

# **DIGITAL LIBRARIES**

**An Overview  
With Special Reference to  
English Studies**

**By**

**R.C. Alston**

**October 2002**

**The summation of human experience is being expanded at a prodigious rate, and the means we use for threading through the consequent maze to the momentarily important items are almost the same as in the days of square-rigged ships. We are being buried in our own product. Tons of printed material are dumped out every week. In this are thoughts, certainly not often as great as Mendel's, but important to our progress. Many of them become lost; many others are repeated over and over. [Vannevar Bush, *Science is not enough*, New York, 1967.]**



Ever since the great library at Alexandria was destroyed there have been periodic attempts to accumulate a truly comprehensive collection of books and manuscripts. John Dee petitioned Queen Mary to found just such a library for the English in 1555, but the embattled Queen had other things on her mind that year. In the 1850s the Society of Arts campaigned for a catalogue of books based on the collections in the British Museum and the major imperial libraries in Europe, and tried to persuade Prince Albert to be its patron. As I write this paper plans are being made for the inauguration of the Bibliotheca Alexandria in Egypt, a project started in 1990 and now getting close to opening.<sup>i[1]</sup> Since 1990, however, librarians have been moving ever more closely towards the notion of a seamless digital library (united by extensive hypertext links) based not in any one place but wherever research resources are to be found. In theory, every library and archive on earth has an opportunity to participate in this venture which, for complexity, rivals that of the Universal Postal Union. Whether this will happen is, of course, uncertain: but there are signs that the first building blocks of such a grand scheme are in place.

There is no need to reiterate here the reasons why libraries with comprehensive collections on the scale of the British Museum or the Library of Congress face an uncertain future: the reasons are well known and concern the ever-increasing costs of metropolitan *space*, and *staff*. Once the bulky card catalogues disappeared from libraries in the early 1990s it was inevitable that ingenuity would be directed at getting rid of the far bulkier books, and the rapid advances in digital photography and mass electronic storage made it seem that conversion to digital format was the only practicable solution. The logic of such a solution has seemed self-evident to many library administrators, even though it ignores two profoundly significant facts about the materials housed in libraries: these materials, whether printed or manuscript, are artifacts with important characteristics not easily reproducible (e.g. paper/parchment), and not all copies of books printed before 1900 are identical. The digital solution may, indeed, simplify problems of access to a text, but for scholarly purposes it will still be necessary to preserve all identified copies of literary texts. In other words, the contemporary librarian is being expected to subscribe to the virtues of building a network of digital collections accessible from anywhere on earth as well as preserving the artifacts inherited from the past. This, I suspect, will prove to be an insupportable burden.

There are persuasive arguments favouring the creation of digital libraries, and some of them can be traced back to Vannevar Bush who, in his landmark essay on the future “As We may Think” published in *Atlantic Monthly* in 1945, said that extracting information from the great research libraries was like “a stone adze in the hands of a cabinet maker”. Bush dreamed of a device (which he called Memex) that would make all knowledge

readily available when and where one needed it. In some respects, the World Wide Web is a gigantic Memex, with information pouring into it from millions of sources and consulted, at any time of the day by countless millions of people in search of *something*. Given the apparently inexorable reduction, in real terms, of library budgets in the past ten years, it has seemed to some that a collaboratively constructed network of digital versions of what research libraries possess on paper or parchment might well provide future scholars with more information, readily available from the office or home, than has been possible for researchers using the traditional means of gathering information. This is undoubtedly possible, but whether it is practical, or affordable, is a different matter. Take some of the problems which must be overcome before there is any likelihood of there being a practical, *affordable* alternative to things as they stand.

1. Until every document/text has been accurately converted to digital form it will be necessary to continue maintaining the present stock of all such libraries and archives which possess materials germane to historical research or enquiry. In other words, even if the most optimistic forecasts come true, we have probably a century of effort and funding before the digital libraries of the world can supply the needs of those engaged in research of any kind. It is simply not possible to conjecture the burden of cost this will place on the largest of our research libraries.<sup>ii[2]</sup>

2. Conversion to digital form must be accompanied by adequate cataloguing, preferably to internationally agreed standards. This will be no trivial task, since cataloguing an electronic source is just as demanding as cataloguing a book, a microfilm, a film, or a video. While MARC standards exist for virtually all library and archive materials, there is not, as yet, any agreed standard for digital conversions. Such a standard would seek to provide information comparable with a library catalogue, but with additional technical data: type of conversion (bit-mapped/ASCII), resolution (dpi), type of file (binary/ASCII), source, &c. For materials printed before ca. 1850 it will be necessary to specify the precise copy used in the conversion, since copies of hand-printed books are seldom identical. A serious problem may prove to be one of the characteristics of the Web: the instability of URLs.<sup>iii[3]</sup>

3. A question not yet resolved concerns the identification of what has been achieved. This is difficult enough at the moment, with a relatively trivial number of conversions.<sup>iv[4]</sup> Will national libraries be expected to be responsible for the national digital output?

4. Problems associated with procedure are, of course, capable of solution: what is not at all clear is where the money is going to be found to continue funding libraries and archives as we do in the year 2000 with the additional burden, for perhaps a century, of converting the world's research materials to digital form.<sup>v[5]</sup>



In the early days of the Internet – before the World Wide Web became universally adopted by academic institutions – the predominant users belonged to scientific and medical disciplines. By the early 1990s university libraries in the United States began making their catalogues available via Telnet, with a confusing variety of protocols. By 1994 access to bibliographical information seemed to be becoming more rather than less difficult to master: by 1996 some twenty different OPAC systems were in use. The Web,

as everyone knows, was developed by CERN for the scientific community as a means to enable international communication using a mark-up code which could be understood throughout the world. That code (SGML) gave way to a more sophisticated code (HTML) capable of handling complex graphics as soon as processor speed and chip design made 16- and then 32-bit processing available on desktop and laptop computers.<sup>vi[6]</sup> At that point in time – I would put it about mid-1995 – academics in the humanities and social sciences began to perceive the possibilities provided by the Web for engaging in what can only be described as large-scale publishing programmes. Conventional publishing methods could be by-passed and, providing one was prepared to master HTML, academics could provide for their students a wealth of source material not otherwise easily available. Indeed, the use of the Web by scholars in English Studies is quite remarkable, and some of the most imaginative sites are run by professors of English. Digital collections are, therefore, the result of two complementary objectives: to provide teaching materials, and to provide research materials. The problem is that these objectives make different demands on the provider. A collection of texts and images for the study of a poet (say Milton) might not find it necessary to provide variant texts of *Paradise Lost*, while a collection intended to support research into Milton's text would regard variants as essential. Indeed, there are those who believe that the creation of *variorum* editions is one application of digital technology with an undoubtedly beneficial future.



It was during the summer of 1995 that Michael Lesk, then Head of Computer Research at Bellcore in Morristown, New Jersey, made one of his regular visits to London, and the British Library. We had been in close touch with each other for several years, and often spent many hours trying to determine the future of the conventional library, with its demands on space and the constant need to develop preservation techniques to cope with the rapidly deteriorating stock of printed materials. The time, it seemed, was ripe for a bold initiative to develop what became known as the *digital library*. Lesk recalled that in a lecture I had given in January, 1991 at the Royal Institution I had even then envisioned the concept of the digital library.

**In its latest report, published this summer, the Commission on Preservation and Access addresses "the feasibility of a project to study the means, costs and benefits of converting large quantities of preserved library materials from microfilm to digital images". This report, by Donald Waters at Yale, follows on from an earlier report (July 1990) in which Michael Lesk observed that image digitisation would soon be both relatively cheap and available. Lesk and I have been talking about ways of digitising film since 1987, and I am pleased to report that it has taken only four years to get this idea accepted as a possibility. Since then, the Mekel 400M and 400F cameras for converting film and fiche to compressed digitised form have appeared and four 400M cameras are being used in the Virginia State Archives to convert existing negative stock as well as new film. In all, there are now fifty machines in operation in America, and each machine can process a 100 foot roll of film in one hour. There is not a single installation in Europe outside the world of commerce.**

**The point of departure, quite properly understood at Yale, is concern for access and two major investigations of the methods used by scholars in the humanities and sciences (the Research Library Group's Program for Research Information Management and the Faxon Institute's report presented at its 1991 conference at Reston, Virginia in April this year) confirm that scholars highly value and tend to favor information that is readily at hand and ... a critical measure of success for libraries charting a course into the future is how readily they**

steer information into the hands of their clients. ... The mission of the access-oriented library is to generate, preserve, and improve access to collections of recorded knowledge. This mission guides the fundamental relationship between the access services and the library collections.

When flatbed scanners were first introduced four years ago they aroused considerable interest in America, and they have been used to convert large quantities of unbound paper materials (they are currently being used in the White House). At the Microcomputer Symposium which I organised at the British Library in 1987 the capability of scanners to handle materials as disparate as maps, manuscripts and palm leaves was demonstrated. But they are quite impractical for use in most archives for two reasons: (1) they are unsuitable for bound volumes; and (2) they do not permit sophisticated techniques for image enhancement and optical filtration necessary with faded or discoloured documents or documents with annotations in pencil. They have the added disadvantage of failing to provide a backup in the event of some unforeseen electronic disaster. For these reasons it is now clear that digitisation is best achieved using microfilm as a first stage. With vast quantities of archive-standard microfilm to hand (presumably filmed because of its historical importance) it means that the transformation from analogue to digital format can proceed without subjecting fragile materials to further destructive handling. Such a cautious approach should appeal to a brotherhood distinguished for its conservatism in all things!

If I am right - more important if Yale University and the Commission on Preservation and Access are right - then it would seem that before very long we shall have whole libraries of historical source material available for consultation in digitized form. The interesting question we must ask is: on what sources will these libraries be based? The answer to that question will have less to do, I suspect, with what archivists imagine to be important and more to do with evolving trends in historical studies. In this respect archivists share a problem with librarians, for the electronic distribution of research materials is going to be as important in the future as the electronic distribution of catalogues and finding aids with which we are now concerned. On the other hand it is my impression that archivists, protected as they have been during the past ten years from the management theories of the Harvard Business School, are in closer touch with the scholarly community using their documents than are some librarians.<sup>vii[7]</sup>

Lesk agreed to write a monograph on the practicality of developing digital libraries, and within a year I had received a draft of what became *Practical Digital Libraries: Books, Bytes, & Bucks*.<sup>viii[8]</sup> Lesk's book is, as far as I am aware, the only serious study of the manifold issues raised by the concept of the digital library, though there are a number of notes of general guidance available on the Web – these are discussed in greater detail below. Encouraged by Lesk I persuaded the British Library to acquire a Melk 400 microfilm scanner, and began testing various types of negative microfilm from the vast stock in the collections. After about three months, just as I was about to report to the Director General on what I considered to be valuable experience and clearly demonstrating the potential of microfilm scanning for creating substantial digital collections of research material the project became absorbed in the Preservation Service. One product of the British Library's not very successful entry into the digital library arena was the publication in 1998 of *Towards the Digital Library: the British Library's Initiatives for Access Programme*.<sup>ix[9]</sup> This is really a collection of essays on various experimental sub-projects and contributions by various authors on subjects pertinent to the concept of a digital library. With the exception of the *Electronic Beowulf* project none of the initiatives for access have borne fruit of any consequence. The Digital Library Programme was put out to tender in 1998; by the middle of 1999 it was abandoned: neither the consortia interested nor the British Library could agree on objectives or terms and conditions. Remarkably enough Lesk's book is nowhere mentioned.

Quite apart from its tentative and ambiguous title, *Towards the Digital Library*, is a strange compilation, part concerned with document supply, part with the problems of transferring rare research materials into digital form. There can be no doubt that the latter is what will interest most scholars in this century; yet we seem no nearer to realizing the dream of Universal Availability of Publications (UAP) than we were all those years ago when it became a core activity of IFLA. Even today, as I shall show, much of what can be regarded as contributing to the concept of the digital library remains tentative, exploratory and limited. The reasons for this are not actually all that difficult to understand.

The first book which attempted to come to terms with digitization emerged from the 1996 celebration of the founding of the New York Public Library: *Books, Bricks & Bytes: Libraries in the Twenty-First Century*, edited by Stephen R. Graubard (Emeritus Professor of History, Brown University) and Paul LeClerc (President of the New York Public Library). This is a disappointing book since only the first essay “What is a Digital Library” by Peter Lyman addresses the issues implicit in the title. Most of the essays are concerned with traditional librarianship. It is not easy to grasp from most of the contributions what sorts of libraries we will have by middle of the twenty-first century.

For at least ten years libraries have been faced with the conflicting demands of those requiring remote access to research materials (monograph/serial, printed/manuscript) and the requirement to preserve and make accessible original materials acquired at vast expense, against a drift towards shrinking budgets. If a librarian found juggling resources difficult in 1980, when the first storm clouds began to appear, by 1990 not even those who approached budgeting with imagination could balance the books. The concept of the digital library was born in desperation. For Michael Lesk “The answer should not be despair but organization. A digital library, a collection of information that is both digitized and organized, gives us powers we never had with traditional libraries.” This, I believe, can be shown to be true, but only if the material digitized is processed with accuracy and fidelity to the original. These two conditions seem to me to be of primary importance if digital libraries are to reduce the increasingly costly and inconvenient business of travelling from Los Angeles to London to study the Chadwick Tracts.<sup>x[10]</sup> While Lesk’s book deals with most of the issues central to digitizing the world’s literature very little attention is paid to that aspect of scholarly use of sources which we call textual criticism. As anyone familiar with scholarly editing knows books and manuscripts are not simply carriers of information: they are, indeed, carriers, but no one copy of one edition of a historical text tells us the whole story. This is as true, as Adrian Johns has shown, for the history of science as it is for the history of a Dickens novel.<sup>xi[11]</sup>

That the technology exists to convert print of every kind and manuscript [text] to digital form is not in doubt. I have three concerns: (1) the choice of what [text] is converted; (2) the manner in which it [text] is converted; (3) the means by which it [text] is accessible. Each of these raises considerable associated problems.



During the heyday of the facsimile reprint [1960-75] publishers frequently reprinted texts known to be needed by research libraries from editions that had little or no textual authority, and did so for no other reason than that the correct edition was probably quite

rare and would have necessitated negotiating with a library which would want a share of the profit. Better, therefore, to acquire a copy of any edition available through the book trade; disbind it; photograph it; rebind it with some of the profit from the facsimile; re-sell the original through the trade. Such a procedure works quite well with a late edition of *Tom Jones*, but is pointless if what is needed is the first quarto edition of *Hamlet*. The world's research libraries have hundreds of expensive facsimile reprints of dubious pedigree: we should endeavour, with the new technology, not to repeat the mistakes of the past. In selecting what books to convert publishers must graduate from the mindset which reproduces every item in STC, or Wing, or ESTC, or NSTC. Sir Charles Chadwyck-Healey did departments of English a valuable service by reprinting all the English poetry in *the Cambridge Bibliography of English Literature*,<sup>xii[12]</sup> but many scholars have complained about the choice of editions used to create the database, especially for authors where the best editions are still in copyright but were not used for financial reasons. Libraries, too, since the Web is hospitable to entrepreneurial publishing, must graduate from the ready-made package based on particular collections of material: local history photographs and maps; pamphlets; ephemera; donated collections; &c.<sup>xiii[13]</sup> It is, I suppose, inevitable that libraries of almost any size will want to make available materials they believe to be particularly important for the remote user, but such sources are often difficult to find.<sup>xiv[14]</sup>

The method used to convert research materials is now being seen as crucial. The casual and uncritical methods adopted at an early stage by Project Gutenberg are clearly of little use for scholarly purposes: the electronic texts for the first few years of this extraordinary enterprise are textually inaccurate and frequently based on editions of doubtful value, yet Michael Hart (the creator of Project Gutenberg) has certainly created a digital library of sorts in the 1000+ texts available for FTP transfer at a number of sites maintained by American universities.<sup>xv[15]</sup> On the other hand, careful scrutiny of texts available from academic sites are not always dependable: of thirty electronic texts (mainly literary and historical) I checked against original texts in the British Library 75% were found to contain errors of substance. In 1992 I checked available electronic versions of Lewis Carroll's celebrated *Jabberwocky* poem, and found no single version that corresponded with the first edition: the situation is no better today!

That there are model websites with electronic texts as good as anything available in print is not in doubt, and I would cite the electronic projects at the University of Virginia (Rossetti, Whitman, *Piers Plowman*) as exemplary. These are, however, but a small percentage of the total number of purveyors of etexts, academic and commercial. Living writers are trying the Web, as did Stephen King in March when he launched his novelette *Riding the Bullet*. This was featured on several sites, including that maintained by Simon & Schuster (<http://www.simonsays.com>). Also advertised on the Simon & Schuster homepage are a variety of software add-ons to make reading ebooks more pleasant: Glassbook<sup>xvi[16]</sup>, NetLibrary<sup>xvii[17]</sup>, Rocket eBook<sup>xviii[18]</sup>, Peanut Press<sup>xix[19]</sup>, Softbook<sup>xx[20]</sup>, and Softlock<sup>xxi[21]</sup>. Microsoft is planning to launch its own software called ClearType in May for use with its own proprietary PDA.<sup>xxii[22]</sup> So far, e-books have been marketed for use with PCs and laptops, but the new generation of palm-type devices has attracted a significant amount of software development. Marketing fiction by distributing the first chapter with newspapers has now given way to providing free "Chapter One" texts electronically, a common feature on the websites maintained by the major American

publishers. Stephen King's success with *Riding the Bullet* might lead established writers to follow suit: why settle for a modest royalty, if you can grab all the revenue? Establishing a website and the mechanisms for collecting money and distributing a text via ftp is now both straightforward and inexpensive, though protecting a site from marauding hackers is still problematical. More serious is the problem associated with third party illegal trading, with thousands of websites situated in countries beyond the reach of copyright law. We do not yet have Internet Police, and there are few safeguards against unscrupulous exploitation of the Web by those who flourish in anarchic situations.

## Copyright

While most academic sites are conscientious about securing permission from copyright holders, I have encountered numerous instances of flagrant breaches for writers still in copyright (e.g. Yeats, Eliot, Spender). The law is clear and is available free at the U.S. Copyright Office at the Library of Congress: <http://lcweb.loc.gov/copyright/>. A detailed paper on "Copyright Law in the Electronic Environment" by Georgia Harper is at: [http://www.utsystem.edu/OGC/Intellectual\\_Property/faculty.htm](http://www.utsystem.edu/OGC/Intellectual_Property/faculty.htm). While *fair use* still operates in an electronic environment (as it has done for photocopying) establishing fairness can present some nice difficulties. Since it is now accepted practice for academics to put texts they require their students to read on university-maintained sites the danger exists that remote users (students in another university) will simply plunder such texts for their own purposes. Mechanisms do exist to restrict use of such texts to the host university for databases marketed on a licensed basis, but few academics like such restrictions, preferring to let their colleagues in other universities take note of their effort. Given the fact that university and college libraries are unwilling to allocate funds to providing multiple copies of works used in freshman and sophomore courses the electronic solution has seemed both practical and economic. Furthermore, teachers are no longer dependent on publishers of student editions for texts: a collection of texts for minor seventeenth century poetry can be created quite simply. Such collections form the majority of literary and historical texts available on the Web: I know of at least two hundred such, mostly at America universities and colleges.<sup>xxiii[23]</sup> For the most part it is obvious that they have been created in haste, and suffer inaccuracy: their architects would presumably argue that almost *any* text of a particular poem is preferable to a text only available in the main library in a single copy. This was what enabled the Chadwyck-Healey English Poetry Database to be marketed so effectively throughout the world: while the editions used may not always have been the ones a fastidious scholar would choose, the texts were acceptable for student use. Intellectually, such a philosophy is disreputable.

## Integrity

If the use of texts based on an edition without any authority is to be avoided, so must the integrity of what is made available electronically be zealously safeguarded. As anyone who has edited a text knows only too well, accuracy is difficult to achieve, but unless users can depend on the integrity of an electronic text then it is virtually useless. As Thomas Tanselle has reminded us:

**Microfilms and other reproductions can be helpful to scholarship if their proper use is recognized: but equating them with originals undermines scholarship by allowing precision to**



be replaced with approximation and secondary evidence to be confused with primary. The texts of many documents that once existed are now lost forever, and the texts of others are known only in copies. We use whatever there is; but when there are originals, we must not let substitutes supplant them as the best evidence we can have for recovering statements from the past.<sup>xxiv[24]</sup>

The quest for fidelity can yield interesting results: nowhere is this better seen than in the vastly expensive, yet exciting, forensic digital photography of the *Beowulf* manuscript in the British Library.<sup>xxv[25]</sup> What we have now is evidently an improvement on the original, for detail not visible to the eye has been captured, enabling doubtful readings to be correctly interpreted. However, at a cost of over £1000 per page this is unlikely to be seen as a methodology for building a digital research library! There are numerous sites at which students can consult pages from medieval and renaissance manuscripts: sometimes the reproductions are complete, more often they consist of selected pages.<sup>xxvi[26]</sup>

## Control

It is not easy to see how international control can be brought to bear on which texts are digitized, any more than it has proven impossible to control preservation microfilming. One has only to consult the National Register of Microfilm Masters to realize that different copies of the same text have been filmed by different institutions at different times. As far as I am aware, there exists no mechanism for ensuring that different copies of the same text are not digitized more than once. And it must not be forgotten that, although a MARC standard was developed for cataloguing microfilms, no such standard exists for cataloguing digital texts.<sup>xxvii[27]</sup> I list below several websites which attempt to bring some measure of control to the prevailing anarchy, but it seems not to have been appreciated that cataloguing what is available electronically is as important as carrying out the conversion to digital form.

For some sixty years now commercial microfilm companies and libraries have been filming rare books, manuscripts, newspapers for two principal reasons: to preserve the texts they contain and to make them more easily available to scholars and students. Many of these microfilms have turned out to be useless, because they have deteriorated and can no longer be read.<sup>xxviii[28]</sup> In all, the efforts of publishers and libraries have made available approximately 1% of the world's stock of printed research materials; materials in manuscript probably amount to less than 1%. Given that published output in the developed world is actually rising, it is difficult to see how it might be possible to alter this small proportion.

The concept of the digital library of the future remains a huge challenge, but before it is considered either as desirable, or practicable, we must decide what we propose to do with the world's existing stock of print. In spite of vast sums of money spent on devising techniques for the mass de-acidification of printed books, each year that passes sees the disappearance of books printed between 1820 and 1900. Stemming the seemingly inevitable deterioration in the world's book stock has given rise to vast sums of money being spent to microfilm collections at risk, notably by the Andrew Mellon Foundation, which has also supported JSTOR [Journal Storage]. The Chadwyck-Healey *Nineteenth Century Microfiche Project*, of which I am Editorial Director, has, in a decade, filmed over 12,000 nineteenth-century texts and provided them with full MARC records; but this represents a very small percentage of the texts at risk because of acidic paper.<sup>xxix[29]</sup> It

seems as if librarians must find truly staggering sums of money to pay for keeping their books and manuscripts as artifacts, and at the same time making them remotely available via the Internet. There seems to me no way in which these competing demands can be met without drastically altering library budgets.

Lesk's *Practical Digital Libraries* concludes that the dream of Vannevar Bush "is about to be realized":

**More and more, it will be the digital version that is used. Just as the reading of old newspapers moved from paper to microfilm, and music moved from performance to recordings, and the theatre from stage to cinema, we can expect a major shift toward digital reading. Just as in these other examples, the old will survive, but the new will be dominant.**

In December 1995 *Wired* magazine solicited the opinions of several American experts on "The Future of Libraries". The questions were: (1) by when will half of the Library of Congress collections be digitized? (2) when will we see the first "Virtual Large Library"? The answers were as follows: (1) some time between 2020 and 2065; (2) 2005-2030.<sup>xxx[30]</sup> Lesk is optimistic that

**Vannevar Bush's dream is going to be achieved, and in one lifetime. Seventy years from 1945, when he wrote his paper, it will be 2015, and it is clear that before then we will have the equivalent of a major research library on each desk. And it will have searching capabilities beyond those Bush imagined. ... We still lack a clear picture of how we're going to pay for all of this, but the explosion of the Web cannot be turned back. Whatever combination of greed and fear winds up supporting millions of Web sites, we will find some solution.**<sup>xxxi[31]</sup>

I find this depressing: not only will it spell doom for those who believe (as I do) that the world's literature still needs careful scrutiny, but it will widen the gap between the "have" nations and the "have nots". Only an engineer, it seems to me, could have reached such a conclusion; or, at least, someone for whom the distinction between what is information and what is knowledge is not understood.

Although no library can continue to seek universal coverage in its collection policy, Alexandria still sings her siren song. We have Bibliotheca Universalis; the Alexandria Digital Library; and the Bibliotheca Alexandrina itself. The dream is still there awaiting realization.

## APPENDIX

The following sites were visited for a final check during the last two weeks of September 2002. Brief mission statements are reproduced verbatim; others are given in summary form. Addresses are correct as of this going to press. Where provided the date of last update is provided. The following list does not cover resources for science, technology and medicine, and lists sources for subjects in the Humanities and Social Sciences, primarily in English. It does not attempt to cover sources for East Asia, for example, such as the valuable East Asian Library of Chinese, Japanese and Korean maintained by the Rutgers University Libraries:

<http://www.libraries.rutgers.edu/rulib/abtlib/easlib/ealib.htm>

or the Tamil Electronic Library

[www.geocities.com/Athens/5180/index.html](http://www.geocities.com/Athens/5180/index.html)

It excludes CD-ROM databases and resources.<sup>xxxii[32]</sup>

## Methodology

Useful guides available online are:

Peter Graham, Bibliography on Electronic Library/Digital Library Issues:

<http://aultnis.rutgers.edu/texts/ElectLibBib.html>;

Ben Gross, Digital Library Related Information and Resources:  
<http://interspace.grainger.uiuc.edu/~bgross/digital-libraries.html>;

Jann Lynn-George, Digitization: Technical Processes, Applications and Issues:  
[http://www.library.ualberta.ca/library\\_html/libraries/law/digit2.html](http://www.library.ualberta.ca/library_html/libraries/law/digit2.html).

For references to printed works on digitization see below: [http://www.lita.org/ital/1603\\_klemperer.htm](http://www.lita.org/ital/1603_klemperer.htm).  
A very useful bibliography - "Scholarly Electronic Publishing" - is maintained by Charles W. Bailey at the University of Houston Library: <http://info.lib.uh.edu/sepb/sepb.html>. A bibliography of works dealing with evaluation of web sources is at: <http://www2.widener.edu/>

The following are the only books on digital libraries of particular interest:

Lesk, M. *Practical Digital Libraries: Books, Bytes, and Bucks*. Morgan Kaufmann, 1997.

Leclerc, P. & S.R. Graubard. *Books, Bricks and Bytes: Libraries in the Twenty-First Century*. Transaction Publishers, 1997.

Lee, S.H. *Collection Development in a Digital Environment*. Haworth Press, 1999.

Janes, J. [et al]. *The Internet Public Library Handbook: a Guide for Building and Monitoring Virtual Libraries*. Neal-Schuman, 1999.

Stielow, F.J. *Creating a Virtual Library*. Neal-Schuman, 1999.

Hunter, G.S. *Preserving Digital Information*. Neal-Schuman, 1999.

Dewitt, D.L. *Going Digital: Strategies for Access, Preservation, and Conversion of Collections to a Digital Format*. Haworth Press, 1998.

Kessler, J. *Internet Digital Libraries*. Artech House, 1996.

Pastine, M. *Collection Development: Access in the Virtual Library*. Haworth Press, 1998.

Stern, D. *Digital Libraries: Philosophies, Technical Design Considerations*. Haworth Press, 1999.

Rothenberg, J. *Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation*. Council on Library & Information Resources, 1999.

### **Text Encoding Initiative**

"The Text Encoding Initiative (TEI) is an international project to develop guidelines for the preparation and interchange of electronic texts for scholarly research, and to satisfy a broad range of uses by the language industries more generally." Maintained by the University of Illinois at Chicago. The site has a list of projects using TEI, technical manuals and guidelines, tutorials, &c. There are links to French and Italian versions.

<http://www.uic.edu/orgs/tei/>

### **Memoriae Mundi Series Bohemica**

Part of UNESCO's Memory of the World Programme and Bibliotheca Universalis (see below). A valuable text is available on image compression.

<http://www.nkp.cz/altnkeng.htm>

<http://www.nkp.cz/start/knihcin/>

### **Memory of the World (UNESCO)**

A useful collection of references to books on digitization; unfortunately the list has not been updated since September 1997!

<http://www.unesco.org/webworld/memory/basictexts.htm>

## **Gateway Sites**

There are now literally hundreds of gateway sites, some of the best ones being those maintained by universities. For a survey of British sites note Brian Kelly in issue 22 of *Ariadne* published in December 1999 - available online at [www.ariadne.ac.uk/issue22/](http://www.ariadne.ac.uk/issue22/) A model academic gateway is that maintained at Oxford: <http://www.bodley.ox.ac.uk/olig/> University sites have been omitted from the following list, with the exception of that for English Studies maintained at Dusseldorf.

### **CyberDewey**

A novel approach to providing access to a multitude of Web resources, using the Dewey decimal scheme. It is an electronic version of David A. Mundie's book *CyberDewey: a Catalogue of the World Wide Web*. Pittsburgh, 1995.

<http://users.telarama.com/~mundie/CyberDewey/>

### **CyberStacks(sm)**

An alternative to CyberDewey CyberStacks “is a centralized, integrated, and unified collection of significant World Wide Web (WWW) and other Internet resources categorized using the Library of Congress classification scheme. ... All of the selected resources in CyberStacks(sm) are full-text, hypertext, or, hypermedia, and of a research or scholarly nature.”

<http://www.public.iastate.edu/~CYBERSTACKS/>

### **Adam**

A gateway resource for Art, Design, Architecture and Media “a searchable catalogue of 2546 Internet resources that have been carefully selected and catalogued by professional librarians for the benefit of the UK Higher Education community.”

<http://www.adam.ac.uk/>

### **Sosig**

Covers a wide spectrum: Business & Management, Economics, Education, Environmental Sciences, Ethnology, Geography, Government, Law, Philosophy, Politics, Psychology, Social Science, Social Welfare, Sociology, Statistics, Women’s Studies.

<http://www.sosig.ac.uk/>

### **Voice of the Shuttle**

Begun at the University of California at Santa Barbara by Alan Liu, who still administers the project. Became available on the Web in 1995. The site has links to just about every web resources a humanist would find interesting. There is an interesting essay by Liu entitled “Globalizing the Humanities – Voice of the Shuttle”, published in *Humanities Collections*, 1, no.1 (1998): 41-56.

<http://vos.ucsb.edu/shuttle/>

### **Humbul**

The HUMANities BULLETin Board was started ca.1985, and administered by the Office for Humanities Computing at Bath University. Funded by JISC “to redevelop Humbul as the Humanities Hub for the Resource Discovery Network.” Currently in process of being “transformed”. This is a veteran site for humanities studies, with links to various disciplines (anthropology, archaeology, classical studies, film/drama/media, history, humanities, hypermedia, linguistics, libraries, literature, medieval studies, music, philosophy, religious studies, visual arts), also dictionaries on-line and electronic resources. There are links to: ALEX – A catalogue of electronic texts [see below]; ATHENA – Electronic texts in various languages [see below]; BOOKS ONLINE [see below].

<http://users.ox.ac.uk/~humbul/>

### **Niss**

National Information Services and Systems. A comprehensive resource for U.K. higher education. Covers all subjects. Links world-wide.

<http://www.niss.ac.uk/>

### **UKOLN**

The UK Office for Library and Information Networking based at the University of Bath. Links world-wide.

<http://www.ukoln.ac.uk/>

### **Dusseldorfer Virtuelle Bibliothek**

A site for English Studies world-wide, with links to other sites in Germany: Erfurt Electronic Studies in English; Institut für Anglistik, Leipzig; Anglo-American Literature Guide, Göttingen. The site is maintained by Dr Annemarie Nilges of the University Library for the Anglistisches Institut der Heinrich-Heine-Universität Dusseldorf. The best site for English Studies on the Continent.

<http://www.uni-duesseldorf.de/WWW/ulb/ang.html>

## Digital Library Projects

There are numerous sites which encompass the concept of the digital library. A useful essay and links to the referenced sites is at: [www.lita.org/ital/1602\\_klemperer.htm](http://www.lita.org/ital/1602_klemperer.htm), compiled by Katharina Klemperer (a consultant) and Stephen Chapman (Harvard University). Other useful bibliographies can be found at:

<http://robotics.stanford.edu/users/ketchpel/annbib.html>

<http://aultnis.rutgers.edu/texts/ElectLibBib.html>

<http://alexandria.sdc.ucsb.edu/public-documents/bibliography>

<http://www.nlc-bnc.ca/ifla/II/diglib.htm>

<http://www.ifla.org/ifla/II/diglib.htm>

<http://www.acm.org/dl>

<http://scils.rutgers.edu/~woernerc/biblio.html>

<http://www.cs.wpi.edu/~webbib>

<http://sunsite.berkeley.edu/CurrentCites>

A critical list of books on cyberculture in its early years (1990-95) is by David Silver: <http://www.otal.umd.edu/~rccs/biblio.html>. A larger (annotated) 22 page list of links is maintained by digitalLibrary.net at: <http://digitallibrary.net/resources/>. Michael Lesk's various papers on digital libraries are at: <http://www.lesk.com/mlesk/diglib.html>.

The Digital Library Federation maintains a site that is part of the Council on Library and Information Resources, a pages providing "a working definition of digital library" and "Institutional goals for digital libraries" – [www.clir.org/diglib/](http://www.clir.org/diglib/) A substantial list of links to digital resources on the Web ("Digital Librarian") is maintained by Margaret Vail Anderson at <http://www.servtech.com/public/mvail/> The fifteen page listing includes many minor, though interesting digital projects currently underway in the U.S. A useful list of "Selected References on the Virtual Library" is maintained by the Special Libraries Association at <http://www.sla.org/> The most comprehensive bibliography is that maintained on the IFLA site at <http://www.ifla.org/II/diglib.htm>

### Iflanet – Digital Libraries

A site that has grown remarkably in the past two years and now provides a wide spectrum of information on all aspects of digital libraries, including links to active projects in Europe and the Americas.

<http://www.ifla.org/>

### D-Lib Forum.

A monthly electronic journal, based at the Corporation for National Research Initiatives and sponsored by DARPA on behalf of the Digital Libraries Initiative. Contains numerous important contributions to the theory and practice of digital libraries. "The emergence of the networked information system environment has allowed us to envision digital library systems that transcend the limits of individual collections to embrace collections and services that are independent of both location and format." D-Lib has published papers from the Xerox Palo Alto Research Center (PARC) on digital libraries, including "The Changing Social Roles of Documents" by Marti A. Hearst, May, 1996.

<http://www.dlib.org/>

### Bibliotheca Universalis

A partnership project of the national libraries of France, Japan, U.S., Canada, Italy, Germany, and U.K. (the founding partners); Belgium, Czech Republic, the Netherlands, Portugal, Spain, and Switzerland. "The main objective of Bibliotheca Universalis is to make the major works of the world's scientific and cultural heritage accessible to a vast public via multimedia technologies, hence fostering the exchange of knowledge and dialogue over national and international borders, The aim is to exploit existing digitization programmes in order to build up a large distributed virtual collection of knowledge and make it available via the global communication networks, enhancing the services to the end users." Started in 1995 at a meeting held in Paris. It remains to be seen whether this ambitious international initiative will get beyond

the planning stage: the list of digitization programmes is not very encouraging: active seem to be *Gallica* (France), *Memoria Hispanica* (Spain), and *American Memory* (U.S.).

<http://www.konbib.nl/gabriel/bibliotheca-uiversalis/bibuniv.htm>

### **British Library Digital Library Programme**

Started in 1993. Except for the Beowulf project – published on CD-ROM in March 2000 – most of the original twenty projects covered by the *Strategic Objectives* (published in 1993) have lapsed. Part of Bibliotheca Universalis (see above).

<http://www.bl.uk/diglib/>

<http://portico.bl.uk/services/ric/diglib/digilib.html>

### **American Memory**

Part of the Digital Libraries Initiative – Phase II, sponsored by the National Science Foundation and other U.S. agencies. “By stimulating a dialog among the tool builders, the content providers, and the users of digital materials, we hope to establish a tighter feedback loop for sharing findings associated with building, maintaining, using, and sustaining digital libraries.” “A digital library is more than a database, and the future National Digital Library will be much more than a universal union catalogue. We envision the National Digital Library as a set of distributed repositories of managed content and a set of interfaces (some of which will resemble traditional catalogs) to that content. The American Memory project has been long in the planning: it started in 1989. Its objectives are deeply rooted in primary and secondary education in the U.S. Many documents associated with this project are now available at:

<http://memory.loc.gov/ammem/ftpfiles.html>

<http://memory.loc.gov/ammem/dli2/index.html>

### **Arts and Humanities Data Service (AHDS)**

“Working on behalf of the academic community to collect, catalogue, manage, preserve and promote the re-use of scholarly digital resources.” Funded by the Joint Information Systems Committee (JISC) of the UK’s Higher Education Funding Councils and the Arts and Humanities Research Board. A national service covering archaeology, history, literary, linguistic and other textual studies, the visual arts, and the performing arts. “The Service Providers collect, preserve, catalogue, and distribute digital resources which are relevant to their academic disciplines, facilitate good practice in their creation and use, and offer some user services.

<http://ahds.ac.uk/bkgd/what.html>

### **Perseus Project**

“An evolving digital library.” An exemplary site, rich in style and substance – as it should be given the funding it receives from several national agencies and commercial organizations. Based in the Department of Classics at Tufts University. Perseus entered the electronic scene at an early stage (1987), publishing its first CD-ROM (Yale University Press) in 1992. The second CD-ROM appeared in 1996. The site currently covers all aspects of the ancient world with over 30,000 images freely available. Includes an online edition of the works of Christopher Marlowe and the 1623 Folio text of *Julius Caesar*. The current Editor-in-Chief is Professor Gregory Crane. There are mirror sites maintained at Somerville, Massachusetts, Berlin (<http://perseus/mpiwg-berlin.mpg.de>), and Oxford (<http://perseus/csad.ox.ac.uk/>).

<http://www.perseus.tufts.edu/>

### **Alexandria Digital Library**

Maintained by the University of California, Santa Barbara. Sponsored by the National Science Foundation. The site contains a useful bibliography of citations for digital libraries “neither definitive or comprehensive”. Not to be confused with the website for the Bibliotheca Alexandrina in Egypt!

<http://alexandria.sdc.ucsb.edu/>

### **Center for the Study of Digital Libraries**

Established at Texas A&M University in 1995, to build upon research developed in the Hypermedia Research Laboratory established in 1987. “Its mission is to foster pioneering research on the theory and



application of digital libraries and to create flexible and efficient new technologies for their use.” The Center organized the first two Conferences on the Theory and Practice of Digital Libraries (1994 and 1995) and the *Proceedings* are available online.

<http://www.csdl.tamu.edu/>

### **Illinois Digital Libraries Initiative**

Maintained by the University of Illinois at Urbana-Champaign. Sponsored by the National Science Foundation. For a detailed study of DLI projects 1994-1999 by Edward A. Fox see <http://www.asis.org/Bulletin/Oct-99/fox.html>. The National Science Foundation site is: <http://www.dli2.nsf.gov/>. “Our efforts were concentrated on building an experimental testbed with tens of thousands of full-text journal articles from physics, engineering, and computer science, and making these articles available over the World Wide Web, often before they were available in print. The DLI Testbed focused on using document structure to provide federated search across publisher collections. There are links to other DLI projects and some digital library sites. A useful report from the National Research Council is “Computing the Future.”

<http://surya.grainger.uiuc.edu/dli>

<http://dli.grainger.uiuc.edu/>

### **New York Public Library Digital Library Collection**

“The New York Public Library Digital Library Collections represent a growing body of primary source materials created for the Web. The Digital Library Collections provide the public with digital versions of books, manuscripts, photographs, engravings, and other items as well as tools to browse, search, and analyze these materials remotely via the Internet.” For printed books the *Schomburg African American Writers of the 19<sup>th</sup> Century* reproduces the text of some 52 works previously published as the *Schomburg Library* in 1988. For historical photographs there is the Dennis collection of Small-Town America, and *Images of African Americans from the 19<sup>th</sup> Century*. A useful text is “Planning Digital Projects for Historical Collections”, available online.

<http://digital.nypl.org/>

### **Stanford Digital Library**

Maintained by Stanford University. Sponsored by the National Science Foundation. “The goal of this Project is [to?] design and implement the infrastructure and services needed for collaboratively creating, disseminating, sharing and managing information in a digital library context.” Started in July 1999, so it is still early days. Seen as part of the California Digital Library – currently being funded by the Universities of California at Berkeley and Santa Barbara, Stanford, and the San Diego Supercomputer Center. “All three projects together yield a synergistic and comprehensive digital libraries project.” “The Stanford component of this effort will develop the base technologies that are required to overcome the most critical barriers to effective digital libraries. One of these barriers is the heterogeneity of information and services. Another impediment is the lack of powerful filtering mechanisms that let users find truly valuable information. The continuous access to information is restricted by the unavailability of library interfaces and tools that effectively operate on portable devices. A fourth barrier is the lack of a solid economic infrastructure that encourages providers to make information available, and give users privacy guarantees.” It is obvious that not all these issues can be resolved by technology.

<http://diglib.stanford.edu/>

### **Berkeley Digital Library – Sunsite Digital Collections**

Maintained by the University of California at Berkeley. Sponsored by the National Science Foundation. A primary site for the study of digital libraries with links to digital text collections, image collections.

<http://sunsite.berkeley.edu/>

<http://elib.cs.berkeley.edu/>

### **Michigan Digital Library**

Maintained by the School of Information and Library Studies, the University of Michigan. Sponsored by the National Science Foundation.

<http://http2.sils.umich.edu/UMDL/HomePage.html>

## Indexes & Directories

### The On-Line Books Page

Established in 1993 and now perhaps the most comprehensive index (with links) to electronic books on the Web. Started by John M. Ockerbloom at Carnegie Mellon University, since 1999 at the University of Pennsylvania. Lists 11,000+ texts, searchable by author/title/subject. Copyright is claimed.

<http://digital.library.upenn.edu/>

### Literary Resources on the Net

Links to all the major sites "dealing especially with English and American literature". Maintained by Jack Lynch.

<http://andromeda.rutgers.edu/~jlynch/>

### The Universal Library

Hosted by Carnegie Mellon University. Extensive links to other sites. "Vision for the Universal Library – Access to all human knowledge – anytime anywhere."

<http://www.ul.cs.cmu.edu/>

### CETH Directory of Electronic Text Centers

Compiled by Mary Mallery. Includes links to 19 U.S. Electronic Text Centers, as well as some Digital Library Projects. Not very current!

<http://scc01.rutgers.edu/ceth/infosrv/ectdir.html>.

### Dial-A-Book

Provides links to Chapter One files in various U.S. newspapers of new books.

<http://www.dialabook.com/>

### Columbia Index to Electronic Texts on the Internet

A list of major online text collections with links. Includes Manuscript Studies, Medieval Studies, Philosophy, Religion, Women and Gender Studies.

<http://www.columbia.edu/cu/libraries/indiv/ets/offsite.subject.html>

### Dscriptorium

"Dscriptorium is devoted to collecting, storing and distributing digital images of medieval manuscripts". Compiled by Jesse D. Hurlbut, Brigham Young University. Numerous links. Not very current. A casualty?

<http://www.byu.edu/~hurlbut/dscriptorium/>

### ALEX Catalogue of Electronic Texts

A "collection of digital documents collected in the subject areas of English literature, American literature, and Western philosophy." Started as a gopher service in July 1994 by Hunter Monroe at Radcliffe Science Library in Oxford (<gopher://rsl.ox.ac.uk:70/11/lib-corn/hunter>). The old list, known as "Alex Classic" is still available. Also links to electronic text collections. Compiled by Eric L. Morgan. "The Catalogue's primary purpose is to assist me in demonstrating a concept I call arscience, a process of understanding using methods from art and science. The Catalogue's secondary purpose is to provide value-added access to some of the world's great literature in turn providing the means for enhancing education." Searchable by author and title. Currently lists 133 authors.

<http://sunsite.berkeley.edu/alex/>

### Argus Clearinghouse

Includes guides to Arts & Humanities, Business & Employment, Communication, Computers & Information Technology, Education, Engineering, Environment, Government & Law, Health & Medicine, Places & Peoples, Recreation, Science & Mathematics, Social Sciences & Social Issues. The section on Arts & Humanities is wide-ranging. Links to a large number of sites.



<http://www.clearinghouse.net/>

## **Victoria Research Web: 19<sup>th</sup>-Century Britain**

Compiled by Patrick Leary at Indiana University. A very good site, though not as comprehensive as George Landow's Victorian Web (see below).

<http://www.indiana.edu/~victoria>

## **Model Editions Partnership**

"The purpose of the Model Editions Partnership is to explore ways of creating editions of historical documents which meet the standards scholars traditionally use in preparing printed editions. Equally important is our goal of making these materials more widely available via the Web. Currently provides exemplary editions of Lincoln's legal papers, General Nathanael Greene, Margaret Sanger, Elizabeth Cady Stanton, Susan B. Anthony, Henry Laurens, and the First Federal Congress. Maintained at the University of South Carolina. Supported by USC, the University of Illinois at Chicago, and the University of Alberta.

<http://mep.cla.sc.edu/>

## **Digital Scriptorium**

"A Prototype Image Database & Visual Union Catalog of Medieval and Renaissance Manuscripts." "The Digital Scriptorium was conceived as an image database of dated and datable medieval and renaissance manuscripts, intended to unite scattered resources into an international tool for teaching and scholarly research. It has evolved into a general union catalog designed for the use of paleographers, codicologists, art historians, textual scholars to verify with their own eyes cataloguing information about places and dates of origin, scripts, artistic styles, and quality." Started in 1996 at the Bancroft Library (Berkeley) with funding from the Andrew W. Mellon Foundation. Currently holds some 8,500 colour images based on the collections at Berkeley, Columbia, the Union Theological Seminary and the De Bellis Collection at San Francisco State University. Funded by the N.E.H. in 1999 to add materials from the Huntington Library, the University of Texas, and the New York Public Library. An ambitious project which has the promise of being a very valuable site.

<http://sunsite.berkeley.edu/Scriptorium/>

## **The Online Medieval & Classical Library**

Created by Douglas B. Killings. Links to several other websites concerned with medieval/classical studies. Provided with a good search engine.

<http://sunsite.berkeley.edu/OMACL/>

## **Luminarium**

Renaissance English literature. Compiled by Anniina Jokinen, University of California at Los Angeles.

<http://www.luminarium.org/renlit/>

## **Exemplary Technologies**

Electronic books for Windows. Some erotica (Sir Richard Burton's translations, *Fanny Hill*, &c.). A commercial enterprise.

<http://www.exemplary.net/publishing/index.html>

## **Electronic Library Resources**

"A subject guide to selected resources on the Internet." Maintained by Nottingham Trent University Department of Library & Information Services. Resources include: Art & Design, Built Environment, Business & Management, Computing, Education, Engineering, Health & Human Services, Humanities, Law, Science & Mathematics, Social Sciences, Sport, Statistical Information, U.K. Official Information. Includes large collection of on- and off-line resources, mainly European in origin.

<http://www.ntu.ac.uk/lis/elr.htm>

## Center for Electronic Texts in the Humanities

Maintained at Rutgers University. Includes a Directory of Electronic Text Centers compiled by Mary Mallery, and the six Digital Library Projects sponsored by the National Science foundation: Santa Barbara, Carnegie Mellon, Illinois, Stanford, Berkeley, and Michigan.

<http://scc01.rutgers.edu/ceth/>

## Clearinghouse of Image Databases

Alphabetical list of image databases maintained by the Library at Arizona University. Regularly updated. Though most of the sites indexed are U.S. there are several from the U.K., Australia, New Zealand.

<http://dizzy.library.arizona.edu/cgi-bin/clearinghouse-to-html.pl?/>

## Inventory of Canadian Digital Initiatives

Maintained by the National Library of Canada. Limited in scope at present. Dependent on libraries/vendors notifying the NLC.

<http://www.nlc-bnc.ca/initiatives/erella.htm>

## Universal Availability of Publications Core Programme (IFLA)

Directory of digitised collections. The aim is “to compile a comprehensive listing of digitised documents held by libraries worldwide.” Begun in 1998.

<http://www.ifla.org/VI/2/p1/desc.htm>

## Lorre Smith

A personal site maintained by a librarian at Albany University. Basically, a list of Berkeley resources, but there is also a substantial list of “Digital Library Resources”, as well as web links to electronic text centers, bibliographies, journals, conferences, metadata, &c.

<http://www.albany.edu/~ls973/>

## Electronic Text Centres

Electronic Text Centres are facilities provided by a number of universities in the U.S. intended primarily as teaching and research resources, with access to electronic literature both on-line and on CD-ROM. I have listed here only those that are sufficiently ambitious to have attracted international interest. All such facilities provide extensive links to pertinent sites and access to electronic books.

## CTI Centre for Textual Studies

“The Computers in Teaching Initiative (CTI) was funded from 1989-1999 by the UK higher education funding bodies to promote and support the use of computers in university teaching. The CTI comprised twenty-four subject centres based within universities within the UK. Arts and Humanities subjects were supported by five centres: music at Lancaster, history and archaeology at Glasgow, modern languages at Hull, art and design at Brighton, and textual studies at Oxford. The CTI Centre for Textual Studies had a wide remit, providing support for any discipline within the humanities in which the study of texts formed a significant activity.” The printed *Guide to Digital Resources for the Humanities* is now in its third edition [1999] and lists over 360 resources and a set of introductory essays; the bibliography lists 370 articles, selected for their importance in humanities computing. CTI Textual Studies will close in 2000 to be replaced by another network of subject centres: the *Learning and Teaching Support Network* (LTSN). The Oxford Text Archive remains one of the pioneer enterprises in humanities computing, founded by Lou Barnard in 1976, and its recent on-line catalogue (26 pages – February 2000) lists hundreds of texts freely available in electronic form. The quality of these has always been of the highest order. It does not produce digital resources, but relies on deposits from the wider academic community as the primary source of high-quality materials.

<http://info.ox.ac.uk/ctitext/>

[http://ota.ahds.ac.uk/main\\_ie4.html](http://ota.ahds.ac.uk/main_ie4.html)

## **Virginia Electronic Text Center**

Maintained at the University of Virginia. A veteran site, started in 1992, with “approximately 45,000 on- and off-line humanities texts in twelve languages, with more than 50,000 related images.” The Research Sources section includes journals, bibliographies and reference works. Many commercial databases and collections are licensed and therefore only available to members of the Virtual Library of Virginia (VIVA). The languages included are: English, French, German, Latin, Chinese, Japanese, Russian, Greek, Hebrew, Spanish, Tibetan, Icelandic, Italian, Portuguese.

<http://www.lib.virginia.edu/uvaonline.html>

<http://etext.lib.virginia.edu/>

## **Michigan Electronic Library**

Sponsored by the Michigan State Library and the University of Michigan.

<http://mel.org/>

## **Electronic Text Center.**

Beck Center for Electronic Collections and Services, Emory University.

<http://chaucer.library.emory.edu/>

## **Center for Text and Technology, Georgetown University**

Largely the creation of Michael Neuman, and one of the first such facilities to be created. Includes electronic standard critical editions in philosophy, including Hegel and Feuerbach, as well as specialist resources for Medieval Studies, American Studies, Politics in the Americas. The Catalogue of Projects in Electronic Text will shortly be available on this site.

<http://www.georgetown.edu>

## **Scholarly Digital Resource Center, University of Iowa**

Maintained by the University Library, “established to foster the creation and use of digitized collections and resources of interest to the University of Iowa community ... and many of its projects are in partnership with faculty from a wide range of disciplines.” In fact a rather thin site as yet – the Arts and Humanities Digital Resources Center has very little to interest anyone apart from those interested in Dada or Chautauqua. There are a large number of comparable sites making extravagant claims on their home pages.

<http://sdrc.lib.uiowa.edu/>

## **Humanities Text Initiative, University of Michigan**

Started in 1994 by the University of Michigan Press. Includes the Corpus of Middle English Prose and Verse, the Middle English Dictionary, and the American Verse Project. Includes the Hyperbibliography of Middle English. “The Hyperbibliography includes all the Middle English materials which are cited in the Middle English Dictionary. Although this is not an exhaustive bibliography of Middle English writing it offers the most comprehensive single list of ME materials at present available in electronic form, and is searchable in multiple ways.” The search engine is powerful and enables searching by dialect as well as author and bibliographical reference.

<http://www.hti.umich.edu/>

## **Irish Literary Sources and Resources**

An archive of Irish texts maintained by Michael Sundermeier at Creighton University. The text of Maria Edgeworth’s *Castle Rackrent* (1800) was checked against the three editions of 1800 and found to be very accurate. “Electronic Editions 2000” contains works by John Banim, William Carleton, Gerald Griffin, James Joyce, J.M. Synge, and Emily Lawless

<http://mockingbird.creighton.edu/english/micsun/index.htm>

## **CELT – Irish Electronic Texts**

Maintained at University College Cork. “The CELT project aims to produce an online database of contemporary and historical topics from many areas, including literature and the other arts. ... Texts are taken from the best printed editions, scanned, and proofread. Markup for structural and analytic features is added according to the recommendations of the Text Encoding Initiative (TEI).” The aim is to “bring the wealth of Irish literary and historical culture (in Irish, Latin, Old Norse, Anglo-Norman French, and English) to the Internet in a rigorously scholarly project that is, at the same time, user-friendly for the widest possible range of readers and researchers.” The list of texts extends to 30 pages, with options to read, SGML, download, plain, and print.

<http://www.ucc.ie/celt/about.html>

## **Irish Resources in the Humanities**

A list of sources covering archaeology, architecture, archives & Libraries, bibliographies, geography, history, Irish language, literature, newspapers, music, and theatre. An ambitious undertaking, maintained by Susan Schreibman, New Jersey Institute of Technology for University College, Dublin.

<http://www.ucd.ie/irh/>

## **Documenting the American South**

Maintained at the University of North Carolina, Chapel Hill. “Provides access to digitized primary materials that offer Southern perspectives on American history and culture.” Includes slave narratives, first-person narratives, Southern literature, Confederate imprints, and “materials related to the church in the black community. Another project featuring North Caroliniana is in development.”

<http://metalab.unc.edu/docsouth/aboutdas.html>

## **Poets’ Corner**

“Our goal is to create the largest, most diverse, and most user-friendly public library of poetic works ever assembled. The materials on display are selected from an inventory of thousands of works by hundreds of authors, transcribed and gathered here by the Editors and by many volunteer contributors from around the world.” Currently 5715 works by 656 poets. Maintained by Bob Blair (Texas), Jon Lachelt (Colorado), Nelson Miller (Georgia), and Steve Spanoudis (Florida). Does not include works in copyright in the U.S.

<http://www.geocities.com/~spanoudi/poems/index.html>

## **EMule.com**

A poetry archive, currently with 3636 poems by 137 poets. A commercial venture.

<http://www.emule.com/poetry/>

## **Individual Authors**

The following are just a few of the large number of digital resources for the study of individual authors. Included are a few of the very best, a few rather less ambitious, and some that illustrate the way in which virtually every library and archive in the U.S. is beginning to exploit their collections on the Web. The problem, as one might expect, is knowing how to find that specialized resource dealing with a particular author.

### **The Complete Writings and Pictures of Dante Gabriel Rossetti: a Hypermedia Research Archive**

Created by Jerome J. McGann. A model of its kind. The archive has been under construction since 1993. Of considerable importance is McGann’s often-cited essay, “The rationale of Hypertext” (1995). It is planned to issue the archive on CD-ROM by the University of Michigan Press some time in 2000.

<http://jefferson.village.virginia.edu/rossetti/tour/index.html>

<http://jefferson.village.virginia.edu/public/jjm2f/rationale.html>

## **The William Blake Archive**

“A hypermedia archive sponsored by the Library of Congress and supported by the Getty Grant Program, the Institute for Advanced Technology in the Humanities at the University of Virginia, Sun Microsystems, and Inso Corporation.” An exemplary site administered by Morris Eaves (University of Rochester), Robert Essick (University of California, Riverside), and Joseph Viscomi (University of North Carolina). As of April 2000 “the Archive now contains fully searchable and scalable electronic editions of 41 copies of 18 of Blake’s 19 illuminated books in the context of full bibliographic information about each work, careful diplomatic transcriptions of all texts, detailed descriptions of all images, and extensive bibliographies.”

<http://jefferson.village.virginia.edu/blake/>

## **The Milton Reading Room**

Started in 1997 by Thomas Luxon at Dartmouth College and his students. Now contains the electronic text of most of Milton’s poetry and some of his prose. “Milton, I believe, should be read and studied from inside a library, and the web makes this possible on a scale only recently imagined. I hope that, in the future, the Milton Reading Room will be redesigned to offer an elaborate search mechanism, access to early manuscript and print versions of the poems and prose, and a virtual forum for discussion and study of Milton’s complete works, including those in ancient and modern languages other than English.”

[http://www.dartmouth.edu/~milton/reading\\_room/contents/](http://www.dartmouth.edu/~milton/reading_room/contents/)

## **Canterbury Tales Project**

The project, based at De Montfort University, is the brainchild of Peter Robinson. It aims to “establish a system of transcription for all the manuscripts and early printed books of the *Canterbury Tales* into computer-readable form. Transcribe the manuscripts using this system. Compare all the manuscripts, creating a record of their agreements and disagreements with a computer collation program. Use computer-based methods to help reconstruct the history of the text from this record of agreements and disagreements. Publish all the materials, the results of our analysis, and the tools which we use in electronic form.” The tangible fruit of this project, begun some years ago when Robinson was at Oxford are CD-ROMs of individual *Tales*. A full description of the project is available on the website.

<http://www.cta.dmu.ac.uk/projects/ctp/>

## **Aberdeen Bestiary Project**

An electronic version of Aberdeen University Library MS. 24, an illuminated bestiary dating from about 1200. The electronic facsimile is accompanied by commentaries, a transcription and translation of the original Latin.

[http://www.clues.abdn.ac.uk/besttest/alt/comment/best\\_toc.html](http://www.clues.abdn.ac.uk/besttest/alt/comment/best_toc.html)

## **Electronic Beowulf**

MS. Cotton Vitellius A. xv. Published by the British Library and the University of Michigan Press on two CD-ROMs. Includes electronic facsimiles of the two Thorkelin transcripts (A+B), the transcripts of Conybeare and Madden, the Southwick Codex and the Nowell Codex. The project is jointly sponsored by the British Library and the University of Kentucky, and edited by Kevin Kiernan and Andrew Prescott. Published by the British Library and the University of Michigan Press on two CDs (February, 2000).

<http://www.uky.edu/%Ekiernan/eBeowulf/content.htm>

<http://www.uky.edu/ArtsSciences/English/Beowulf/eBeowulf/guide.htm>

## **The Walt Whitman Hypertext Archive**

Edited by Kenneth M. Price and Ed Folsom at the University of Virginia. “The Walt Whitman Hypertext Archive is a hypermedia environment for studying the works of the nineteenth-century American poet Walt Whitman. The archive is a structured database holding digitized images of Whitman’s works in their original documentary forms.” Exemplary site.

<http://jefferson.village.virginia.edu/whitman/>

## **WOMDA – The Wilfred Owen Multimedia Digital Archive**

An Oxford site modeled on similar archives at the University of Virginia. Exemplary. Prepared by Paul Groves.

<http://www.hcu.ox.ac.uk/jtap/>

## **Project Whistlestop – Truman Presidential Library Digital Archives**

Includes original documents for the following political events in Truman's period of office: "The Decision to Drop the Atomic Bomb", "The Recognition of Israel", "The Marshall Plan", "Desegregation of the Armed Forces", "The Korean War", "Berlin Airlift", "North Atlantic Treaty Organization". Within the constraints of national security one would hope for similar sites based on Presidential libraries!

<http://www.whistlestop.org/>

## **The Calvin Shedd Papers**

"The letters of Calvin Shedd, edited and reproduced here for the first time, tell a story of personal integrity and sacrifice in the words of a simple man who lived in a turbulent, complicated time. The Shedd letters add another fascinating source to our national reservoir of primary source materials relating to the Civil War." Reproduced from originals in the University of Miami and Dartmouth College. *Dartmouth College Library Bulletin*, xxxviii, no. 1, Nov. 1997, 28-45.

<http://www.library.miami.edu/archives/shedd/>

## **Electronic Books**

### **The Electronic Literature Foundation**

"The mission ... is to produce advanced electronic texts to be used by students, scholars, and admirers of literature around the world. Our goal is to provide free access to a variety of texts from world literature available in several languages and/or editions, with forums for communication regarding these works, for all types of readers." Includes works by Dante, Jane Austen, the Brontes, Chaucer, Hardy, Milton, Poe, Shakespeare, Mary Shelley, Verne, and Voltaire.

<http://elf.chaoscape.com/>

### **Victorian Women Writers Project**

A site of primary importance for the subject. New texts regularly added. "The goal of the Victorian Women Writers Project is to produce highly accurate transcriptions of works by British women writers of the 19<sup>th</sup> century, encoded using the Standard Generalized Markup Language (SGML)." Includes a list of works in preparation. Edited by Perry Willett.

<http://www.indiana.edu/~letrs/vwwp/>

### **The English Server**

A comprehensive site for students and researchers in English literature. Started in 1991 in the English Department at Carnegie Mellon University. The section devoted to Eighteenth Century Studies is devoted almost exclusively to 18<sup>th</sup> century texts in all genres. An extensive list of titles. Maintained at Carnegie Mellon University.

<http://englis-server.hss.cmu.edu/18th/>

### **Bibliomania**

Electronic texts by English and American writers.

<http://www.bibliomania.com/>

### **Alive & Free**

An irreverent site with excerpts from modern American writers, as well as complete texts. The links to other sites badly needs revision as they are mostly wrong! Presumably a casualty?

<http://www.c3f.com/alivfree.html>

## **Nerdworld**

An offbeat site with extensive links (over 315,000) to many subjects, including Literature.

<http://www.nerdworld.com/>

## **The Gilded Page**

“A collection of electronic texts written by US authors or widely read by Americans in the Gilded Age (loosely defined here as 1866-1901). I assign these as primary sources for my William and Mary students doing projects in my postbellum US history courses.” Compiled by Scott Nelson.

<http://www.wm.edu/~srnels/giltext.html>

## **Secular Web Library – Internet Infidels**

Mostly concerned with “freethought” literature. Published by the Internet Infidels, “an educational nonprofit organization of unpaid volunteers dedicated to the growth and maintenance of the most comprehensive freethought web site on the Internet. Our mission is to defend and promote metaphysical naturalism, the view that our natural world is all that there is, a closed system in no need of an explanation and sufficient unto itself. ... With over 6,000 documents and more than 190,000 unique visitors per month, the Secular Web is the largest and most heavily visited nontheistic web site on the Internet.”

<http://www.infidels.org/library/>

## **Making of America**

“A digital library of primary sources in American social history from the antebellum period through reconstruction. The collection is particularly strong in the subject areas of education, psychology, American history, sociology, religion, and science and technology. The collection currently contains approximately 1,600 books and 50,000 journal articles with 19<sup>th</sup> century imprints.” Maintained at the University of Michigan. The project plans to add 7,500 volumes in the next two years. Supported by the Andrew W. Mellon Foundation.

<http://moa.umdl.umich.edu/>

## **The Internet Classics Archive**

A project based at MIT. Currently some 400 works of classical literature by 59 authors. All texts in English. Compiled by Daniel C. Stevenson. Claims copyright.

<http://classics.mit.edu/>

## **Historical Text Archive**

A project which had its roots in pre-Web days when academic exchange was usually by ftp and gopher. The brainchild of Don Mabry. A site rich in resources for the study of history.

<http://www.geocities.com/Athens/Forum/9061/>

## **Representative Poetry Online**

Published by the University of Toronto Library. Electronic versions of an anthology of English poetry first published by the University of Toronto Press in various editions between 1912 and 1967. The first electronic version was available in 1994. The current version includes 2,100 poems by 331 poets from Caedmon to the early twentieth century. An exemplary site. Edited by Professor Ian Lancashire.

<http://www.library.utoronto.ca/utel/rp/>

## **Early Canadiana Online**

“A project to provide enhanced access to Canada’s published heritage.” “Early Canadiana Online (ECO) is a full text online collection of more than 3,000 books and pamphlets documenting Canadian history from the first European contact to the late 19<sup>th</sup> century.” Funded by numerous public and private organisations and agencies. Jointly sponsored by the Canadian Institute for Historical Microreproductions, the National Library of Canada, the Universite Laval Library, the University of Toronto Library, and the Bibliotheque nationale du Quebec. Directed by Karen Turko.



<http://www.canadiana.org/>

### **Christian Classics Ethereal Library**

“Classic Christian books in electronic format.” With an online author index, and a search engine. A CD-ROM was published in February, 2000.

<http://www.ccel.org/>

### **The Internet Public Library**

An ambitious site maintained by the School of Information at the University of Michigan. Sponsored by Bell & Howell. “The Internet is a mess. Since nobody runs it, that’s no surprise. There are a lot of interesting, worthwhile and valuable things out there – and a lot that are a complete waste of time. ... The Internet Public Library is the first public library of the Internet.” This claim is, in fact, a dubious one as the IPL makes available a great deal of what is available elsewhere. The published date of last update is not very encouraging! Contains a useful set of links to other sites.

<http://www.ipl.org/>

### **Bartleby.com**

“Great Books Online.” Begun in 1993 by Steven H. van Leeuwen with an electronic version of Whitman’s *Leaves of Grass*. Reference works and major English & American authors, including several still in copyright (Eliot, Yeats). It is not entirely clear whether the electronic versions of authors still in copyright have the permission of the copyright holders. The texts are, for the most part, accurate and the choice of editions appropriate. *The Waste Land*, for example, is from the first Boni & Livewright edition of 1922. Bartleby bibliographical records for the texts published have ISBNs and copyright is claimed on the electronic edition, which I find bizarre. There is a six page user’s agreement.

<http://www.bartleby.com/>

### **Books and Stories Online**

Links to other electronic sites, with a few texts found only on the Web.

<http://www.geocities.com/SoHo/7786/books.htm>

### **The Etext Archives**

“Home to electronic texts of all kinds, from the sacred to the profane, from the political to the personal.” Started in 1992 by Paul Southworth at the University of Michigan (not in any way the project’s sponsor). Dedicated originally to archiving e-zines, the site now has sections devoted to Politics, Fiction, Religion, and Poetry. Currently stores over 3 GB of text.

<http://www.etetx.org/>

### **netLibrary**

“The next generation library. NetLibrary offers the world’s largest library of eBooks. Our “electronic books” give you the power to read and research. From anywhere. At any time of the day.” *New York Times*, January 18, 2000: “netLibrary will enter the market for supplying digitized books to public libraries with a program aimed at allowing the libraries to sample e-books for six months without charge.” *Library Journal*, January 17, 2000: “netLibrary donating 150,000 eBooks to 100 top Public Libraries.” Browsable index of eBooks by category. Language & Linguistics: 157 titles; American Literature: 333 titles; Classical Literature: 39 titles; European Literature: 234 titles.

<http://www.netlibrary.com/>

### **Ebooks on the Net**

Fiction; Children; Humor; Non fiction; Poetry; Mystery; Plays; Horror/Sci-Fi; Medical/Psychology. Based in Bar Harbor, Maine: “Dynamically presenting Acadia and Maine to the Entire World.”

<http://www.ebooksonthe.net/>



## **Scribblers eBookshoppe**

“As the site grows in leaps and bounds, it’s impossible to put all what’s happening on the main page ...”  
Sells ebooks from 25 electronic publishers.

<http://www.ebookshoppe.com/>

## **TeleRead: Bring the E-Books Home**

Founded by David Rothman “national coordinator of TeleRead and a long-time advocate of electronic books and a well-stocked national digital library.” “Teleread is a nonpartisan plan to get electronic books into American homes – through a national digital library and small sharp-screened computers – in an era of declining literacy. With links to other sites with comparable aims, including Project Gutenberg, Project Bartleby, American Literary Classics, The On-Line Books Page, The Electric Book, The English Server, Online Originals, Boson Books, ReadToMe, &c.

<http://www.teleread.org/>

## **Napoleonic Literature**

A site devoted to Napoleonic studies.

<http://napoleonic-literature.simplenet.com/>

## **Michael Gamer**

A Professor in the English Department at the University of Pennsylvania whose personal website contains, amongst other electronic goodies, links to other literature sites and extensive links to texts for the Romantic writers.

<http://www.english.upenn.edu/~mgamer/>

## **Publishers of Electronic Books**

While most major publishers that have e-texts on their lists provide for their purchase online, the number of small enterprises (almost the electronic counterpart to the private press) is growing weekly. Every conceivable patronage is catered for, including pornography, fiction, romance, as well as non-fiction for most hobbies and activities. I list just a representative sample here.

### **Antique Book Library**

A small player so far with a limited list. “Instead of just uploading ASCII text, we’ve given you full color scans of these books, so that you can enjoy the weathered character that makes them so fascinating.” The site is administered by Antique Books Inc. in Pittsburgh.

<http://www.antiquebooks.net/readpage.html>

### **Etext – College Electronic Textbook Publishing**

A division of Mozena Multimedia Publishing. “Pay for information, not paper and binding!” Currently 188 “Eclassics”, mostly derived from Project Gutenberg and MIT. “Why pay for pulic [!] domain classic [!] when you can download them for free from the internet?”

<http://www.ertext.net/cgi-bin/Eclassics.cgi>

### **Petals of Life**

“Where all things are possible!” Started by Michele J. Johnson, a spiritual healer. There are free and priced ebooks on this site, fiction and nonfiction.

<http://www.petalsoflife.com/>

### **Simon & Schuster**

The Simonsays site gives links to various vendors of ebooks and ebook readers.

<http://www.simonsays.com/ebooks>

## **EBooknet**

Launched the Rocket-Library on February 4, 2000. Compiled by volunteers, with 3,000 titles on offer – “the second largest repository of free eBooks on the Web”. The site features online discussion forums of current issues in electronic publishing.

<http://www.ebooknet.com/>

## **Immortal Publications**

Almost entirely fiction written for online sale. Claims copyright.

<http://www.immortalpublications.com/>

## **Electronic Poetry Center**

Kinetic and hypertext poetry. Quite substantial list. Maintained by SUNY Buffalo.

<http://wings.buffalo.edu/epc/>

## **Ongoing Tales**

“Out of copyright books beautifully enhanced with electronic [!] media, these ebooks include both classic works and little known writings from the past.”

<http://www.teleport.com/~writers/SERIALS/oldtime/mybooks.html>

## **Tartan Hen**

Publishes fiction in conventional format. Specimen chapters (normally no. 1) available online.

<http://tartanhen.co.uk/>

## **E-Fiction**

“A new art form is about to be born – interactive, electronic works that are about people, story, and theme. It will do computer entertainment what Henry Fielding’s Tom Jones did for novels and D.W. Griffith’s “The Birth of a Nation” did for movies.” Developed by Chris Crawford, doyen of computer games. Crawford’s authoring software is known as the “Erasmatron” – available from the author.

<http://www.thuntek.net/~scg/sts/efiction.htm>

## **OpeneBook Forum**

“The Open eBook Forum (OEBF) is an association of hardware and software companies, publishers and users of electronic books and related organizations whose goals are to establish common specifications for electronic book systems.” A commercial version of AHDS (above). Started in October 1998 at the first electronic book conference sponsored by the National Institute of Standards and Technology at Gaithersburg, Maryland. A meeting held in San Francisco in January 2000 led to the creation of the Open eBook Forum.

<http://www.openebook.org/>

## **Softbook Press**

Founded by Jim Sachs and Tom Pomeroy “to redefine the way large volumes of information are distributed and read. ... Softbook Press has developed patented technologies that are transforming the way organizations of all sizes distribute, access, and use reading material. The company is currently building strategic partnerships with leading publishers and content providers to make available an ever-expanding library of online material. Through the SoftBook system, SoftBook Press provides a cost-effective way for any organization to distribute and manage information – without paper.” Bought by Gemstar International in January, 2000. Based in Redwood City, California.

<http://www.softbook.com/>

## **References**

---

---

<sup>i[1]</sup> The website devoted to this extraordinary venture is still under construction (20-04-00), but the home page states clearly: “1) The Revival of the Ancient Library of Alexandria Project aims at building a universal modern public research library to be a center of culture, science and academic research. 2) The Library is to provide both the national and international communities of scholars and researchers with unique collections and facilities focusing on Alexandrian, Egyptian, ancient and medieval civilizations as well as on contemporary disciplines. ... 3) The Bibliotheca Alexandrina shall sponsor intensive studies on the historical and contemporary cultural heritage of the region.” <http://www.greece.org/alexandria/library/> and <http://www.bibalex.gov.eg> and <http://www.unesco.org/webworld/alex/>

<sup>ii[2]</sup> It is probably sensible to regard “major” collections as those with more than one million items.

<sup>iii[3]</sup> Universal Resource Locators: the Web address used to locate the conversion. Web sites often migrate from one location to another, and this is likely to prove problematical for any cataloguing agency to deal with.

<sup>iv[4]</sup> I estimate that the present total for electronic texts is of the order of 50,000. This is less than one percent of the total holdings of the principal university libraries in the U.K.

<sup>v[5]</sup> Lesk, Michael (1997) *Practical Digital Libraries*. San Francisco, Morgan Kaufman, 1997. Lesk never really addresses the funding question, though sections 9.3 and 9.4 (pp. 204-211) look at possible models for funding digital libraries, including, as with videos, the rental model. Historically, this makes some sense, since between 1700 and 1890 there were literally thousands of circulating and subscription libraries worldwide which brought expensive print into the hands of the working classes.

<sup>vi[6]</sup> SGML is often referred to as the parent language, with HTML [HyperText Markup Language] and now XML [Extensible Markup Language] as special derivatives. XML is not a fixed format, and is designed to enable SGML on the Web. For a helpful paper – “Frequently Asked Questions about the Extensible Markup Language – is available at: <http://www.ucc.ie/xml/>. For a similar paper on SGML note: <http://www.infosys.utas.edu.au/info/sgmlfaq.txt>.

<sup>vii[7]</sup> Alston, R.C. (1992) ‘Preserving the Record’. *Archives*, xx(88), October 1992.

<sup>viii[8]</sup> San Francisco, Morgan Kaufmann, 1997.

<sup>ix[9]</sup> Edited by Leona Carpenter, Simon Shaw and Andrew Prescott. [London], The British Library, 1998.

<sup>x[10]</sup> The collection of tracts on health and sanitation (inter alia) collected by Edwin Chadwick in the British Library, bound in 535 volumes (CT1-535).

<sup>xi[11]</sup> Johns, Adrian (1998). *The Nature of the Book: Print and Knowledge in the Making*. Chicago & London, 1998. Johns’ book is the “first real attempt to portray print culture in the making”: without the survival of different copies of the works with which it is concerned it could not have been written.

<sup>xii[12]</sup> An idea I first suggested to him in 1986, but which seemed impossibly ambitious at that time; as did my suggestion that he publish in machine-readable form Migne’s *Patrologia*.

<sup>xiii[13]</sup> Every week I receive information of library proposals to digitize collections of materials of minor importance.

<sup>xiv[14]</sup> One reason for this is the fact that it is not always understood that the metadata which form part of a website, in which the materials are classified in such a way as to ensure their capture by the major search engines, is crucially important. The most effective metadata protocol is what is known as “Dublin Core”, developed by OCLC. See: “Dublin Core Metadata Initiative”, <http://purl.oclc.org/dc/>.

<sup>xv[15]</sup> Project Gutenberg was, from the start, designed as a supplier of etexts via ftp [File Transfer Protocol]. These have been either keyed (early days) or scanned using OCR software, proofed and sent to the project’s base at Champaign Illinois. A text may have several corrected versions: Milton’s *Areopagitica* had, by 1996, ten corrected versions available for ftp download. I have never understood who use these texts, and for what purpose: in the current list of titles available 90% are in print.

<sup>xvi[16]</sup> There is the Glassbook Reader, the Glassbook Plus Reader, and other software products for book publishers, distributors and booksellers, as well as libraries. There is even Glassbook Kiosk – “An e-book purchasing station for bookstores, coffee shops, or anywhere book lovers gather.” <http://www.glassbook.com/>.

<sup>xvii[17]</sup> NetLibrary for PCs – <http://www.netlibrary.com>.

<sup>xviii[18]</sup> Rocket eBook for palm devices and PDAs – <http://www.rocket-ebook.com>.

<sup>xix[19]</sup> Peanut Press for PDAs – <http://peanutpress.com>.

---

xx[20] Softbook for PDAs – <http://www.softbook.com>.

xxi[21] Softlock for PCs – <http://www.softlock.com>.

xxii[22] <http://www.microsoft.com/opentype/cleartype/>

xxiii[23] These sites are, of course, for students whose principal language is English; there are hundreds more for universities in Germany, France, Holland, Scandinavia, Italy. In 1999 a number of sites came online for oriental studies.

xxiv[24] Tanselle, G. Thomas (1998) 'Reproductions and Scholarship'. *Literature and Artifacts*. Charlottesville, 1998, p. 88. Tanselle's strict adherence to the importance of original texts runs as a theme throughout this collection of essays and is the most coherent statement of the dangers implicit in the transfer of texts from paper to electronic form.

xxv[25] <http://www.uky.edu/ArtsSciences/English/Beowulf/eBeowulf/guide.htm>. This is the online *Guide* to the CDROM edition of the text and full supporting documentation, edited by Kevin Kiernan. The Electronic Beowulf proved an expensive project because of the number of high quality colour images. The camera used was a Kontron, now marketed as the JenOptik ProgRes 3008. Typical file sizes for the JPEG colour pages is 20MB. The CDROM images are severely compressed.

xxvi[26] For example, the Cambridge University Library site (<http://www.lib.cam.ac.uk/Digital/>) makes available MS. Ee.3.59, and Anglo-Norman verse life of Edward the Confessor, written ca. 1235. Several illustrations from manuscripts in the Vatican Library are available at the Library of Congress site "Rome Reborn" (<http://lcweb.loc.gov/exhibits/vatican/toc.html>). Unfortunately, no site I am aware of provides technical information on the conversions.

xxvii[27] The best standard is Dublin Core: see note 10.

xxviii[28] My own sampling of films made of items in STC (1475-1640) and Wing (1641-1700) by University Microfilms suggests that about 50% of the films made before 1980 are difficult to read, and certainly could not be used for conversion to digital form using any of the currently available microfilm scanners.

xxix[29] A serious problem has affected the choice of texts for this project: namely, the preservation treatment of books at risk, making them impossible to film, especially if lamination has been prescribed. The massive use of lamination in the British Library makes it unlikely that books treated this way will ever be digitized. It is ironic that literally thousands of items which might have been ideal candidates for a digital library project must now be preserved as they are for ever.

xxx[30] "Reality Check – The Future of Libraries." The experts were: Ken Dowlin, former librarian of the San Francisco Public Library; Hector Garcia-Molina, Professor at Stanford University, and Principal Investigator for the Stanford Digital Library Project; Clifford Lynch, Director of Library Automation, University of California; Ellen Poisson, Assistant Director, New York Public Library; Robert Zich, Director of Electronic Programs, Library of Congress. [http://www.wired.com/wired/archive/3.12/reality\\_check.html](http://www.wired.com/wired/archive/3.12/reality_check.html).

xxxi[31] Lesk, Michael (1997) *Practical Digital Libraries*, p. 270. Bush's landmark essay "As We May Think" was published in *The Atlantic Monthly*, July, 1945. In 1967 Vannevar Bush published *Science is not Enough*. An essay in that work is entitled "Memex Revisited" in which he, like Lesk, believes that a radical alteration of the ways in which we learn and think is just round the corner.

xxxii[32] The 11<sup>th</sup> edition of the *CD-ROM Directory94* was published in paper format in 1993 (London, TFPL) with 1073 pages. It is now published only electronically on CD-ROM.

---

[HOME](#)