РОССИЙСКИЙ НАУЧНЫЙ ЭЛЕКТРОННЫЙ ЖУРНАЛ
Электронные библиотеки
Издается с 1998 года
ISSN 1562-5419

Russian Digital Libraries Journal - 1999 - Vol 2 - Issue 3

Towards the Hybrid Library *)

Chris Rusbridge
Electronic Libraries Programme

Introduction

Achievements in research programs are seldom what was envisaged at the start. It is still too early to judge the success or failure of the first phases of the UK Electronic Libraries Program (eLib), as several projects are still not completed. Despite this, some lessons learned are becoming apparent. In this article, I reflect on a little of what has been learned, and explore some of the implications. The eLib program has a heavy evaluation component, but the views expressed here are personal, deriving as much from living and working with the projects for the last three years as from any formal evaluations at this stage. From these reflections, I look forward briefly to the latest phase of eLib. My aim is to argue for the concept of the *hybrid library* as a logical follow on from current developments. Institutions should remain an important focus for digital library activities, and users in those institutions require the sort of integration of digital library services which the hybrid library promises.

Background and context

Many *D-Lib* readers will be very familiar with the US Digital Libraries Initiative (DLI) [1], funded by three agencies of the USA federal government: the National Science Foundation, the Defense Advanced Research Projects Agency, and the National Aeronautics and Space Administration (NSF/DARPA/NASA). I understand this has been mostly a large-scale computer science research program. But the title and the subject area make it an obvious point of contrast with eLib. After attending a couple of their conferences, my sense is the word seemed to be "ASQ-not", meaning "don't automate the status quo". The participants aimed (properly) to be innovative and free-thinking, leaving aside the constraints of existing practice. The results are exciting and extraordinarily interesting, but it is very hard to determine how many of these ideas might be effectively deployed in real life situations. It is notoriously difficult to transfer new technology from experiment to practice, but this is clearly harder the more distant the experimental context from real life.

By contrast, the eLib program characterised itself right from the start as "development" rather than research. The Joint Information Systems Committee (JISC) [2] , which is the parent organisation of eLib, does not fund research in the same way the NSF does (or, for example, the UK Research Councils do). Rather, the mission of the JISC is *to stimulate and enable the cost effective exploitation of information systems and to provide a high quality national network infrastructure for the UK higher education and research councils communities*; in this context, JISC funds a number of development programs aimed at supporting universities by piloting the use of appropriate new technologies. Unlike the fundamental research characteristics of the NSF and similar agencies, JISC's projects are concentrated at the near-market, practical application end of the spectrum. Both are needed. The eLib work is still research, despite a curious disdain for the word in some quarters.

An important precursor of eLib was the UK Higher Education Libraries Review. This was set up by the Higher Education Funding Councils (HEFCs, which fund universities in the UK) to look into the needs of libraries across the sector. It followed major changes which occurred in the late 1980s, notably the doubling of the number of universities through re-classification of tertiary level but largely teaching institutions. The Review, chaired by Professor Brian Follett, reported in 1993 [3]; the report was one of the most influential of recent times and resulted in several initiatives, large and small, unlocking significant spending (totaling more than J100 million).

Ironically, nearly half of this spending went on a capital program extending, re-furbishing and re-equipping library buildings. The other large component was the Non-formula-funded program for humanities special collections and archives (known infelicitously as the "NFF" program). The eLib program (phases 1 and 2) cost just J15 million over 3 years; in this period JISC (a joint committee of those same funding councils, and funder of the eLib program) also spent about J24 million on data services and probably J50 million on networks.

Web2PDF
converted by Web2PDFConvert.com

eLib phases 1 and 2

Chapter 7 of the Follett report referred to above (and written, as mentioned, in 1993) related to the implications of information technology in improving operations and capabilities of over-stretched libraries. The original eLib program grew out of the effort to implement that Chapter 7, and needs to be understood in that context. The program did not, in itself, attempt to create the electronic library, or even to project a particularly coherent view of such an endeavour. Instead, it aimed to address a number of important possibilities which had been identified in the review process. Some 5 years later, although our understanding has moved on in many ways, several of those early projects have still a few months to run. It takes a long time to get big programs rolling.

There are two phases to this early part of eLib; the first call was issued in 1994 [4], and the second in 1995 [5]. The latter was mostly to cover areas which reflected perceived weaknesses in the first phase. Effectively, these two phases can be considered together.

What lessons can be drawn from these early phases of eLib? At the detailed level, there are far too many to cover in this article. But a few stand out which help to understand our subsequent thinking. At first glance, they may seem obvious or commonsensical. They were nevertheless more or less painfully achieved, reminding us how rare a trait common sense can be (at least, in foresight rather than hind-sight).

The program areas of these phases and a few of the lessons are given in the table below:

Electronic publishing	Electronic journals	the Internet plays havoc with business models
	Pre-prints & grey literature	electronic publishing is not free
	Quality assurance	quality is not cheap
	Digitisation and images	digitisation is expensive; any particular item is often hard to justify; critical mass is important in both the size of the image base and the size of the markets
Learning and teaching	On demand publishing & electronic reserve	time is against us, and copyright clearance takes too long; OCR (proof-reading) costs too much to justify for single use; images of texts cost too much to down-load and print
Resources access	Access to network resources	quality costs, but users demand quality; business models are not obvious; and producing and relying on software in the same program is risky (stupid?)
	Document delivery	new services are slow to establish; new systems are hard to produce to service (not pilot) quality; supply services need comprehensive catalogues/systems (to find the stuff before buying it)
Supporting studies		human systems resist change
Training and awareness		people, not technology, represent the important issue

Information about eLib and the projects can be found at URL http://www.ukoln.ac.uk/services/elib/, and many aspects of it have been covered in the electronic magazine *Ariadne* [6].

There are many general lessons, including problems of inexperience in project management, difficulties of communication in consortium projects, and many difficulties in dissemination. However, the most outstanding lesson from this phase and from JISC's other work in provision of datasets is the need for integration.

Dissemination

The problem of dissemination of results and learning has been hinted at already. This has been a major issue for eLib and is certainly not resolved. We see dissemination as including not just telling about results and writing about them, but as Colin Harris, Librarian at Manchester Metropolitan University, puts it: "supported and assisted take-up and implementation" (verbal communication). In practice, this has not

Web2PDF
converted by Web2PDFConvert.com

been delivered successfully in most cases, but is an aspect we are emphasising much more strongly in later programs.

Evaluation

The eLib program has, as noted, a major evaluation component, devised by the Tavistock Institute. Each project has evaluation elements, and the program has a formative evaluation component (effectively, to help us do later bits better by learning from the earlier bits) and a summative component (to be conducted later, to determine how well we spent the resources on delivering our brief).

Supporting studies

As well as a small number of significant-sized supporting studies, there have been a large number of small studies on a wide variety of topics, which we categorise as evaluation, preservation and general. Many of these studies are being published in print form, and most will be available for personal use freely from the eLib web site at URL http://www.ukoln.ac.uk/services/elib/papers/supporting/.

The problem of integration

As well as dividing digital library (DL) activities today into research and development, I see a division between those centred on resources and those centred on organisational or service contexts. When building a DL around a specific set of resources, often selected with a useful degree of coherence, it is easy (or at least, solvable) to create usable and useful systems for users (we need to distinguish carefully between people-based services and IT-based services, which in this article I will call "systems").

For librarians struggling to create usable and useful systems and services for their institutions and patrons, life is not so easy. The jacket cover from *Gateways to Knowledge* [7] has an interesting quote: "Struggling to define the library of the future, librarians have too often bolted new technology, programs and services onto existing library functions".

It is right that librarians have taken this path. Each new system offering in this rapidly changing environment is a separate decision, and usually cannot be rejected on grounds of non-compatibility with other offerings. But there comes a time when we must review our range of systems and cry, "Enough". Time to take a stand; we must find a way to reduce the range of system interfaces, to a small number if not to one. (To reduce to one is unlikely and probably not even desirable; it is clear that Geographical Information Systems call for a very different set of interfaces from bibliographic systems.)

Details will differ in different countries, but the general principle of what follows will remain. Consider the following scenario. In a British university like the University of Warwick where the eLib team happens to be based, library patrons face a huge range of interfaces, including:

- local OPAC (telnet/web)
- COPAC union catalogue (telnet/web/Z39.50)
- Regional virtual union catalogue (web/Z39.50)
- stand-alone CD-ROMs (29 at Warwick)
- off-line CD-ROMs and diskettes
- networked CD-ROMs (62 at Warwick)
- full text services
- electronic reserve systems
- remote datasets at the community data centres (BIDS, EDINA and MIDAS; 12 or so at Warwick)
- remote datasets at other universities
- remote commercial datasets
- local datasets e.g., bibliographies, pamphlet collections and archives (9 at Warwick)
- local web-based documents, library and institutional
- local web resource gateways
- remote web subject/resource gateways
- remote web resources
- remotely mounted electronic journals (EJs)
- local and remotely mounted e-books

It's interesting that it is not easy to get statistics on many of these. On reflection, the list above singles out digital media in a way that most researchers would not. To be fair and even-handed on the media question, one should perhaps add to this list:

- books: loanable, reference, and available via interlibrary loan (ILL)
- paper-based journals
- special collections, maps, slides, sound recordings and videos

Clearly, there are good reasons why some of these resources are treated differently and have different

Web2PDF
converted by Web2PDFConvert.com

interfaces. Maps, for example, require a different finding and storing interface from text, even in the paper-based world. But many of the differences are incidental and are not functionally driven. The differences between the many CD-ROMs devoted to abstracting and indexing products are generally only excusable through the accident of authorship.

The web is beginning to provide an interface paradigm which overcomes some of this heterogeneity. Any resource made accessible over the web is likely to show consistency in such ancillary functions as printing and down-loading (the areas where CD-ROMs show most divergence), unless of course the web browser merely delivers the resource to another system such as a word processor or Adobe Acrobat (although even here graphical user interfaces provide some consistency). But this presentational level of integration still often leaves a great deal of annoying and meaningless (frustrating?) incompatibility.

As an example, consider yourself as a researcher interested in a matter which crosses several disciplinary boundaries, in this case engineering, medicine and social conditions (as, for example, the effects on social conditions of improving health resulting from the great nineteenth century sanitary engineering developments). You might wish to search three subject-based gateways provided by eLib projects: EEVL [8], OMNI [9] , and SOSIG [10]. Although there is a common set of facilities provided by each of these services, there is a significant degree of difference in the structure of the pages. I so often find myself bemused by the interface presented by a resource, and distracted from the research mission which brought me there in the first place. We have not yet converged to the well-known structures which have evolved over centuries on the print world.

A related idea in JISC is the Distributed National Electronic Resource. It is little more than an idea at this stage, but it embodies ideas of integration in the large scale.

I will return to some of these issues in the later section on technology and architectures.

Work spaces

Richard Heseltine writes: "Communications and Information Technology enables the integration of information and work. They no longer have to occupy separate spaces... A student's work and information space might include learning objectives, learning tasks, modes of assessment, methods of feedback, and a variety of learning resources, including tutors, laboratories, documents and datasets, and modes of audio, video and textual communication. All of these components are capable of being reproduced electronically..." [11].

This concept of integration of information and work spaces is important. There are two ways of looking at how we use resources (data) in the electronic environment. First, what do we do with or to it, i.e., what is it for? Second, how do we manipulate it?

Faced with electronic information, scholars might want to do some or all of the following actions with it (not all of which are legal without a specific licence), *as information*:

- read it on screen
- study it
- follow links (read it as a pointer)
- use it as a clue to other information
- evaluate quality of it
- assess quality of our own knowledge against it
- write down key facts from it
- print or save to our personal library
- send it to colleagues
- criticise and provide feedback on it
- discuss and debate it
- analyse it
- add to it
- include it in a database
- include information from it in a database
- quote it in research or teaching
- refer to it in research or teaching
- point to it from a web page or other resources
- cause it to point to our resources
- sell it or re-sell it
- sell services based on it
- do funded research based on it (fine distinction there!)
- buy services based on it

These are information utilisation issues. There is also a set of linked technologically-based *processing* issues, which may not match too well. To undertake the information actions noted above, we may also have to

Web2PDF
converted by Web2PDFConvert.com

undertake some or all of the processing actions presented below.

- find the resource
- provide authentication and obtain authorisation to use it, which might involve prior steps
- agree to any specific conditions
- read notes and news
- get general help on using the resource
- navigate some imposed structure (to a search screen, e.g.)
- browse conceptual structures and review results, or
- make searches and review results
- get help in context with problems
- refine searches
- preserve searches
- follow links, perhaps to different resources
- return
- bookmark resources, and manage bookmarks
- save, print or email results
- extract results to re-use in other documents
- receive peripheral information (including possibly advertising)
- order related services
- receive related services
- pay for information and related services
- provide comments and feedback to providers
- contribute to related discussion fora
- end session by choice or automatically

and probably

- indirectly and unwittingly provide information on our activities!

These lists may not be complete, but they say something about the context, especially the terms and conditions on which we need access to resources, and the supporting facilities needed. This is not about collections of passive information, but practical, usable information resources for learning and research.

Given the range of actions of both kinds in the lists above, we need to explore this integration between information and work spaces. Information is there to be used. Although much of it is read and discarded or ignored, and some is read, understood, learned and becomes human knowledge without further processing on our part, a considerable part of it is used in many of the ways indicated in the context of research or study. In these circumstances, users must be able to access the information while carrying out some other activity. This might be writing an article or term paper, but could include many other applications including entering data into spreadsheets and databases, etc.

An unfortunate contrary example to this was the European COPICAT project [12], which created an early example of an extremely secure ECMS environment. In the version demonstrated, you could read the controlled document, but all other applications, printing (except through a controlled interface) and cut and paste were disabled. To write anything about what you had read, you had to leave the COPICAT environment, and hence lose sight of the information. Clearly this is absurd (and illustrates that we need to concentrate our efforts more on compliance than control), but related examples occur which hardly raise an eye-brow; for example, dedicated OPAC or CD-ROM stations where the reader must also write down all references and quotes before moving to some other system to carry on working. Situations like this show the wide-spread need to support real integration between information and work spaces.

It is fashionable to see this need to integrate information and work spaces as liberating the user from the confining walls of the library, and therefore as negating the need for the library. While it is true that an important and valuable liberation can occur, there remain critical and important roles for libraries in institutional contexts, and crucial roles for librarians in many individual and group support contexts. The details may change, but the need remains.

Political divides

Yet more sources of incompatibility arise from the many political divides that sometimes make sensible things difficult. Often of extreme artificiality, due to various governing and funding structures, these divides prevent various providers from working together, or support them in ignoring each other, to the detriment of the user. In the UK these divides exist (1) between the JISC (Information) and many learning and teaching activities funded through the very same set of funding councils; (2) between academic, public and special libraries, archives and museums; and (3) between government, academic and commercial sector publishers and data providers. Richard Heseltine again in the article referred to above says of the first of these: "I understand full well that there are political reasons for that divide, but I also believe it reflects a

Web2PDF
converted by Web2PDFConvert.com

failure of librarians to engage with learning issues, and a failure on the part of academics to appreciate the need radically to rethink the provision of learning resources in a massified and transformed system of higher education." It is important to work to integrate access for end users of information, as far as possible, across these political divides. No doubt similar political divides exist in the USA and many other countries.

Authentication

One of the many issues acting against effective use of electronic and digital material is the wide range of authentication procedures with which we have to comply. To use the huge print resources of a major library, I have either one authentication procedure -- to become in some sense a "member" of the owning institution -- or often no procedure at all, where the library allows "walk-in" access -- as do public libraries, and many academic libraries. There is, of course, an increasing need for authentication even in these contexts as I start to do more expensive things than reading books from the shelf; even to borrow a book will require presentation of a library card in almost any library.

But authentication in the digital world can be a major problem, especially if every resource has a separate and different authentication challenge.

It is pleasing that the need to simplify authentication is being widely recognised, in the UK through JISC's interim ATHENS system, and in the US through the CNI authentication initiative. But it is worth noting that many separate authentication schemes and requirements -- and particularly many different username/password combinations required for a single user -- will greatly retard significant use of a wide range of digital resources, in the process possibly making them uneconomic.

In sum, integration can be understood as embodying three issues: work spaces, political divides, and authentication. The eLib Phase 3 hybrid library projects (to which I will return later) have them at the heart of their agenda.

Service

After integration, we come to the issue of service. Service impacts on digital and electronic libraries in very many ways.

A big issue for many UK universities (whether they are for it or against it) is convergence between computing and library services. Clearly, there is a certain logic to this in the context of digital library development. But studies of convergence in the eLib project *Skills for New Information Professionals* (SKIP) [13] have shown that a major difficulty in convergence is the different attitudes to service. Penny Garrod and Ivan Sidgreaves write: "Library staff in several institutions saw their computing colleagues as having different priorities to them, in relation to user needs, and as being concerned primarily with IT systems, rather than the service provided to users... Some computing staff described library staff as clinging to outmoded notions of professionalism, and as being over-zealous in helping students, who needed to be encouraged to become more independent as learners." [13].

The next quote from the same source illustrates the general problem: "There is strong evidence that there are still many unhelpful barriers between library and computing professionals... In general there is an absence of understanding of what each group is doing, and a lack of respect for each other's professional skills". There appear to be quite different notions of service in the two communities. On the one hand, support for the individual learner, on the other hand a sometimes desperate struggle to keep a (usually) under-funded communications and IT infrastructure operational for the benefit of all learners.

Perhaps professional pride and status is getting in the way, and all parties should instead be focusing on how their services can support the learning, teaching and research strategies of the university.

Clearly, service is a critical issue which as Bernie Sloan notes (article seen in draft [14]) has been largely ignored. He writes: "Human interaction in the digital library is discussed far less frequently. One would almost get the impression that the service tradition of the physical library will be unnecessary and redundant in the digital library environment". He suggests instead: "Digital library proponents must consider the role of people (as users and service providers) if the digital library is to be truly beneficial. Technology and information resources, on their own, cannot make up an effective digital library". Sloan goes on to discuss service mainly in the context of the digital reference librarian, physically separate from the reader, but service also manifests itself in simple issues like information about the range of systems available, documentation on their use, support and help in the face of problems, assistance and training of users in their use. However optimistic we may be, the use of digital library resources is by no means sublimely obvious to many of their potential users, and it is certain that many problems cannot be resolved through appropriate graphical interfaces alone.

Much of the issue here is the role of the library relative to its community, often undertaken by the subject librarian, in engaging with the needs of its community and getting closer to understanding the customer.

There is an enormous range of available information in the world. The role of the library is to select, acquire, organise and make available an appropriate subset of these resources. This focusing in from the universe of resources is significant in many contexts, including for learners. Partly this is an issue of quality control. A substantial part of the HEI community is perhaps ill equipped to assess the quality of the information

Web2PDF
converted by Web2PDFConvert.com

resources they may find. The library has a role here in the digital world as with print -- not just in excluding access to rubbish, but in encouraging access paths to quality.

Certainly, services must be offered by libraries which provide access to material outside their primary purchases. But the goal of the library is to match the institution's needs and budget against the available information and its costs.

Prices, costs and licences

Increasingly, digital information comes with a price tag, sometimes frighteningly large. Many information resources will be out of reach for individuals, and even departments. But the library, negotiating on behalf of the institution (and perhaps as part of a consortium) has a chance of providing access. Again, the match of the mission and needs of the institution against the available budget is the critical role.

An important question here is the scope of the library, and therefore of the licence it signs. The scope of a traditional library is fairly clear -- it has a collection of physical objects which are either removed for loan (one person at a time) or consulted within its walls. We *could* (but prefer not to) define the scope of an electronic library almost exactly as the scope of a traditional library. Consult the electronic collection, one item at a time within the walls of the library. This is such a major restriction on what we want to achieve with the electronic library that it is transparently absurd. Yet it is the only model that faithfully maintains relationships with publishers and, for example, protects revenue from personal subscriptions.

The more the scope is widened, the more problems are introduced. Many want new models for electronic libraries: discipline based, national or even global in scope. The broader the scope, the greater the impact on publishing. In the end, a global library is a global re-publisher, and the price for the single subscription becomes the price for the whole world.

I believe an institutional scope is appropriate, even though it might tend to erode personal subscriptions and copies. It is increasingly likely that scope will be extended to the environs, at least to the homes of staff and students. These changes tend to push up the price charged by publishers, who see the need to recover their fixed costs.

Costs are not necessarily lower than for print publishing, either. Although there is a significant saving in the cost of paper and distribution, obviously the costs of parallel publishing are higher than for print publishing alone. The costs of electronic-only publishing, but with significant value-added or multi-media content are also significantly higher than for print publishing, as *Internet Archaeology* [15] found out. And in the UK, value-added tax (VAT) is added to electronic information but not to print.

With prices come licences. Pretty well all the available information is covered by copyright in one form or another. Copyright is a major issue in digital libraries: a big motivator, the defining issue for the information marketplace, and a big problem. In the digital world, sadly, copyright means licences. Whereas I can own a copy of a book or a journal, in the electronic world I am only licensed to use it. Licences are required for all access, whether implied or explicit. The terms under which I own a book or journal, and what I can do with it are uniform, defined by copyright law; the same rules apply to every one of the millions of volumes in a major library. Unfortunately, the terms of licences in the electronic world are limited only by the imaginations of corporate lawyers. There will be a variety of terms and conditions, a variety of pricing models, and all possibilities are likely to be exploited. Given the likely huge number of licensed digital objects (eventually millions) and the large numbers of copyright owners (many, many thousands) this could result in an absolutely untenable situation for libraries struggling to achieve compliance and to provide a usable service to their patrons.

Some sort of uniformity is desperately needed for the sanity of all of us. Whether this can be done without creating cartel or anti-trust problems remains an open question. We do see libraries getting together to define a small number of model licences (such as, for example, the work in the UK between JISC and the Publishers Association on model licences, fair dealing, and ILL in the electronic environment [16], and elsewhere in ICOLC), and this will help but may not solve the problem. There are always likely to be more publishers unaware of such developments than aware of them (given, for example, Derek Law's comment that the median number of journals per publisher in a UK university is one [verbal communication]).

The variety of these licences requires, therefore, that they be codable as terms and conditions metadata and automatically processable, if we are to build workable systems.

Libraries are well placed to negotiate licences for their institutions, and to provide major information management systems, both digital and human, to facilitate controlled access to material in ways which comply with those licences. Alternatively, where access cannot be controlled appropriately, libraries can warn HEIs against signing licences which can leave them liable to legal action.

Role of the library

I am trying to argue here that the library and the HEI have continuing roles in the digital library environment. It is interesting to read in *Gateways to Knowledge* [7] the argument amongst the writers for and against the metaphor of "place" in the concept of the gateway library (which seems to match in many respects what we call hybrid libraries). Anthony Appiah writes:

Web2PDF
converted by Web2PDFConvert.com

"The library I never go to is already one of the most important places in my life" (Realizing the Virtual Library, Anthony Appiah, in *Gateways to Knowledge*, ed. Lawrence Dowler MIT Press, Cambridge Mass, 1997). But Richard Rockwell writes in the same compendium:

"The gateway library is ... a *process* that delivers services to the user" and "[the] gateway library is not a place ... [it] is as distributed and decentralized as campus computing has become." (The concept of the Gateway Library: a View from the Periphery, Richard C. Rockwell, in *Gateways to Knowledge*, ed. Lawrence Dowler MIT Press, Cambridge Mass, 1997.)

Perhaps the library is to become a cyber-place: an institutional and practical reality which comes to exist as much in cyber-space as in physical space.

Libraries *do* exist as important parts of our institutions. They command public respect, and already provide huge quantities of information, both print and electronic. They also provide other important facilities, including vital study space and an important social environment. This is not just about sex, but mutual support in information-seeking behaviours which the isolated user misses (and will often still miss even if electronically well connected). The library as a place will have a continual role in institutions. It seems unlikely that student use of libraries as places will diminish much in the foreseeable future (despite concerns about the group of students who "don't do libraries"), even if staff use continues to decline as they access resources more from the comfort (and isolation) of their offices and homes. Staff expect the library to come to them.

Overall, whatever the arguments about place as a metaphor in gateway or hybrid libraries, or against place as a determinant for access (community or institution, not site!), the institution and the library as agent for it are critically important in digital libraries. We must provide integrated access for our community to a wide range of resources, placed in a service context.

The cry should not be "ASQ-not". We should instead ask how we can build on the skills and experience already there. New services invented from the ground up are important, but eventually they will have to be fitted into a service context.

Resources

It should be clear from this discussion that the aims of eLib were very different from the aims of the DLI and other advanced research projects. One way of characterising these differences might be that the DLI projects mainly started from resources and worked from there to the users, while eLib attempted to start from the needs of the institution, the user and the librarian and work towards the resources (there are of course many cases where this analysis is far too simple, but it remains an interesting difference in approach).

It may be useful at this point to have a quick look at the range of resources we are contemplating in digital libraries. Different types of resources confront us with different requirements. For this paper, I have used my own classification of resources: *legacy, transition, new,* and *future*.

By *legacy* resources, I mean largely non-digital resources, including manuscript, print, slides, maps, audio and video recordings. Here, we have had the support of digital services in management roles, and these should increase. For example, we can build services which support end user discovery, location, request and loan of these materials on an inter-institutional basis. We can even offer premium services providing service to the academic desk for a price. But the vast majority of existing legacy resources will remain outside the electronic domain for many years to come, despite the huge investments in digitisation. These legacy resources are a major reason why existing libraries remain fundamentally important.

By *transition* resources, I mean legacy resources which are being or have been digitised, making the transition into the digital world. These are resources primarily designed for another medium. Looking just at text, we need to make a distinction between text digitised as images and text converted by OCR or other means into other formats. Or (it's not quite the same thing) between transitions which preserve page fidelity and those which provide the resource in a form suitable for consumption on-screen. It has to be said that many of the transition resources on offer at the moment, wonderful as they are in providing access to previously hidden treasures, are quite awful experiences in use. The mismatch of paper and screen formats, particularly for multi-column text, makes for extremely difficult manoeuvring and militates strongly against a smooth reading experience. If possible, most will use the paper version any time. Sadly, this too often applies to PDF as well, and makes the widespread adoption of this format one of the unhappier compromises we have been forced to accept. However, as a print-on-demand technology, PDF is not too bad and I suspect that is mostly how it is used.

Outside of specialist repositories, most transition material converted to text through OCR is still offered in PDF or HTML. Both give us very weak access to the structure of the information. The SGML community has been arguing for years that we need the rich markup facilities that SGML provides, and that this will aid analysis of the text in new and exciting ways, allowing, for example, citation linking. It is interesting to see the progress made by the eLib Open Journal Framework project [17], and others, in analysing PDF files to provide this sort of support. It can apply as much to HTML files; the project uses the concept of the external linkbase (the link as an externally authored rather than embedded resource). It is never 100% effective,

Web2PDF
converted by Web2PDFConvert.com

however, and SGML would provide a much more accurate, systematic capability. SGML has frightened most of us with its complexity; perhaps XML will provide a richer alternative to HTML in structuring these resources, with less of an entry barrier than SGML proper.

Making the transition for maps (see the eLib *Digimap* project [18]), images (see *HELIX* [19]), audio and video (see *PATRON* [20]) into the digital world raises many further issues, not least the much greater complexity of rights issues in these areas (complex enough for printed text!). But assuming appropriate rights have been negotiated, many further hurdles confront us, particularly in indexing and metadata, in content-based searching, and even in simple presentation. Standards in many of these areas are inadequately developed, and too often ignored.

Transition resources allow us to slightly reduce reliance on physical libraries on the one hand, and provide vastly increased access and often new modes of analysis on the other.

It would be good if we could report that *new* digital resources (whether expressly created as digital or created in parallel with print, etc.) were free of many of the problems in this digital world. But at present, the increased freedom from media constraints in the digital world seems to make things worse, not better (from an integration point of view). Individual resources are designed for a particular use and IT context, with a bit of specialisation, differentiation and branding thrown in, but with little thought for the user who needs to access many such resources. So there is an increasingly wide range of digital resources, from formally published electronic journals and (increasingly) electronic books, through databases and datasets in many formats (bibliographic, full text, image, vector/map, audio/video, statistical and numeric datasets).

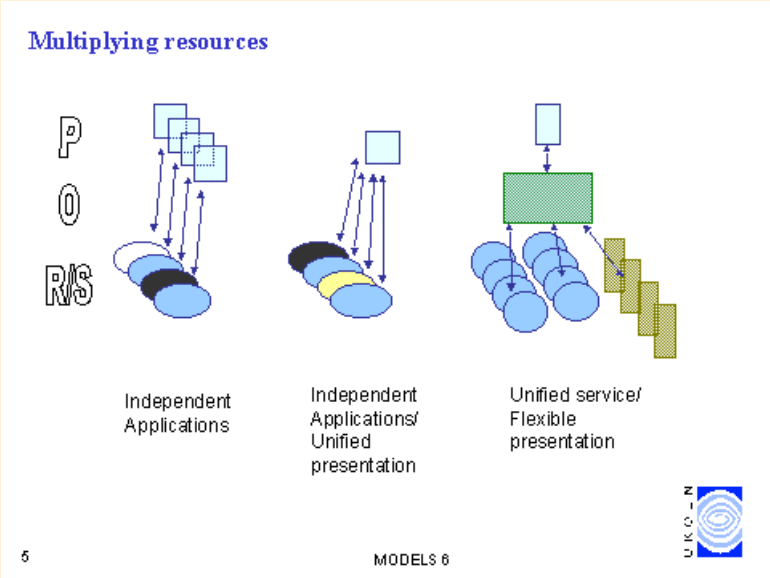
This huge array of datasets is a problem because it tends to assume, but not contain or even formally specify, an access method. We need to decide on the access methods from contexts which are available now but which may not be so even in the medium term future. The object-oriented world of digital objects, packaging the data resources and the access or processing method as an entity (particularly where machine independent code is used for the access methods) holds out the best hope for resources of the *future*.

Technology and architectures

MODELS

It's worth mentioning, at this point, the eLib Supporting Studies project *Moving towards Distributed Electronic Libraries* (MODELS) [21]. Through a series of workshops, this project has been tackling architectural issues in digital libraries. Perhaps best known outside the UK was the second MODELS workshop (URL), which was also the first Dublin Core (DC) workshop outside the US, and which started the development of the DC syntax and came up with the Warwick Framework.

Recently, MODELS workshops have been elaborating the MODELS Information Architecture (MIA). A diagram from a presentation on MIA by Lorcan Dempsey illustrates some of these incompatibility and integration issues.



Source: http://www.ukoln.ac.uk/dlis/models/models6/iap/sld005.htm

The left hand section shows a set of resources, each with its own interface. This is the CD-ROM world, and most other electronic resources until a few years ago. The middle section shows a set of resources accessed through a unifying presentation layer (the web). This is definitely better, but integration as described above is no more than skin deep. The right hand section shows a unified interface to a broker, software which attempts to provide semantic consistency in access to the target resources. Some readers will recognise the CORBA 3-layer model here. Although the broker can unify access to many resources at a general level, in doing so it may lose some functionality. Specialist access for the "power user" must

Web2PDF
converted by Web2PDFConvert.com

therefore remain.

Perhaps the difference, here, is between integration of connection and integration of operation. Indeed, the target is to provide complex systems which do not just display results for human consumption, but have some significant understanding of the results; understanding in the sense of being able to re-use them.

For example, imagine that you have asked a user-oriented document delivery system to search article datasets. Having discovered an article of interest in the *British Journal of Academic Librarianship*, the system will use the title to search catalogues; it should understand the holdings statements returned to decide which supply library holds the particular issue required, before sending the ILL request.

If we look up this journal in COPAC, the physical union catalogue of UK research university libraries [22], we find several entries (mainly due to failures in the de-duplication algorithm). Amongst these, there are several versions of holdings statements which are human-readable but not machine-processable, for example:

> Cambridge - contact Cambridge University Library ; P876.c.151.1- Vol.1 MISSING; Vol.1 (Spring 1986)-
> Edinburgh - Main Library ; SERIALS, Per. .01 Bri. ;Vol.1, no.1 (Spring 1986)-
> Edinburgh - EUL, MAIN, SERIALS ; Per. .01 New. [This different entry refers to the journal's continuation as The New Review of Academic Librarianship, but illustrates changing practice even within one library.]
> Glasgow - Main Library ; Bibliog Per B5250 // Library holds: 1986-1994.
> UCL - Main Library (South) ; LIBRARIANSHIP Periodicals

This particular example clearly shows the need for a degree of standardisation of holdings statements which is not yet available in practice.

Much of MIA can be implemented now, in the form of web gateways, with Z39.50 as a unifying access protocol. The protocol is beginning to have a critical mass in the library world and a few related areas (museums, government information, etc.), but not elsewhere despite its promises of generality in remote database access. Z39.50 may be a horribly complex protocol, but there appears to be nothing better on the horizon at the moment. The complexity of Z39.50 is in part due to its long history, but in significant part due to the complexity of the remote dataset access problem. Nevertheless, Z39.50 as a protocol is showing its age, and one wonders whether daughter of Z39.50 is waiting in the wings somewhere.

Scaling

The big issue with many new technological solutions is whether they will scale from pilots to the real world. As one example, can we design a virtual union catalogue for all of UK HE (using Z39.50 to access any of the 200 institutions) which does not suffer from significant scaling problems? For someone who wants to find the definitive set of resources on a topic in the UK, the temptation will be to try to search the whole lot. Not only would this multiply every transaction into 200 transactions (multiplied again by the networking costs of session startup and teardown implicit in web-Z39.50 gateway designs), but it would also add hugely to the transaction loads for smaller catalogues in the system (e.g., music or art colleges) and might force them to withdraw from the system. How much worse this would be for some 3,500 institutions in the USA. Much needs to be done with the design of interfaces and gateways, perhaps using query routing and automated collection summary or manual description techniques to support an accurate choice of the appropriate subset of catalogues to search.

Readers might be interested to note that a form of query routing is being used in the eLib subject-based gateway projects. In fact, the little scenario I painted above with access to EEVL, OMNI and SOSIG will be overcome shortly through the use of cross-searching using query routing based on centroids (see Jon Knight et al. [23]).

Physical union catalogues, of course, do not suffer from these problems, but are affected by different ones such as the cost of the centralised operation and delays before records newly added in the HEIs are available in the central system.

As an example of scaling problems from a different angle, take the example of COPAC [22]. This is a physical union catalogue of UK research libraries. They report about 5,000 searches or browses on the web interface per day. About one third of these were from all the institutions in UK higher education, while a similar number originated from one university, Karlsruhe in Germany. It appears that access to COPAC has been designed in as an adjunct to OPAC searching at Karlsruhe, initially as a default additional search, more lately requiring a choice by the searcher. While COPAC can cope with this high level of use from one site, changes from individual to systematic access in this way from many institutions could bring it to its knees. Federating systems may not be without systemic problems.

Scaling issues turn up in many more ways. What sorts of systems could cope with millions of images or hundreds of thousands of hours of audio or video? Or the complete set of mapping data for a country (a reasonable resource for study)? What are the implications of some of these for our networking infrastructure? Many of our current systems are mere toys compared with what is realistically needed. There is a tendency to solve some of these problems by mounting resources at one specialised facility, perhaps with a small number of mirrors elsewhere. Libraries, then, only provide access, not the resource. Although "Access, not holdings" has been a catch-cry in the library world for some time, it is dangerous when

Web2PDF
converted by Web2PDFConvert.com

implemented in such a radical form as happens with many electronic resources, where only the publisher or agent and perhaps a few mirrors hold the actual data. It is not clear how this particular paradigm shift will affect the nature of libraries and of scholarship. The parallels with the lost library of Alexandria are very disturbing. In particular, the effect of preservation by diaspora is lost.

In real libraries, we find ways of organising huge quantities of (mostly) print material as the basis of scholarship. In the digital library world, we find the beginnings of systems to do the same with digital material. Hybrid libraries -- real libraries -- struggle to bring the same realistic scale and treatment to the digital as to the print, recognising the place of each and bringing them both to the attention of the scholar.

Phase 3

Our review of eLib phases 1 and 2 was not as conscious as the material indicated above might indicate. But from the beginnings of these ideas, Lynne Brindley and JISC's Committee for Electronic Information derived the eLib Phase 3 program. Whereas phase 2 of eLib was designed to fill gaps in phase 1, phase 3 came somewhat later, with more of an aim to bring many separate developments together. There was a strong feeling amongst the funders that we needed to build on successes without just having more of the same. Phase 3 was designed to have 4 components:

- hybrid libraries
- large scale resource discovery, or clumps
- preservation
- turning early projects into services

Hybrid Library

The hybrid library was designed to bring a range of technologies from different sources together in the context of a working library, and also to begin to explore integrated systems and services in both the electronic and print environments. The hybrid library should integrate access to all four different kinds of resources identified above, using different technologies from the digital library world, and across different media. The name hybrid library is intended to reflect the transitional state of the library, which today can neither be fully print nor fully digital. As we have seen, in so many cases the results of adding technology piece-meal are unsatisfactory. The hybrid library tries to use the technologies available to bring things together into a library reflecting the best of both worlds. There are 5 projects with different approaches:

- Agora - systems led [URL http://hosted.ukoln.ac.uk/agora/]
- BUILDER - institutional [URL http://builder.bham.ac.uk/]
- HEADLINE - learning landscapes [URL http://www.lse.ac.uk/blpes/headline.shtml]
- HYLIFE - wide range of client types [URL http://www.unn.ac.uk/~xcu2/hylife/]
- MALIBU - management implications, humanities area [URL http://www.kcl.ac.uk/humanities/cch/malibu/]

A major article on hybrid libraries is planned for a future edition of *D-Lib Magazine*.

Large scale resource discovery, or clumps

The clumps area was largely derived from early MODELS work, attempting to solve the problem of access to (mostly print) scholarly resources anywhere in the UK. The need for this was brought home by the document delivery projects, for how can you request delivery of a document if you don't know where it is? We felt that extending COPAC to include all 200 institutions would be expensive and was potentially unnecessary, given the potential of Z39.50. Further virtual union catalogues would potentially have access to more current data and more accurate status information than would be possible with a physical union catalogue. The idea was boosted by plans to enhance cooperation between libraries arising from the Anderson Report [24], now coming to fruition through the Research Libraries Strategy of the HEFCs. The clumps projects are:

- CAIRNS, all the universities in Scotland, addressing issues relating to the selection of appropriate subsets to search. [URL http://cairns.lib.gla.ac.uk/]
- M25 Link, a pilot group in the London area, addressing issues particularly relating to serials holdings. [URL http://www.lse.ac.uk/blpes/m25/]
- RIDING, Yorkshire universities and BL Document Supply Centre. [URL http://www.shef.ac.uk/~riding/]
- Music Libraries Online, a subject grouping of music conservatoires. [URL http://www.musiconline.uce.ac.uk/]

Digital preservation

Having caused some of the increasing move of scholarly resources into the digital domain, we felt it was irresponsible to continue to ignore or to sideline preservation issues. Although a number of studies into preservation were funded by JISC (managed for us by the British Library and under the aegis of the National Preservation Office), there were no long term strategies, and proposed legal deposit legislation for

Web2PDF
converted by Web2PDFConvert.com

non-print material seemed to be receding ever further into the distance -- not a political priority for either party. So we sought a project to explore issues with a view to developing recommendations for services. This is CEDARS [25], based in Leeds, Oxford and Cambridge.

Projects to services

Finally, although in general all phase 1/2 projects were expected to work out how they would become self-supporting by the end of the project, we realised this would not be achieved for all, and that some (if not many) would simply disappear if some further level of funding was not provided to help the transition to a self-supporting status. Most of these transitional projects are extensions of existing projects (not specifically listed here), but in two areas we sought new (or differently targeted) work.

The first of these was in building a simple electronic journal support system. This was originally thought of in comparison with High Wire Press, but the eventual project, EPRESS, is much smaller scale and aimed at the "do-it-yourself" startup electronic journal market. They will develop tools from those used to create *Sociological Research Online*, and provide these in two ways. They will manage much of the administration for an EJ (referred to as an in-journal), or provide the tools for academics to start their own EJ (referred to as an out-journal).

The second new project is in the On Demand Publishing and Electronic Reserve program area. A study (Halliday et al. [26]) had identified two major flaws with setting up these (potentially highly desirable) operations on an institutional basis. These flaws include the length of time needed to obtain permissions to use the material (77 days with 6 follow-ups on average), which is quite unrealistic in today's timescales (cf deadlines of 35 days before start of term for print reserve) [27]. The second flaw was the cost of digitisation, particularly if an OCRed version is needed (estimated roughly at J2.50 per page including proof reading and other processing). We felt that the only way to overcome these flaws was to try to build a resource bank of pre-cleared, pre-digitised material of interest in the learning process. We hoped that the resources already made available to the 12 OD/ER projects could (with the hoped for agreement of rights holders) form the basis of the resource bank. To this all new resources would be added. Of course, there is no way to avoid the time and cost to add new resources (although the recent move of the Copyright Licensing Agency into licensing digital versions should help greatly here); however, our hope is that it will build up gradually towards a critical mass of resources where it becomes unavoidably interesting for the teaching and learning process. This project is HERON, and like all the others in the last part of phase 3, it starts in August 1998.

Management lessons

We also tried to learn some management lessons from Phase 1/2. These meant we

- retained emphasis on evaluation
- continued to provide training for incoming project managers in project planning and evaluation
- took a significantly greater interest in the project plans, making the continuation of the project contingent on acceptance of these plans
- increased the emphasis on communication systems between projects
- required specific dissemination events to be built into project plans
- gave as much advance warning as possible of the start of projects, so that the project managers could be recruited ahead of time.

We have yet to see how effective these strategies will be this time around.

More work needed...

By the nature of the program, eLib projects tend to try to apply newly developed solutions, rather than to develop much in the way of new technologies. Areas where significant development is needed, I believe, include

- quiet authentication, providing authenticated access without the intrusion of authentication challenges
- rights or terms and conditions metadata, and automated licences, to allow automated processing of access management to digital library resources
- integrity and authenticity of electronic documents to replace the authority of the published work
- useful copyright management systems (supporting integration of work and information spaces while retaining reasonable control of the copyright material)
- simple information trading, including micro-payments
- more sophisticated electronic commerce
- scalability issues, including collection summaries and query routing
- metadata inter-operability (between different dialects, e.g., in descriptive metadata, or between different types, e,g., identification information between descriptive metadata and terms and conditions)

Web2PDF
converted by Web2PDFConvert.com

- daughter of Z39.50
- machine-independent, object-oriented packaging of digital library material
- information longevity issues -- digital preservation

DLI-2

Meanwhile, it is fascinating to watch the beginnings of the second phase of the DLI, and to wonder whether the additional sponsors (e.g., the Library of Congress, the National Library of Medicine, the National Endowment for the Humanities) with their different orientations will anchor the new developments more in real life situations, while still preserving the fresh originality of the first phase.

Conclusions

The Hybrid Library effort, the attempts to scale digital library resources into production services, plus digital preservation, are the most important parts of the current eLib program. Nothing in this program represents the last word, in the advances of knowledge or of practice in these matters. But the program and its projects represent several more steps forward, sideways and even maybe backwards in our understanding of these issues. It is important to take an ecological approach, to allow failures as well as successes. Visionaries can (and should) design for future. But if we design too much, some of our key bricks may fail and the whole edifice collapse. We arrive at the future day by day, and the small steps taken today to serve information needs now are as important as ones that will happen in ten years time.

I will give the last word to Ian Mowat, who writes: "The Follett Report, which is still guiding our development five years after its appearance, was brilliantly successful for a variety of reasons, not least being the skill of Sir Brian Follett himself in extracting the necessary funding and the openness with which the review was conducted by all involved. For me, however, one of its greatest strengths has been its practical realism. Although there was a very necessary element of visionary thinking, the main thrust of its recommendations was down-to-earth, practical and designed to take us half a step forward to a clearly defined future, rather than forcing us to make giant leaps into the dark."[28]

Notes

[1] NSF/DARPA/NASA Digital Libraries Initiative Projects, http://www.cise.nsf.gov/iis/dli_home.html

[2] Joint Information Systems Committee, http://www.jisc.ac.uk/

[3] Joint Funding Council's Libraries Review. *Report (The Follett Report)*. Bristol: HEFCE, 1993. Also available at http://www.ukoln.ac.uk/services/papers/follett/report/

[4] JISC. *Circular 4/94 FIGIT Framework*. Bristol: JISC, 1994. Available at http://www.jisc.ac.uk/pub/c4_94.html

[5] JISC. *Circular 11/95 Electronics Library Programme (eLib): targeted call for new proposals*. Bristol: JISC, 1995. Available at http://www.jisc.ac.uk/pub/c11_95.html

[6] *Ariadne*, http://www.ariadne.ac.uk/

[7] Lawrence Dowler (ed.). *Gateways to Knowledge: The Role of Academic Libraries in Teaching, Learning and Research*. Cambridge, MA: The MIT Press, 1997.

[8] Edinburgh Engineering Virtual Library (EEVL), http://www.eevl.ac.uk/

[9] Organising Medical Networked Information (OMNI), http://omni.ac.uk/

[10] Social Science Information Gateway (SOSIG), http://sosig.ac.uk/

[11] Richard Heseltine. The Future of the Subject Resource Gateways. *Proceedings of the Library Strategy Workshop*. Bristol: JISC, 1998.

[12] Copyright Ownership Protection In Computer Assisted Training (COPICAT), http://www.mari.co.uk/copicat/

[13] Penny Garrod and Ivan Sidgreaves. *Skills for new Information Professionals: the SKIP Project*. London: LITC, 1998. (In press) Also available at http://www.ukoln.ac.uk/services/elib/papers/other/skip/.

[14] Bernard G. Sloan. *Service Perspectives for the Digital Library: Remote Reference Services*. University of Illinois at Urbana-Champaign, 1997. Forthcoming Library Trends 47[2] (Summer 1998). Currently available at http://alexia.lis.uiuc.edu/~sloan/e-ref.html.

[15] *Internet Archaeology*, http://intarch.ac.uk/

[16] Some of this work is reported at http://www.ukoln.ac.uk/services/elib/papers/pa/, and more recent papers will be added to this collection as they are finalised.

[17] Open Journal Framework project, http://journals.ecs.soton.ac.uk/

[18] Digimap, http://digimap.ed.ac.uk:8081/

[19] HELIX, http://severn.dmu.ac.uk/elib/helix/

Web2PDF
converted by Web2PDFConvert.com

[20] PATRON, http://www.lib.surrey.ac.uk/eLib/Patron/Patron.htm

[21] MODELS, http://www.ukoln.ac.uk/dlis/models/

[22] COPAC, http://copac.ac.uk/copac/

[23] Jon Knight, Dan Brickley, Martin Hamilton, John Kirriemuir, Susan Welsh. Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach. *D-Lib Magazine*, January 1998, http://www.dlib.org/dlib/january98/01kirriemuir.html

[24] Joint Funding Council's Library Review. *Report of the Group on a National/ Regional Strategy for Library Provision for Researchers (The Anderson Report).* Bath: UKOLN, 1995. Available at http://www.ukoln.ac.uk/services/elib/papers/other/anderson/

[25] CEDARS, http://www.curl.ac.uk/cedarsinfo.shtml

[26] Lynda Agili, David Corben, Leah Halliday, Alison McGilvray, Lisa McRory, Caroline Moore, Carolyn Rowlinson. *Impact of on-demand publishing and electronic resource banks on teaching, students and libraries*. London: LITC, 1998. (In press.)

[27] Project ACORN Final Report, Appendix 9 Final Copyright Permissions Report, available from Pilkington Library, Loughborough University.

[28] Ian Mowat. Giant Leaps or Small Steps. *Proceedings of the Library Strategy Workshop*. Bristol: JISC, 1998

---

*) Chris Rusbridge. Towards the Hybrid Library // D-Lib Magazine - July/August 1998. - (Engl.) - URL: http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html

**) Печатается с любезного разрешения автора.

© *Rusbridge C., 1998*

HotLog 418223 +92

Web2PDF
converted by Web2PDFConvert.com