate additional possible errors.

1.2.2 File level cataloguing

The most common way of cataloguing has been at file level, where a file might correspond to a book or an author or a subject. The most straightforward files are represented by the Bodley Head and Hogarth Press catalogues, mostly book files, the contents of which have been described and indexes of correspondents compiled.

1.2.3 File level cataloguing on unicorn

There are five similarly-structured catalogues that are distinguished from the others, and this is because they appear on Reading University Library’s general cataloguing system UNICORN. The five are the archives of: Allen & Unwin, Chatto & Windus, Routledge, Heinemann Educational Books and Jonathan Cape.

   Allen & Unwin
Allen & Unwin has been done in its entirety; it consists of 37,489 records (this is large because all the readers’ reports have been catalogued individually).

   Chatto & Windus
Random House is currently supporting a cataloguer to work on the Chatto & Windus archive, with the possibility of later doing some additional cataloguing of Jonathan Cape. Currently there are 4,160 records.

   Routledge
Routledge has been partially catalogued; currently there are 1,588 records.
So far this consists of records of African and Caribbean Writers series and the Windmill series. Currently there are 869 records.

We are hoping that the Random House cataloguer will finally move from Chatto to Cape and so expand the currently pitifully small set of records. Currently there are just 283 records.

One of the great strengths of all these records being integrated into UNICORN is that they can be searched along with printed books. This not only provides the proper and balanced digital context for book trade mss which clearly need to be studied along with the printed consequences of the processes implied by editorial files and production records, but it also increases the chance of serendipity. For example, last term I was searching UNICORN for printed books issued by the mid-Victorian publisher John Camden Hotten. I knew of and had worked with Reading’s unparalleled collection of Hotten manuscripts c.1867-73 and had a good working knowledge of the Chatto & Windus files up to 1914. However, my broad search of UNICORN for Hotten pulled up a Chatto & Windus file of 1952 which alluded to correspondence about an – up till then – unknown journal kept by Hotten as a young man.

### 1.2.4 Hybrid archives

Some archives have lent themselves to this file level cataloguing in part, but where an archive includes many different types of material, different strategies need to be adopted to deal with them. For instance:

**The George Bell archive**

This is a hybrid archive, consisting of three main types of material: bundles of incoming nineteenth century letters, files of correspondence maintained by each of the firm’s directors, and miscellaneous material including ledgers. The Bell catalogue consists of an inventory of the whole collection with a large index of personal and corporate names covering both the bundles of letters and the correspondence files.
The Longman archive

This has been dealt with in a similar way except that no full index has been prepared—many of the files are in alphabetical sequences and are therefore easy to access.

General

Much material in several collections remains uncatalogued, though handlists of files exist for most deposits; others are arranged in chronological/alphabetical sequences, e.g., much of the Routledge archive, and most of the Heinemann Educational Books archive have no listing at present.

Collection level descriptions are to be created for all the collections to be attached to the somewhat perfunctory descriptions which at present appear on the library’s Special Collections web site. This project is under way and is scheduled for completion by the beginning of next session.

Cataloguing—whether it is done in guardbooks, on 6 x 4 inch cards, in word-processor files, or in full relational databases; whether it is done in MARC or the latest version of EAD—requires substantial amounts of time and money. But, as we all know, the creation of electronic information is not a neutral act of transference. Transmuting data from manuscripts or print into electronic form creates a wholly different and higher set of expectations in the user in terms of speed, access and detail. Faced with this revolution of rising expectations in users we have three possible courses of action. One, we can tough it out: dump what we can offer on to the Web and indicate that scholars are lucky to have it. Two, spend some time engineering the expectations of users so that they understand the enormous constraints under which any library and archive work. Three, raise our game, find more money and go for something much more ambitious than that which has satisfied users in the past. Clearly the third course of action would be the preferred option for most of us, but this presents considerable problems as my next examples will illustrate.
2 THREE PROJECTS

2.1 VADS VISUAL DATABASE: VICTORIAN SHEET MUSIC COVERS

Reading University Library (RUL) has been involved in a recent project with a Book History archival twist. In 1997 the Visual Arts Database Service (VADS) in the UK wished to provide some form of digitized visual material (music covers, fashion, design, photographs mostly nineteenth and twentieth centuries) for use in teaching and research in the higher education sector in the UK. The project was completed in 2000. Reading University Library’s contribution to this was the selection of 800 Victorian sheet music covers derived from the Spellman Collection.

Funding was provided by JISC; scanning done at another university (Hatfield); a centralised group of editors was provided; and quality control also made available. But despite all this external funding and support, the task facing RUL proved quite a challenge.

The database that provided the index and catalogue to this collection was demanding: the visual content had to be indexed and—despite the availability of such reference works as the Getty thesaurus—this was a lengthy job.

Only a selection of the RUL’s collection of music covers was used so the selecting of examples took time and required getting the balance correct by using a set of inter-locking criteria: social history content, interesting in terms of graphic design and colour printing, aesthetic considerations—and also having to find examples that might be regarded as typical.

Experiments were made with scanning at 600 dpi. However, the resultant files were too large to be handled easily. Three hundred dpi was finally selected but even at this rate the 800 music covers took up 55 CD-ROMS-worth of storage.

The digitisation brought many advantages to Reading University Library: it was able to avoid having many users calling for the originals and good cataloguing made searching much easier. The partial downside, as with many digitisation projects, was that it also raised expectations. For instance, the catalogue had included references to Scottish themes but this was not satisfactory for a researcher who was trying to survey the Victorian representation of kilts.
The greatest problem was raised by copyright. Late Victorian music covers might have been designed by people who lived into the 1950s and 1960s and therefore the covers might still be in copyright. This was regarded as so serious a problem by those running the project that, if there were any doubt concerning the copyright of a cover, that cover was left out of the digitized collection.

This last point, of course, raises serious questions about reproducing an unpublished manuscript which is normally regarded as having perpetual copyright until published. Machine-readable catalogues, however high and detailed their level of description, is one thing; verbatim transcription or the publication of a digitized image of the document in question is quite another. How are we to get substantial amounts of mostly manuscript book history material on to the Web when its copyright, at least in the UK and at least in theory, belongs to others who may or may not be traceable and who may or may not give permission?

2.2 BOOKHAD

The rul is also involved in a broader project, BOOKHAD, whose aim is to increase and enhance personal access to research resources in the areas of book history and book design. BOOKHAD is likely to have fewer copyright problems because the majority of its material will be in the form of catalogue entries from its constituent members’ libraries and archives.

As BOOKHAD is simply making mostly existing material—such catalogue records of Book History materials—more readily available, the problem of copyright will not occur very often. However, as it has links to various databases containing digitized images of manuscripts, it is going to have to hope that the host libraries have done their reasonable best to establish copyright permission.

Indeed this is the problem: with most of the material we shall want to digitize, the danger of being taken to court by some irate but, until that time undiscovered, copyright owner is slight. However, in order to make your project safe from litigation, you will have to prove that you have made all reasonable attempts to establish the ownership of copyright. These “reasonable attempts” are likely to involve a large amount of fruitless but necessary administration and that is going to add a distinct, and possibly
substantial, cost to any digitisation project that relies on making copies of previously unpublished manuscript material (such as letters) freely available on the Web.

However effective a project such as BOOKHAD becomes, and however many sources of digitized information it gives us access to, we must remember that digitisation projects in book history must always be regarded as a process of augmentation, not a process of substitution. No digitisation project should justify making the originals on which it is based either wholly inaccessible or, worse still, justify the destruction of those originals. Digitisation can justify an archivist suggesting that copies rather than the originals be used for most forms of research, but there are types of research and types of question that can only be resolved by the analysis of the actual artefact. Paper analysis, binding studies, all questions involving book structures, will rarely if ever be answerable by even the most highly resolved of digital images. Nicolson Baker recently gave a vivid example of clandestine newspapers produced in occupied France. Paper stock was difficult to get hold of and was derived from a range of different sources. The nature of the paper mix between different issues and even within a single issue (and thus likely sources of supply) could only be resolved by close study of the artefact itself. This is an obvious point to make to scholars, but it is not so obvious to politicians, civil servants or those in commerce who are frequently attracted to an IT solution because they fondly believe that ultimately it will save them time and money. In such minds the concepts of “saving” and “efficiency” frequently takes the form of believing that, after digitisation, they will be able to junk the bulky, slowly deteriorating, difficult-to-manage, expensive-to-store originals. In our keenness to acquire government and commercial support we should never be tempted to sell our scholarly birthright for a mess of potage, however favourable the mess of potage to Euro exchange rate happens to be at the time.

2.3 PROPOSED RANDOM HOUSE DIGITISATION PROJECT

A few years ago Random House decided to set up a single location at which it would integrate the records of the various publishing firms that it had taken over. The Random House Archive and Library was established at the Sanders Lodge Industrial Estate, Rushden, Northamptonshire. In the last
couple of years this has moved into a large, purpose-built warehouse on the Estate. Its main function is to support the publishing activities of the Group and thus its preoccupation should be with literary properties that can still be exploited – in other words, with those books that still have active copyrights associated with them. Despite this, Random House’s strategic policy seems inclined to bring to Rushden all the archive material of its constituent companies, even that material published in the 1860s and 1870s that has no copyright significance whatsoever.

In some ways this is admirable: here we have a major publishing company spending significant sums of money on integrating, housing, and cataloguing a diverse range of archives. The problem is that this integration means that major historical archives, such as those housed in the University of Reading, will finally be moved to an obscure location in Northamptonshire miles and miles away from any university that might be able to make use of them.

That having been said, the Company is doing its best to make the material accessible. As part of this process, and perhaps in a rush of high-tech enthusiasm, a year or two ago it was proposed within the Company that Random House should digitize its entire archive. It is not clear how high this proposal got up the hierarchy, or how seriously it was taken, but I can report that the Random House Library Manager, Mrs Jean Rose, has expressed serious doubts about the feasibility of such a grand project. She has proposed a more modest approach that would see a selected range of author files digitized. Even this more modest scheme would have to confront the messy, mixed, multiple and muddled nature of publishers’ archives. As Mrs Rose suggests: “... the sheer amount of paper in each file, duplication of paperwork in files and the vast range of paper sizes ...” is likely to cause all sorts of problems in process control. Currently Random House is migrating its catalogue data to a new database, so the project is on hold.

That having been said, the higher levels of management seem still to harbour greater ambitions and have visions of a grand scheme. As Mrs Rose has remarked to me: “the management would like to do all of [the archives] but they have no concept of how long it will take”. This is surely a useful object lesson in the need to understand the full implications of any digitisation scheme in terms of workload, cost and user expectation. That the sophisticated directorship of Random House displays such a touchingly
innocent optimism—probably driven in part by a fashionable devotion to IT—should alert us to a dismal fact. Namely, that senior management decisions, even ones that often profoundly affect the fate of publishers and their archives, are frequently made on the slenderest evidence and driven by the shortest-lived of management theories or enthusiasms.

It goes without saying that we should be pleased that those who direct publishing firms should be so enthusiastic about increasing accessibility by using large-scale digitisation. Random House’s first aim is, of course, to make their own support services more efficient, but one of the by-products of this process would certainly be improved access to book history materials for scholars. The danger is that—if this process proves to be as difficult and as demanding as our past experience suggests—then the project will go off at half-cock: it will get bogged down, criteria and practices will be changed half-way through, it will run out of steam or it will be abandoned. Worst of all, it may leave a bitter taste in the mouth of such firms as attempt to undertake it, and that will have serious consequences, particularly in the case of possible sources of funding, on any future projects to digitize in a controlled, measured and effective way.

One of the principles that I think we need to embrace is that, once formed and agreed, we need to proselytise our views on the digitisation of book history archives. It is all very well creating international standards and agreed protocols if significant book trade firms know nothing about them and care even less. We should not assume, for instance, that those in charge of a firm’s archives have library or archival training or, even if they had been so trained in the past, that they are up-to-speed in terms of the latest standards.

In the UK, about a couple of decades ago Tim Rix, then CEO of Longman, and Lord Asa Briggs, then just beginning to write the history of Longman, held a series of dinners for publishers. The aim of these dinners was to alert publishers to the importance of their archives. These dinners had some success and certainly increased the flow of archives into the public domain, most noticeably into Reading. I suspect that we shall need another similar, but perhaps international, series of meetings that would allow us to get across to the firms of today the importance of doing digitisation well and consistently. We will have to be pragmatic and down-to-earth, we will have to think of them as businesses first and holders of archives second, but if we
can engage with the publishers then we give ourselves the best chance of creating a network of tolerably consistent, exchangeable digital databases that will make the job of writing national and international histories of the book a much easier and a much more productive process.

Clearly all the exercises and projects so far described are worthy and admirable and, whatever their problems and limitations, deserve our enthusiastic support. In conclusion, however, I want to suggest, not an alternative, but a complementary way of the using the Web to promote the development of book history archives. This latter approach does not simply aim to reproduce in electronic form what libraries and archives have always striven to do, but explores some of the unique features of the Web to do something that would have been almost impossible without it.

### 3 A TENTATIVE PROPOSAL

#### 3.1 A VIRTUALLY RECONSTRUCTED ARCHIVE (VRA)

As a past president of the Society for the History of Authorship, Reading and Publishing (SHARP), I have always been impressed by the ways in which that Society has been able to function as a “distributed system”, that is, one that does not have a single, physical co-ordinating centre. I live in the UK, most of the other Executive Officers are scattered around the North American continent. Our Directors span the globe and our membership is equally widely distributed. Nevertheless our electronic discussion group, SHARP-L, and the wonders of email keep some 1100 members scattered across more than thirty countries in touch with each other on a daily, and when some controversial matter is being discussed on an hourly, basis.

As President it had long been my ambition to add to SHARP’s many activities by creating a SHARP research initiative. Such an initiative would have to reflect SHARP’s interests and its international profile, and would have to play to the Society’s strengths. One of SHARP’s strengths, of course, that it does much with a rather modest income, so any research project would have to be characterised by very low costs. Putting all these things together, I began to think of the possibility of reconstructing a book history archive virtually. Let me explain what I mean by this phrase.
Like many material-based subjects, book history is haunted by the might-have-beens from the Alexandrian library to the most recently pulped publisher’s archive. Let me take an egregious example. The UK publisher and wholesaler Simpkin Marshall was of great importance nationally and internationally throughout most of the nineteenth century and the first half of the twentieth century. In the 1940s it suffered the double whammy of being bombed in the Blitz and its remnants later being taken over by Robert Maxwell. Unsurprisingly, there is no surviving archive.

Simpkin Marshall had dealings with a multitude of other publishers, printers, wholesalers, libraries, etc., throughout its vigorous life. There must be hundreds, probably thousands, of letters from Simpkin Marshall, and copies of letters to the firm, buried in the surviving archives of the book trade in the UK and elsewhere, along with other material that would help us reconstruct this lost company by using database software linked to SHARP Web.

The idea would be that SHARP would nominate a volunteer scholar who would then set up a standard, SHARP-approved Access-based database (with automatic archiving and back-up) on a server at his or her home institution and who would then act as the director of the specified project. Controlled access to the database would be through a research page on SHARP Web and material could be sent to it either through this route or via a standard email. Once the database had an agreed minimum number of records, it could be made available in the public domain for searching.

As colleagues investigate surviving book history archives for their own research topics, they are occasionally likely to come across material relevant to Simpkin Marshall. It may be a document they have to transcribe for their own purposes, it may be something they note in passing (perhaps just a file number). If we had alerted colleagues to the fact that there was a place to which they could email this information, they would be more likely to respond to it and record it, however briefly. If we could offer privileged access to the database to those who contributed to it, then we could provide a not wholly selfless motivation for the collection of the data. The material stored on such a database might consist of just a reference, it might be a partial or full transcription, it might even be a digitised image.

If we could establish such a system, it would be like building up a picture of an extinct creature by studying its fossilised tracks or understanding
Martian climate and geology by studying a meteorite from that planet. Not as good as having the real thing, but a huge advance on our current state of knowledge.

Slowly we would accumulate sufficient information to begin to reconstruct the impact the lost firm had on surrounding firms and the book trade as a whole and, through that, some idea of how it worked internally. These “Virtually Reconstructed Archives” (Vras) could be one model for how co-operative scholarly research for mutual benefit could work on the Web.

The transmission of such data could be done electronically, as could its collection, and that is very much in the tradition of SHARP. It could and should and must be international, and that is also a strength of SHARP. Most contributions to such databases will be from individuals (appropriately acknowledged in the database record) but the database itself will have a growing collective scholarly significance. All this sounds a typically SHARP endeavour to me.

I am proposing to SHARP that it considers initially setting up a small number of Vras that together would cover as wide a range of SHARP interests as possible by covering a number of continents and a number of periods. Not too many as to become diffuse and complicated, but enough so that the project will be broad and inclusive. Say, three to four firms to begin with, each of which would have its own database and a named member responsible for it within SHARP.

I originally launched the idea via SHARP-L and the response was on the whole very positive. As a way of kicking things off I then proposed to the SHARP conference at Williamsburg, Virginia in July 2001 that I set up the Simpkin Marshall Vra at Reading University. Currently plans are going ahead on this basis and I hope to make a formal announcement some time in 2002.

3.2 AN ELECTRONIC COLLECTION POINT (ECP)

In addition to a small number of trial Vras, I am proposing that we also experiment with a variation on this procedure and create one or two “Electronic Collection Points” (ECPs). This would be an attempt to collect together, at a similar virtual point, documentary evidence that had always
been—both in origin and nature—scattered, but that needed to be assembled at one point and in large enough quantities to be of some historical and statistical significance. The Reading Experience Database (RED), which I direct, is an example of such an ECP, as is the Book Production Database 1830-1914 at the Open University.

Other possibilities, however, abound. My SHARP colleague Patrick Leary is particularly interested in using such a system to assemble scattered data about the financial details of contributors to magazines in the nineteenth century, particular rates of pay and nature of contracts. Indeed, a database on author income before and after copyright legislation would be an invaluable source of information for most book historians.

Both the VRA and the ECP offer models of how co-operative book history research for mutual benefit and at relatively low cost might work on the Web. Even more importantly, both VRAs and ECPs would have the power to convert an accident of research into scholarly serendipity—surely one of the happiest and most desirable outcomes of using the Internet as an academic tool.