

5

What is a library? Digital and hybrid libraries

The cost of data storage is coming down rapidly . . . so for the money we can store more data. But all that we will then have is cheaper garbage. (Shank, 1983)

Introduction

Any study of libraries and librarianship at the present time will almost inevitably be dominated by concerns with the impacts of technology in general and ICTs in particular. In the preceding chapters it is clear that whether the view taken is sectoral, theoretical or professional, ICTs have become of supreme importance to the library and information management profession. In later chapters the technological underpinning of future libraries will be examined in some detail, but first it is useful to examine some of the approaches being taken to designing new types of library on technological foundations. While terminology is unsettled – with digital library, electronic library and virtual library vying for prominence – the first of these is preferred here because it appears to encompass the widest scope. In a final section of this chapter, the concept of the *hybrid library*, a kind of half-way house between the traditional and the digital, is examined.

Perspectives of digital library developers and commentators tend to be driven by their backgrounds in librarianship, information science, computer science, systems development, education or other disciplines. The discussion focuses most often on the technology, but also encompasses the development of new economic models, legal frameworks, user perspectives and social impacts. A small selection of the more theoretical approaches are discussed here in order to shed light both on the concept itself and on what it has to contribute to the development and shaping of future libraries. A lengthier discussion of digital library research can be found in Brophy (1999), while a survey by Chowdhury and Chowdhury (1999) and an analysis of underlying concepts by Bawden and Rowlands (1999a, 1999b) and Rowlands and Bawden (1999) have also appeared recently.

Knowledge mediators

A study by Owen and Wiercx of NBBi in the Netherlands, undertaken as a supporting study within the European Commission's Telematics for Libraries Programme, developed what were called 'knowledge models for networked library services': 'libraries, as a component of the information chain, act as a link between knowledge sources and users' (Owen and Wiercx, 1996). The authors suggested that they can therefore best be understood as *knowledge mediators*: 'the process whereby libraries provide users with insight into the existing body of knowledge and assist users in acquiring resources referring to or containing such knowledge'.

In the context of networked information sources, libraries will no longer be 'restricted to the catalogue' but will make use of a wider range of tools in fulfilling this function. Three fundamental functions of the digital library were defined in this work:

- Making available various types of knowledge resources
- Providing resource discovery mechanisms which allow users to identify relevant or requested resources and their locations
- Providing mechanisms for delivery of specific resources to the user; delivery includes both obtaining a resource when it is not already available in the library, and passing it on to the user in a suitable way.

The process of acquiring information, or *resource discovery* in the authors' terminology, was shown in this model to involve three processes: *resource location* through which the user discovers the existence of an item and a location for it; *resource provision*, involving the provision or acquisition of the resource to the library; and, *resource delivery* which involves such processes as on-site consultation, photocopying and downloading. The resource discovery system itself contains document descriptions, pointers to locations, searching or browsing mechanisms and, sometimes, abstracts or full text. Owen and Wiercx pointed out that the choice of resource discovery mechanism restricts the set of resources that the user can access – most obviously, the library catalogue usually restricts the set to those publications held by that library.

Owen and Wiercx developed their model further by examining parallels with traditional library functions (such as user support) and then developed a series of 'application models' to assist libraries to incorporate networked resources alongside traditional services. One of the main issues for libraries that they identified was the management of user expectations, which they recognized as being raised by the ease with which networked resources could be accessed directly: 'This creates an unfavourable situation for both sides: the

user has to cope without the support and quality mechanisms of the modern networked library, and the library loses part of its user base' – this echoes earlier theoretical work by Buckland which is considered in the next chapter. Other issues include staff awareness and skills, and the availability of adequate financial resources needed to migrate to networked services.

Moving to distributed environments for library services

The eLib *Moving to Distributed Environments for Library Services* or MODELS study has developed a 'MODELS Information Architecture' (MIA) as a way of describing systems that unify access to service providers through an intermediary while providing flexibility of data presentation to the user – and where the 'user' may in fact be software that processes, analyses and possibly re-uses results in some way on behalf of the human end-user(s). The MIA has been described as both 'a conceptual, heuristic tool for the library community' and 'a tool to assist developers as they think about future systems work' (Dempsey, Russell and Murray, 1999).

In brief, the MIA characterizes libraries as 'brokers' that both hide the complexities and differences of underlying resource discovery services from their users and facilitate data flows so as to enable processes to be automated. The broker is a *trading place*, 'where service requests and service providers come together'. A generalized description of such services includes the provision of:

- user access, including the presentation of an 'information landscape' and support for user profiles
- an applications framework consisting of software and data needed to manage the services, passing data between functions
- distributed service interfaces, which determine and control how requests are presented to underlying services
- access control, including the authentication of users and commercial transactions such as payments.

New or revised underlying services are handled by the applications framework and distributed service interfaces without requiring changes to the user access layer, since the service must operate in an environment of rapidly changing target services. Thus adding a new service should be cost-effective (and the library both scalable and sustainable) since it does not require a new user interface to be built.

The applications framework can be defined in terms of four key functions (similar to Owen and Wiercx's analysis described above): resource discovery, location, request and delivery. These require descriptions of the underlying ser-

vices, including collection descriptions and interface descriptions (ie what information is available and the protocols needed to access it), and profiles of users that enable the system to determine access rights, preferences and so on. The MIA provides an underlying theoretical framework for the UK's Distributed National Electronic Resource (DNER) – described in Chapter 9 – and is of enormous significance for both digital and hybrid library developments.

The architecture of the digital library

A series of papers produced as part of the US Computer Science Technical Reports Project, associated with the Library of Congress's National Digital Library Program (NDLP), and its precursor programme, American Memory, have suggested a basic architecture for digital library services (Kahn and Wilenski, 1995, W Y Arms, 1995; C R Arms, 1996; W Y Arms, Blanchi and Overly, 1997 – see also <http://memory.loc.gov/ammem/ftpfiles.html>). In essence the digital library is seen as constituted from digital objects – as described briefly in Chapter 1 – 'handles' or unique object identifiers, and repositories. The architecture has been implemented in developing the digitization of the Library of Congress's historic collections.

The handle system is central to the implementation of the concept in real-life systems, and has developed over recent years into an internationally significant service to the networked information community as a whole. Its essence lies in the development of persistent identifiers for digital objects that are independent of the current location or other 'state' of the resource, but can be resolved into a location by a resolution service. These issues are examined more extensively in Chapter 8 below. The handle system has been developed by the Corporation for National Research Initiatives (CNRI), which is also the publisher of the influential *D-Lib Magazine*, a monthly publication on digital library research and development.

An interesting aspect of this work is the way in which different types of object are placed in explicit categories, such as web pages or digitized audio, and are then subject to specific category rules. Arms, Blanchi and Overly (1997) describe this approach as follows: 'the rules describe the digital objects that are used to represent material in the library, how each is represented, how they are grouped as a set of digital objects, the internal structure of each digital object, the associated metadata, and the conventions for naming the digital objects'. With the categorization and rules in place, a query system – whether a user interface or another intermediary system – can determine how to access, store and display any object, including correctly interpreting its relationship to other objects. It is explicit that the digital object delivered to the user is conceptually

not the same as the stored object but may be delivered by processing. Arms (1995) suggests that eight principles should inform the development of digital libraries:

- 1 The technical framework exists within a legal and social framework
- 2 Understanding of digital library concepts is hampered by terminology
- 3 The underlying architecture should be separate from the content stored in the library
- 4 Names and identifiers are the basic building block for the digital library
- 5 Digital library objects are more than collections of bits
- 6 The digital library object that is used is different from the stored object
- 7 Repositories must look after the information they hold
- 8 Users want intellectual works, not digital objects.

Reference architectures

Related to the National Digital Library Project described above, work at Cornell University has explored issues related to the development of distributed digital libraries. The Cornell Reference Architecture for Distributed Digital Libraries (CRADDL) defines a number of core services of the digital library (Lagoze and Fielding, 1998). These include:

- a *repository service*, where digital objects can be deposited and stored, and to which the digital library provides access
- a *naming service*, which ensures that each digital object has a globally unique name and that this can be resolved into at least one physical location
- an *index service*, which enables sets of digital objects to be described and queries to be resolved into result sets containing unique names
- a *collection service*, which performs selection against defined criteria, tailored and specialized catalogues or other resource discovery aids, and administrative processes.

Not surprisingly, as the teams have worked together closely, there are many similarities between this and the approaches described above, but there are additional emphases that are important to the development of more generic models.

CRADDL emphasizes that without effective *naming* conventions and systems effective retrieval is impossible: in the CRADDL model this element consists of the application of Uniform Resource Names (URNs) to digital objects, coupled with the provision of a *naming service* that links the URN to physical

locations – implemented using the handle system described in the last section, ie CNRI's *Handle System*® (<http://www.handle.net>). It thus takes on the traditional library's catalogue-as-finding-tool role, but within a distributed environment. Again it is important to note that the architecture separates *naming* from *location*, which tend to be concatenated in the traditional library catalogue.

CRADDL's definition of *index services* provides the mechanism to enable users to perform structured queries and receive result sets: essentially they are collections of metadata defined according to a criterion that may relate to a physical or virtual 'repository'.

CRADDL also defines the *collection* as 'a set of criteria for selecting resources from the broader information space'. This is a very important approach because it removes the concept of 'collection' from that of 'physical location' (whether physical artefacts or electronic entities on a particular server) back to the idea of the creation of collections by selection from the universe of information objects available. In the CRADDL model the collection need have no existence beyond the criteria for its selection: items may be selected for the collection dynamically from a wide range of sources as they are needed. Furthermore, the 'collection' may, through a carefully crafted set of criteria, grow or shrink with the development of its subject. Lagoze and Fielding suggest that this provides three key advantages: location and administrative independence, dynamic information object membership and extensibility.

The control zone

Observing that one of the key features of the traditional library is the careful selection of stock to add to the collection, Atkinson (1996) argues that 'it is time . . . for the academic community to begin work on the creation and management of a single, virtual, distributed, international digital library, a library that has [conceptual, virtual] boundaries, that defines its service operationally on the basis of the opposition between what is inside and outside those boundaries, and that bases that service on the traditional social ethic that has motivated all library operations in modern times'. He argues that the role of the library is to control a systematically selected sub-set of published information, and then to ensure that 'such a subset remains stable and accessible over time'. To achieve this he defines five characteristics of such a 'control zone':

- *Core definition* – materials selected by the library on the basis of the value of their content and against which material outside the core may be evaluated.

- *Particularization* – a combination of the *level* of a work and its *significance*, as attested by its being accepted by specialists and scholars in the field – in essence a form of peer review.
- *Maintenance* – the library accepts responsibility for maintenance of the integrity of all objects admitted to the zone, both in terms of physical integrity and authentication – they remain as they were when originally ‘acquired’.
- *Certification* – moving an information object into the control zone becomes the equivalent of publishing it, so university presses and university libraries are amalgamated in the digital environment.
- *Standardization and coordination* – the use of agreed protocols and standards.

This analysis is interesting in addressing one of the major issues of the digital environment, namely how the long-term authority and preservation of digital objects is to be achieved. It suggests that merging publisher and library functions, and undertaking an explicit ‘maintenance’ function, should be seen as important digital library functions. Furthermore it offers a possible framework for distinguishing ‘useful’ digital content from the dross of which users so often complain by offering what is in effect a ‘kitemark’ for each individual information object.

Conceptual frameworks

Most of the above studies consider the digital library from a technological perspective. In a series of recent reports (Bawden and Rowlands, 1999a; Bawden and Rowlands, 1999b; Rowlands and Bawden, 1999) a conceptual framework is put forward based on Yates’s work-oriented perspective (Yates, 1989) and a model adapted from the work of Reid (1999). An important and useful aspect of this model is its explicit identification of three ‘domains’ within which understanding of digital libraries can be developed:

- the *social domain*, which considers human factors including such issues as information skills and literacy, social impacts on the information transfer chain and information law and policy
- the *informational domain*, including knowledge organization and discovery (including such issues as description) and implications from a documentary perspective for the information transfer chain; a particular emphasis within this domain is on the role of metadata
- the *systems domain*, including human–computer interaction, software agents and systems architectures; issues within this domain that Bawden and Rowlands isolate as of special significance are scalability and interoper-

ability – again these are considered in greater detail in Chapter 8, where sustainability and personalization are added as key issues.

This work includes an attempt to place the concept of the digital library in the broader context of ‘library’ development, using a diagram with analogue–digital and localized–distributed axes. The authors are led to suggest the use of the term ‘complex library’, based on work by Crawford (1999), which they prefer to ‘hybrid library’ which they feel suggests ‘an awkward transitional phase’. Given the widespread use of the latter terminology, however, there may be some doubt as to whether their preferred term will catch on.

Summary of digital library concepts

There are very many approaches to modelling the digital library, but a number of common features emerge from the research and development that has taken place in recent years. These are:

- recognition that in the real world the information of interest to digital library users is to be found in a range of heterogeneous databases and collections, physically distributed but connected by electronic networks and containing objects of many different types
- distributed ownership and rights, including complex intellectual property rights
- a need to provide organization, provenance and authority for items and collections
- a wide range of users with a variety of client systems operating within a broad selection of environments and pursuing many different purposes
- a business need to control access to resources although the models, and especially the economic models, are far from clear
- a role, variously defined, for a broker or other intermediary which connects the users to the resources of interest to them
- a range of standards and protocols, for describing resources, for encoding them and for delivering them, including for searching and retrieval; many existing systems cannot interoperate effectively and there is thus a pressing need to agree open standards.

The hybrid library

A pragmatic view of the future of libraries would suggest that it is likely that most libraries will for the foreseeable future – which after all is not all that long – base their services on a mix of physical objects (books, paper-based journals,

videotapes and the like) and electronic 'stuff' (web pages, remote data services, CD-ROMs etc). For this reason, approaches that stress the management and delivery of 'hybrid library' services would appear to be the most useful. Crawford (1999), who argues somewhat controversially that 'the glory days of the all-digital brigade seem to be in the past', refers to 'ever-shifting complex combinations of digital and analog collections' which, as noted above, he terms 'complex libraries'.

The most widely used definition of the hybrid library is that provided by Rusbridge (1998):

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 . . . kinds of resources . . . using different technologies from the digital library world, and across different media.

Brophy and Fisher (1998) comment:

It follows that most users will continue to be offered a mix of formats via a mix of delivery systems. The challenge for library managers is to create integrated services which provide a 'seamless' service to the user. The user should be able to access services through consistent interfaces which provide compatible features (so that, for example, the user does not have to adjust her search strategy and syntax each time she wishes to use a different source). The hybrid library should not, however, be homogeneous: it should be able to adjust its services to the needs and rights of each user – for example by 'remembering' previous search strategies, by storing details of the user groups to which the individual belongs and by 'knowing' the user's willingness to pay for premium services.

Work on hybrid libraries has also embraced the wider resources available, including the need to bring together descriptions of items in collections situated either in geographical proximity or with common subject strengths. The Electronic Libraries Programme's 'clump' projects were designed to explore this territory (Brack, 1999):

The clump projects are based on the need to aid discovery of, and increase access to, the vast scholarly bibliographic resources available to the Higher Education community, as proposed in the Anderson Report [Higher Education Funding Council for England, 1995b], which examined library provision for researchers in UK higher education. In the current financial climate it is impossible for individual institutions to acquire adequate research collections within their libraries, and although union

catalogues have been available as central resources for many years, they have not necessarily taken advantage of network technology. Since the beginning of 1998 the clumps have been investigating the use of such technology in opening up access to these bibliographic resources.

A 'clump' is a term that was coined by the 3rd MODELS (Moving to Distributed Environments for Library Services) Workshop [see <http://www.ukoln.ac.uk/services/elib/projects/models/>] and is used to describe an aggregation of library catalogues. The clump may be 'physical' – in traditional terminology a union catalogue, such as COPAC (the CURL OPAC) or the BLCMP union catalogue – or it may be 'virtual', being created at the time of searching.

The hybrid library concept is important precisely because it stresses that in the real world both traditional and digital resources can and must be managed together. The UK parliamentary investigation of public libraries, which reported in May 2000, put it this way (House of Commons, 2000):

There is a continuing tendency in some analyses of trends in library services to stress the competition between the book and new technology. This is a false antithesis. Their development must be complementary not competitive. We are convinced that the book will survive for the foreseeable future. It will be supplemented, not superseded. The challenge for the library sector is to ensure that the development of information technology in libraries broadens library services and does not take place at the expense of the book.

Conclusion

There are a considerable number of concepts being explored in digital and hybrid library research and development. In essence, though, the focus is on distributed and local collections of information objects – in the hybrid library including analogue as well as digital objects – and on ways of identifying objects of interest to a user and arranging for the user to access them. Although the concentration has tended to be on the objects themselves, on their description and on their organization, many of the research teams have pointed to the wider social contexts and the need to understand the human as well as the technological domain. This broadening of perspective, although not yet greatly in evidence, will undoubtedly become more prominent in the future.

In the meantime it is useful to note the close parallels between the digital library models and the key characteristics of modern libraries that emerged from earlier discussions of traditional library approaches. The twin concern with information and with its use is common to all these models, while the modelling of processes such as resource discovery, resource location and

resource delivery formalizes operational concerns that have concerned the designers of library services from earliest times.