

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

## A UTOPIAN PLACE OF CRITICISM? BROKERING ACCESS TO NETWORK INFORMATION

LORCAN DEMPSEY *and* ROSEMARY RUSSELL  
{l.dempsey; r.russell}@ukoln.ac.uk

*UK Office for Library and Information Networking, University of Bath  
Bath BA2 7AY*

ROBIN MURRAY  
rmurray@fdgroup.co.uk

*Fretwell-Downing Informatics, Ecclesall Rd  
Sheffield S11 7AE.*

The management of autonomous, heterogeneous network resources and services provides new challenges which libraries are now addressing. This paper outlines an approach based on the construction of broker services which mediate access to resources. It outlines a framework – the MODELS Information Architecture – for thinking about the components of broker services and their logical arrangement. It describes several development projects and services which show how brokers are developing. It uses examples drawn from the serials environment to describe some of the issues.

*Technologists understand that they must build more stable and unobtrusive media. They must establish more coherent contexts into which the technology may disappear.*

*Malcolm McCullough. Abstracting craft: the practised digital hand [1].*

### 1. INTRODUCTION

#### 1.1 Libraries

Alberto Manguel has a chapter on libraries and librarians in his recent *A history of reading* [2]. He calls them ‘ordainers of the universe’, an epithet used, he tells us, by the Sumerians. He dwells on the efforts of Callimachus to ordain the order of books at The Library of Alexandria, and notes that:

According to the French critic Christian Jacob, Callimachus’s library was the first example of a ‘utopian place of criticism, in which the texts can be compared, opened side by side’ [3]. With Callimachus, the library became an organized reading-space.

For many years, libraries have refined the techniques developed by Callimachus. They have created physical places and intellectual reading-spaces which connect

their users to resources in useful ways. They have evolved well-understood internal practices and procedures for management, and predictable ways of presenting services for their users. In this they have been assisted by the evolving technologies of print and publishing, as well as by internal library technologies. Books and journals come in accepted formats, which support some consistency of treatment and arrangement, which allow the advance construction of shelves and processing equipment, the assignment of space, and so on. They only exceptionally require separate introduction or special treatment: these particular technologies have become unobtrusive, experience of them submerged in the practice of reading. Libraries have also developed an intellectual apparatus for the organisation of their resources in various ways and with various goals (to collocate works by authors, for example). The physical and intellectual apparatuses are meshed together in different ways.

Libraries are now faced with the challenge of recreating this role in a new kind of space, the space of flows supported by the worldwide span of networks. Writing about access to networked information, Richard Heseltine remarked:

What I am more concerned about is the need to make the desktop working environment of the end-user simple and easy to operate. End-users are being confronted now by a multiplicity of systems and services: for obtaining information; for communicating; for taking delivery of documents, and for producing documents. We need to have much better models of how all these services should fit together from the point of view of the end-user. What are the key standards? What are the most effective means of presenting services? This is not just a matter of user interfaces but of the means of bringing everything together in a real working environment [4].

Heseltine's concern is echoed in a recent UK Higher Education policy document [5] which lays out a view of how national information services should develop alongside local provision as part of a Distributed National Electronic Resource (DNER):

Integration is the key, allowing the user to move more easily between different information functions; more easily across all services (. . .); and to use from the desktop, the emerging tools for exploiting networked information, for more intelligent and standardised searching and retrieval, for locating material, requesting and receiving it, and for making appropriate use of all forms in further analysis and research.

What is being suggested in each case is an 'organised space', in which resources may be used 'side-by-side' in a 'real working environment'. However, this environment is not limited by the PC or the local library, it is an environment which may reach out in space and time: in space because the network spans the globe; in time, because users may be supported more persistently – by some combinations of personal profiles, agents, alerting services, or configurable, adaptive environments. It is also a space in which there are new divisions of labour in the learning and information domains (as for example in document supply, where publishers, libraries and aggregators are realigning the pattern of delivery), and new forms of user behaviour and expectation (as for example, where communication technologies are reaching into writing and learning environments).

How such 'integration' will be constructed is a research and development challenge. Current digital environments are in early stages of development. What integration there is tends to be rather shallow, typically at the user access level where the web has become the approach of choice. Pages of links, perhaps a database of resource descriptions [6], provide a level of integration at the discovery stage, but resources themselves are differently presented, accessed, structured. They remain individual, unconnected opportunities. A user may have to interact with quite different information systems to carry out a full search, for example. Furthermore, individual tasks rather than end-to-end processes are automated: the emphasis has been on getting individual systems to work – the ILL system, access to BIDS, the CD-ROM network – rather than seeing them as part of a wider information environment which needs to be linked in various ways. These systems do not connect to each other. Information flows intermittently through supply chains, which are fragmented and incomplete.

Part of the challenge is that 'standardization efforts are lagging behind the development of digital library services' [7]. The standards infrastructure is not yet achieved enough to make the management and use of electronic resources routinely predictable in the way that the management and use of print resources is. We do not yet have a 'coherent context' into which the technology can disappear: difference and distraction are very much on the surface, data do not cross system boundaries. Libraries have evolved ways of combining components to provide services. In the digital environment, components do not work well together.

This is a serious issue, and, we suggest, the most significant barrier to pervasive deployment of networked information systems as part of users' normal working practice. It is this development deficit that is driving the current interest in digital library research.

### *1.2 Access to serials – an example of some of the issues*

It would be useful to make some of the 'integration' issues more concrete with specific examples. The serials literature presents some particular challenges [8], which highlight more general issues, and we use it throughout this article to illustrate issues and trends. Take a simple example, which we have introduced elsewhere [9]. A project group wishes to discover journal articles and books about Roman Bath. In a well-stocked library, they can scan the shelves. Say they want to do a more thorough 'discovery' of material. They can look in the catalogue. They can look in databases on CD-ROM. They might have access to some remote databases over the Internet. But each of these is delivered through a separate interface, they may have to move between machines, they may have to print out or write down results. They also have to know which databases to look in: in fact there may be very relevant resources which they will fail to use if they are not directed to them by staff. Figure 1 shows some of the resources which may typically be available to users.

Once they have discovered a selection of materials, they may have to find out where they are. Typically, they will have to return to the catalogue and redo searches for the desired titles. Say they are in a library which has an arrangement for reciprocal borrowing with several neighbouring institutions: they will have to redo searches for unfound titles in those libraries' catalogues. They might bring

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

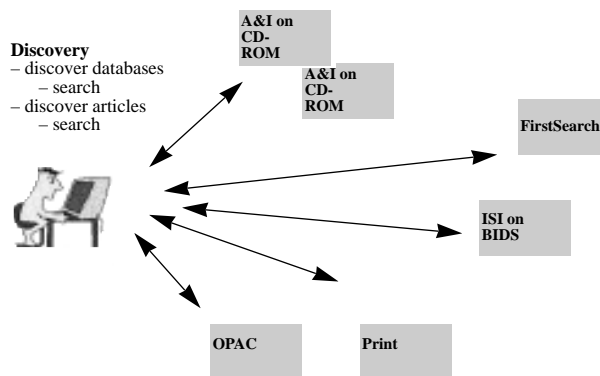


Figure 1. *Discovery – heterogeneous resources*

other items to the interlibrary loan department, where they may have to write down the details again. Then the ILL staff may repeat some of the operations already carried out by the users. Requests may then be sent for materials. Bibliographic details may be rekeyed for transmission.

Within the MODELS project (described further below), it has been found that the type of behaviour that we have just described can be represented in terms of a limited number of logical functions: 'discover', 'locate', 'request', and 'deliver'. These map well onto user behaviour and the services required to support it. Clearly, this list could be extended, to include 'use' for example, but we have focused our attention on these four, which cover a large number of cross sectoral service scenarios and business requirements. They represent services provided in current library environments. (One of the issues which libraries need to address in this new environment, and it is part of the rationale for the development of 'learning resource centres', is support for use, analysis and reuse of digital resources, but this is not treated further here.)

We have described an environment in which there is a variety of boundaries – between functions, between users and library – which are not interconnected by systems, and across which data do not flow. The 'connections' are made by human effort: users repeat operations on different systems, consult staff, transcribe details. This wastes time and imposes barriers to full use. Some of the services which may have to be interacted with are shown in Figure 2. The current situation has some important characteristics:

- It is composed of heterogeneous resources which are growing in volume and variety. Services have different access characteristics, use different formats, and so on.
- Services are autonomously managed; they have developed independently, responsive to different service and business goals.
- Services represent different aggregations of function. For example, a union catalogue allows people to discover (and locate) journal articles. An integrated service from the British Library or BIDS allows a user to discover, locate, and request documents and have them delivered.

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

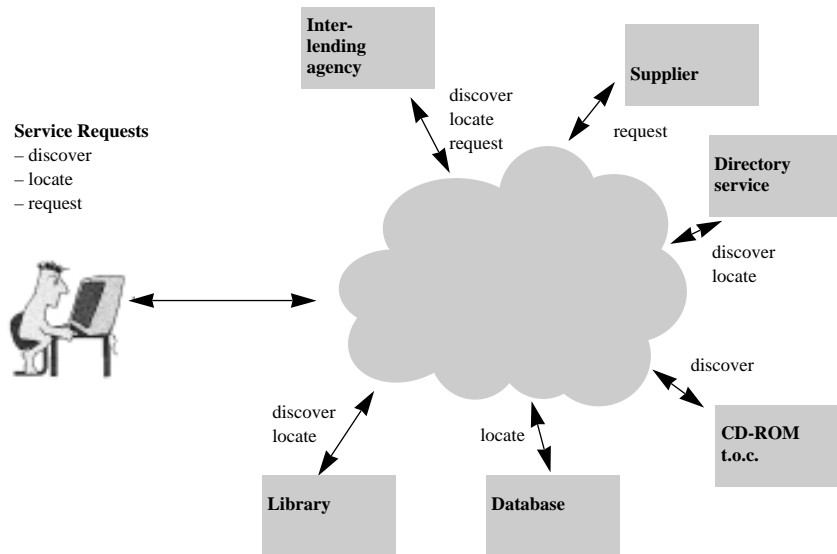


Figure 2. *Heterogeneous service providers*

An abstracting and indexing service allows users to discover the existence of documents. And so on.

- No service will meet all needs. Whilst some organisations now offer article services which include discover, locate, request and deliver facilities, they are still just components within this potentially distributed document supply service since no server will meet all coverage or quality of service criteria. This is not to say that 'one stop shops' are not locally useful, or an inappropriate response to the problems we have identified, but they provide inevitably partial solutions.

There are clear integration challenges: service autonomy, heterogeneity, and geographic distribution will remain the rule. This is the case with the serials literature. It is very much more so when we consider the wider range of resources that are of potential interest to a user.

The current situation, as sketched above, is unsustainable for two broad reasons: it is wasteful of users' time and energies, and it poses growing management problems for libraries which need to move to routine ways of managing hybrid collections. Current electronic information services are underdeveloped: they do not allow easy integration with users' working environments, and they present case-by-case problems to libraries. This leads us to argue that a growing factor in the assessment of individual services and resources will be the ease with which they can become part of a wider environment of use. This will be a managed environment, which delivers added value to users and which develops procedures and practices which support economies of operation.

In terms of Figure 2, the challenge is to construct the cloud in the middle. This is 'middleware', an additional service layer, which shields the user from

complexity, heterogeneity and multiple mechanical actions. It provides a higher level interface, creating a federated resource from underlying heterogeneity and mediating access to it. In this article, we use the term 'broker' for this cloud. To create this service requires high levels of inter-operability, both between servers providing the same service (e.g. two search services) and between different services (e.g. between locate and request).

We refer to these two types of interoperability as 'intra-function' and 'inter-function' respectively. Intra-function interoperability suggests the ability, for example, to search across several resources, or to have a request relayed to several different request systems. Inter-function interoperability is where these functions communicate with each other by passing data between them. There is a high level of inter-function integration in a simple web environment, which is what makes it such an attractive tool to users, who can discover, locate, request, and have delivered materials with a few clicks. The environment we have discussed is a more heterogeneous one, in technical and service terms. There is currently much greater 'friction' in the system, as intra- and inter-function integration are lacking.

### *1.3 The MODELS project and the MODELS Information Architecture (MIA)*

Interoperation benefits from standard, predictable application interfaces (for search and retrieve, for requesting, and so on, which will allow data to flow between applications and between applications and user systems), and on meta-data (data which support operations on resources, and which help users and applications behave sensibly). More recently there has been significant renewed interest in identifiers [10, 11], and in issues of authentication and authorisation [12, 13]. However, we are always likely to be some way from a completely standards-based environment, and customised links also have to be created. Together, these form the infrastructure or 'glue' that will allow the construction of broker services. The focus has been on individual technologies; attention is now turning to how they work together to support useful services. This in turn has created significant interest and development work to provide support for communicating software components. CORBA (Common Object Request Broker Architecture) and DCOM (Distributed Component Object Model) have emerged as principal specifications for such environments.

Our approach has been somewhat influenced by such developments, but is more concerned with modelling aspects of the information environment in which applications will be built. Some advantages of such an approach are that:

- it encourages thinking about support for business processes rather than about particular technologies or systems;
- it helps align thinking about priorities;
- it provides a common vocabulary and frame of reference which support communication and discussion;
- it identifies areas where development or consensus are lacking.

The MODELS Information Architecture (MIA) aims to provide a framework within which such issues can be addressed. The MIA is aligned with wider work which sees the development of 'middleware' or 'broker' services as a central part

January 1999

BROKER SERVICES

of how the information environment will develop. It is concerned with the types of function such 'broker' services need to provide as they help project a unified service over a distributed, heterogeneous set of network services. It has a dual focus: as a conceptual heuristic tool for the library community which helps clarify thinking and acts as a lever for development, and as a tool to assist developers as they think about future systems work. The main emphasis has been on the former aspect. The MIA investigates the functional components of viable digital information environments and arranges them in a logical architecture: it does not yet specify how components will be implemented, or concrete interfaces.

MODELS (Moving to Distributed Environments for Library Services) [14] is a UKOLN initiative with additional support from the Electronic Libraries Programme, with some support also from the British Library. It has progressed through a series of workshops, background research and technical consultancy to work towards consensus on these issues. It has initiated several studies and has influenced policy and emerging services in the UK.

This paper further develops a discussion of the current information environments, introduces the MIA, and describes several current development initiatives which show how MIA-like service environments are emerging. Throughout the MODELS project, it has become clear that many of the issues addressed are generic across several information domains: they are not library-specific [15]. While the focus of this article is on library issues, discussion of current development work points to a wider context.

## 2. RELATED WORK

MODELS has focused on the 'hybrid' library environment, characterised by the aim of constructing unified information spaces across existing, place-based library services, and emerging digital information spaces. These issues have had to be addressed in the various applied research and technical development programmes: the EU Libraries Programme (a colloquial term for activities managed by the Libraries Unit of DGXIII through Framework Programmes III and IV) [16], the Electronic Libraries Programme of the Joint Information Systems Committee of the Higher Education Funding Councils in the UK (eLib) [17], and the US Digital Libraries Initiative [18]. We trace the emergence of distributed library services in Europe elsewhere, with some focus on developments within the European Libraries Programme [19]. Work here has tended to be implementation- and demonstrator-oriented, and to have proceeded with a project-based focus, although there has been some investigation of library application models in the light of changing network environments [20]. There has been collaboration on a programme-wide basis, and further, on particular technologies, but rather less on broader architectural issues. The approach developed in this paper provides a good framework for understanding and relating many of these developments. This is especially so as the initial emphasis on the construction of component building blocks has now switched to the construction of production services based on integrated systems and products.

A similar situation exists within eLib. In this case, participants are working with MODELS as a tool in developing service scenarios, and the 'clumps' initiative

flows from it. A 'clump' is a term used for services based on unified access to physically distributed catalogues, several of which are being funded under eLib [21]. The integration ideas underlying the DNER have been influenced by MODELS and we will continue to work with policy and service actors in this area.

The Digital Libraries Initiative in the US has had a different remit to either of the above programmes. It has been rather more research-led, with a computer science focus. It has also been rather more 'digital' than 'hybrid'. This has meant that there has been a more explicit architectural focus. In particular, the Stanford Digital Library Initiative has a specific emphasis on interoperability, and is pursuing a suite of related investigations within a distributed object environment, based on CORBA (Common Object Request Broker Architecture). Project collaborators note [22]:

In the Stanford Digital Library project, we view long-term digital library systems as collections of widely distributed autonomously maintained services. While searching services are valuable, they are not the only kind of service in the digital library of the future. Remotely usable information processing facilities are also important digital library services. These services provide support for activities such as document summarization, indexing, collaborative annotation, format conversion, bibliography maintenance, and copyright clearance.

Our project has focused on developing an infrastructure in which these disparate services can communicate and interoperate with one another. Our digital library testbed is providing an infrastructure that affords interoperability among these heterogeneous, autonomous components, much like a hardware bus enables interaction between disparate hardware elements. We call this infrastructure the *InfoBus*.

In building the InfoBus, we needed to provide services for finding resources likely to satisfy a given query, for formulating queries that are appropriate for multiple sources, for translating queries, and for making sense of query results.

The Stanford project has a research and implementation focus. In developing a distributed object model, it is consciously in advance of current implementation and experience.

Paepcke and colleagues from Stanford have recently presented an interesting discussion of digital library interoperability issues [23]. They identify a variety of modes of interoperability. We briefly consider three here to locate our discussion in a wider context. The first is 'external mediation', where what they call the 'interoperability machinery' is located outside the participating local systems to mediate between components. This machinery translates formats and modes of interaction, and where systems are standards based such mediation is facilitated: 'For example, in an external mediation system providing interoperation for highly autonomous search components a single mediation facility covers all Z39.50 resources at once'. A second is 'specification-based interaction', where the goal is to 'describe each component's requirements, assumptions and services so the components can interact with one another after inspecting and reasoning about each others' specifications.' A third they call 'mobile functionality', which



All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

accomplishes interoperability by 'exchanging code that "does the right thing" to communicate successfully among components'. The latter two approaches will become more common in a distributed object environment. Discussion in this article leans towards the 'external mediation' mode, not because the MIA assumes it, but because most of the development work we consider follows that mode. The discussion shows the growing importance of metadata, to describe services, applications and other components in a distributed environment.

MODELS has been developed within a library context, but it is clear that many issues are generic across other domains which involve managed access to, and use of, network resources. There are points of contact or comparison with ongoing work in other domains: other curatorial traditions (e.g. archives, museums [24]), geospatial data [25], humanities computing [26], distributed indexing of Internet resources [27], electronic product catalogues [28], and other areas.

### 3. ASPECTS OF THE NETWORK INFORMATION ENVIRONMENT

In this section we put the above discussion in a wider context, by linking it to a more general account of the emerging technical environment, by considering in more detail the issues of providing distributed article discovery and supply, by briefly considering the wider information environment, and by introducing some of the services a broker might provide. We continue to consider the emergence of broker-based services alongside an examination of access to the serials literature.

#### 3.1 *The technical environment*

Consider the progressive development of network information systems as presented in Figure 3 and Table 1. We suggest three layers: user layer (U), organisation layer (O), and a layer of resources and services (R/S). In a first phase, monolithic applications are accessed by separate interfaces, often indeed, by separate terminals. By monolithic, we mean that data and software are integrated and that the user communicates via terminal access; there is no flow of data between applications: there is no integration at that level. In this phase, there is little support at user or organisation layer. In a second phase, which is the current dominant phase, the web

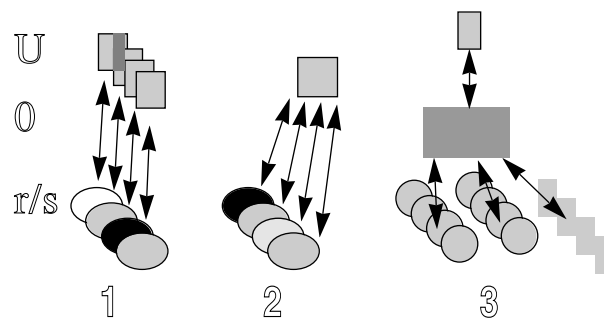


Figure 3. *Multiplying resources*

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

Table 1. *Three phases*

	<i>User presentation</i>	<i>Organisation</i>	<i>Resources and services</i>
Phase 1	Independent	–	Monolithic applications
Phase 2	Unified	Shallow – presentation only	Monolithic applications – data and application
Phase 3	Flexible	Unified	Server based services, may be accessible in several ways.

provides a unifying user layer. Increasingly, the web is the *de facto* desktop user route into resources. The web also provides a very shallow framework for organisation of the underlying resources and services. In phase 2, we see unified web-based presentation to the user over largely monolithic, heterogeneous applications.

In phase 3, the emergent phase, we see the introduction of broker services, which support flexible presentation of a range of underlying services. This layer adds value to the lower layer of resources and services by supporting their presentation as a unified service according to a particular business logic, rather than as a set of separate opportunities. We have discussed the motivation for this development above.

It is now a familiar way of organising network applications, and is supported in a variety of business and application domains.

In the *current* network environment, services are typically made network accessible in one of several ways, which are successively characteristic of behaviour in each of these phases:

- Terminal access to remote services: users login to a monolithic application with its own user interface and request commands. Data may be downloaded using special separate procedures or by some form of screen capture.
- WWW access: this is currently essentially the same architecture as terminal access – the service provider presents its own user interface and request commands, this time through a web form. The gain is that there is some consistency of presentation in a consistently navigable environment.
- Client/Server access: this architecture is significantly different in that the client is intelligent – it has some understanding of the data it is handling, which means it can reuse it to support various operations.

These modes of access correspond approximately to the three phases identified above. In the WWW and terminal examples the client software just responds to user interface directives such as ‘display this text in bold italics’ etc. These services are largely oriented around providing services to human users, who then have to process the results.

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

When the client software understands the data it is handling, they can be reused in various ways, the results can be processed. The client software is responsible for the representation of those data to the user, and is capable of shielding the user from differences between servers. It can be reformatted for display alongside records from other resources. It is also capable of providing increased levels of inter-operability between multiple servers that perform discreet activities within, for example, the discovery and request operations we discussed above. For example a document identifier or citation found from a discovery session on one server can be passed to a 'locate application' which may use it to query against a holdings file.

It is unlikely that users will have application clients on their desktops; typically they will have web browsers. Client functionality will be built into intermediate systems or brokers (i) which access end systems (e) on users' behalf (Figure 4).

This suggests that increasingly services will be provided which may be accessible to human users through the web, or to intermediate systems through some protocol. In the latter case, machine-readable structured data will be returned for reuse in some context; in the former human-readable results will be returned for reading. So for example an OPAC may have a web interface for human access, and a Z39.50 interface for 'clump' access. Brokers talk to machine interfaces.

It should be noted that this outline represents the current web environment. In this, the web is primarily a medium for passive information transfer, with some support for interactivity through CGI scripts and other techniques. We are seeing significant development in the web community which will provide support for the exchange of structured data, for distributed objects, and for a range of security and other services which will significantly enhance the development of web-based information processing applications. The development of XML (extensible markup language) is providing the basis for a variety of applications based on structured data, and together with support for communicating software objects (for example, the Internet Inter-ORB Protocol), will transform the web into a platform which supports distributed transaction-oriented applications [29].

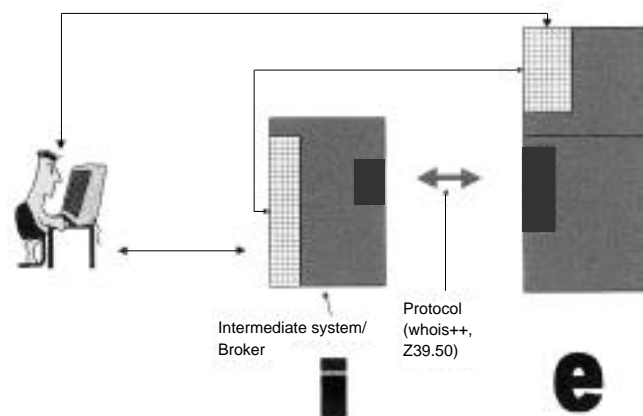


Figure 4. *User interface and server interface*

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

In passing, it should be noted that the emergence of the 'structured web' raises issues for the further development of distributed library and information services and the approaches which have evolved in their support. It will be interesting, for example, to see whether a search protocol is developed in the web environment, and what impact this has on Z39.50 or other specialist approaches which predate current developments. Given the strategic importance of the web, it is likely that many communities will reengineer existing approaches to take advantage of developments there [30].

### 3.2 *The broker*

We have identified some broad-based service goals for brokers:

- to hide the difference and mechanics of interaction from users, whether end users or 'intermediaries', so as to save time and simplify procedures. This includes intra-function integration;
- to facilitate the flow of data between applications so as to automate processes. This includes inter-function integration.

Some examples of such services can be provided from current developments. In each case, an application, or 'external mediator', mediates access to diverse resources and supports data flow. The broker will be designed to support a particular business need. For example, the eLib EDDIS project has developed an application that will provide some of the integration in the serials area that we have suggested is lacking above [31]. It brokers access to discovery, locate, request and delivery services in a distributed document supply context. Agora, to which we return in more detail, is an eLib hybrid libraries project which will extend EDDIS to look at other media and resource types. It will be designed to support access and use of print and electronic content of the 'hybrid library'. The Arts and Humanities Data Service (AHDS) Gateway, also described in more detail below, projects a unified picture of the AHDS based on a federation of five underlying, autonomously managed service providers. The Gateway provides a service which hides the different access mechanisms and data formats of the heterogeneous systems in use at service provider sites, and provides authenticated document requesting services. Another example can be provided by the 'clump' projects, already mentioned; these are funded by eLib to provide virtual union catalogue services across different underlying OPACs. Yet another example is provided by the ROADS cross-searching service. ROADS (Resource Organisation And Discovery in Subject-based Services) is an eLib funded project which is providing a set of tools for the UK subject gateways, databases of descriptions of Internet resources. The cross-searching service provides a query routing and referral service between the autonomously managed subject gateways [32]. Gazebo, developed by the National Centre for Supercomputing Applications, is Internet server software which allows clients to query and browse results from multiple data sources simultaneously [33]. These provide early examples of the type of facility that is likely to become more common across a wider range of resources.

The focus of some of these developments is discovery: to varying degrees, they hide differences and collate the results from several different underlying discovery

January 1999

BROKER SERVICES

systems. Some go beyond this to address several functional areas and allow data to flow between them. For example, EDDIS passes data about selected articles from a discovery to a 'locate' function where they may be matched against some holdings data; data may then be passed to a request function, where they form the basis of a request message. It hides the difference between different discovery systems and different request systems. There have been 'gateways' and other intermediary services in the past. What characterises these initiatives is that they are based on standard machine interfaces which allows them to interact with arbitrary resources which support the same interfaces. They are a move towards a 'plug and play' environment, which is seen to be essential in the type of distributed, indefinitely large resource space we have discussed. They currently largely operate with data that are well structured, or amenable to being so structured. While they are largely 'bespoke' applications with standard interfaces to the outside world, we will gradually see more fully distributed applications emerge based on distributed components specialised by function and assembled as required.

MODELS has generalised the services provided by such 'brokers' in the following way:

- *User access – managing interaction with the user.* 'Information landscape' is a term that was introduced within the MODELS context to describe the presentation of rich information environments to users. The landscape will project the underlying business process in a user-oriented way. The 'information landscape' will describe the resources available, provide navigation and selection support, and may be configured to reflect particular styles, policies and collections. A minimal landscape may have links to resources; richer landscapes will gradually emerge which are constructed on the basis of forward knowledge of available services, user profiles, and other data. In a hybrid environment, it will also be important to provide routes into the print and the digital environments here. The landscape is dependent on the other services the broker can support.
- *Applications framework – programs and data to support business logic.* There needs to be a framework which orchestrates the other parts; which brings together the user requests with appropriate services; which merges and manipulates data for presentation to a user or to pass between functions; which implements the particular business processes needed (e.g. distributed document delivery, clump, cross-domain resource discovery, etc.). To do this it requires 'forward knowledge' about formats, applications and collections to which it brokers access. This will be provided by metadata to which it has access.
- *Distributed service interfaces – managing interaction with services and resources.* Various resources and services require different modes of interaction. Ideally, the broker will abstract the mechanics of these interactions. A reader may request an item to be delivered; the system will decide whether this needs an HTTP 'get', an ILL request, or a note saying to go to the reference collection. A reader may wish to locate some items; the system will present some options for searching, will open up a Z39.50 session, or a web browser, or whatever is required. The system will deal

appropriately with delivered items. The variety of services and resources means that a variety of modes of interaction will be possible. The broker will support interfaces appropriate to its application area. In a real-world environment, the absence of widespread deployment of standard machine interfaces means that a broker may have customised or specialised interfaces to particular services. There are clearly advantages in minimising the number of interfaces which have to be supported.

- *Control of access – authentication, commerce.* The library as a place has certain controls built in: membership cards, single points of entry and exit, supervision. Open distributed control in the digital information space is still a research and development challenge. Authentication and authorisation information needs to be exchanged at various stages. Support for commercial activity will have to be provided. The integrity and privacy of exchanged information may have to be assured. The usefulness of an information landscape will be severely mitigated without distributed authentication services which mean that the client need only 'prove' himself or herself once. Multiple challenges (passwords) erect fences in the landscape which inhibit use. We do not discuss issues of control further here. They are potentially relevant at various points in information exchanges.

These are logical functional groups which have worked well when measured against a range of emerging developments, and are further explained in discussion of the MIA. The advantage of such an approach is that it separates different aspects. So if a service is offered through a different protocol, or if a new service is added, it should not be necessary to change the user access level. The appropriate transformations will be effected in the middle layer. Similarly, users may see available resources through different landscapes without having to alter the way in which those resources are organised. We begin to see how new resources might be routinely 'shelved' by being added to the lower layer. We also begin to see how the flexibility introduced in the user access layer makes it plausible to consider a variety of customised approaches into the available resources.

### *3.3 The service environment*

*3.3.1 The example of serials.* We have explored how the current library electronic information environment is fragmented in much the way we described in phase two above, with special reference to the journal literature. Some of the issues can be made clearer if we look at the actual technical environment of article discovery and supply in some more detail, and relate these issues to our description of the broker services. To do so is to be reminded of remarks quoted earlier of how standardisation is lagging behind actual digital library development requirements. (We explore some of the standardisation issues for serials elsewhere [34].)

*3.3.1.1 Discovery.* In current automated systems, the 'discovery' system is very underdeveloped. It tends to be limited to a series of one-off searches in discrete databases. Future services will need to be richer in various ways:

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

- From a distributed system point of view artificial constraints are placed on this process by physical database boundaries. In some scenarios the user should be able to search multiple physical databases simultaneously and transparently. It should be noted that such a development is not incompatible with continuing access to native interfaces for specific purposes.
- Once presented with this capability a higher level issue is raised of knowing which databases to search. It is clearly not practical from a user or systems perspective to search all online databases simultaneously, hence there is a requirement for some 'database directory' or directories that can be used to identify potential targets for the user's search. Metadata describing available services and collections will be required.
- Current systems are very passive: they do not remember a user's requirements or initiate actions on their behalf. It is likely that some way of profiling user requirements, preferences and permissions will be deployed.

The discovery function may depend on a single search in a bibliographic database, or it may depend on processing of several potential searches and their results:

- bibliographic data ('item' level metadata);
- data about bibliographic data sources, including intellectual content, terms and conditions, technical access characteristics (this is metadata at the 'collection' or 'service' level);
- user profiles (this is metadata for people).

These data could be served up in different ways. Bibliographic data might be available through Z39.50 servers, data about bibliographic data sources might be available through the explain service of Z39.50, they might be made available through a directory service (e.g. LDAP – Lightweight Directory Access Protocol). User profiles might be made available in a number of ways including a directory service.

We can identify several current impediments. Currently, there is no widely deployed directory-type service in the UK; this means that local systems have to be configured with environment information. There is no agreed way of representing user profiles or database content. Experience with Z39.50 is still limited. Although there has been a steady growth in the number of Z39.50-based services and of Z39.50 client capability, there is little production use as yet in the UK. This may change in the light of the eLib phase three development discussed above. The Z39.50 'explain service' is not yet widely deployed. Other search and retrieve protocols, Whois++ for example, do not have widespread use in this environment.

*3.3.1.2 Locate.* Within libraries, the locate function presents a challenge which arises from a historical accident in the way libraries provide access to the print literature [35]. Typically, the library catalogue provides access to journals at the journal title and issue levels, rather than at the article level. It says which titles are in the library, and sometimes which issues.

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

Abstracting and indexing services say which articles are in the world, but they do not typically say where they are or who might provide them. There are no tools which 'locate' articles in the library, and, in the absence of widely deployed concise identifiers for journal articles, location is largely a manual process. Automation of this operation would confer many benefits, and will be increasingly necessary in emerging service environments where documents may be sourced from multiple locations (for example, from a resource sharing consortium, from publisher databases, from document suppliers).

Depending on the level of indexing employed at a specific database, users wishing to ascertain the potential availability of a required article may need to interact with bibliographic information at three distinct levels :

- article level – the individual article may be indexed with associated holdings information. (An example of this would be a holdings file associated with an electronic document delivery service.);
- issue level – there may be no record for the article, but there may be a record pertaining to the journal issue containing it. Holdings information for the issue may also be available;
- journal title level – the only indication of the existence of the article may be through analysing the holdings summary statement at the journal title level.

Depending on the level of indexing and the level of standardisation employed, inter-operability between the discovery and locate functions may be performed in a number of ways:

- The SICI (Serial Item and Contribution Identifier) represents a standardised, concise mechanism for transferring unique identification for both articles and issues between discovery and locate functions (and then onto the request phase). More recently, we have seen the emergence of the Digital Object Identifier [36], an initiative of the Association of American Publishers which may identify an object or some method of acquiring it [13].
- In the absence of a SICI, the article will be identified to the locate service in some less concise non-standardised unstructured manner (e.g. a citation format). Location services (automated or manual) must then rely on fuzzy matching techniques to equate the request with their own holdings; this may also rely on some intelligent analysis of holdings summary statements.

Having ascertained that a particular supplier potentially has the required article, holdings level information may provide the current availability status – this may either be the 'circulation status' for physical items or format information for electronically available items. At this point the user may wish to know the choices in terms of availability for the item; this is especially important where multiple potential suppliers are being consulted simultaneously.

Again we can identify some impediments in the current environment. The locate function depends on search services, and a name resolution service (which returns bibliographic details for a SICI) or special procedures to match records.



All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

The SICI is not now widely deployed. Holdings data for serial articles are not widely standardised [37]; in fact, in a UK environment, there is considerable diversity of technical and service approaches. Nor are there agreed ways of representing terms and conditions data [38].

*3.3.1.3 Request.* The identification of the required item, along with the preferred terms of availability and delivery mechanisms must be passed from the locate to the request process. The request itself must also carry order information (delivery address, billing address etc.) to the supplier. A request transaction may involve many messages passing between requester and supplier since it must cover the entire lifecycle from a request being placed to successful/unsuccessful completion. For inter-operability at the business as well as system level there must be a common understanding of the meaning of these messages and the effect they have on the status of the request. Interesting issues are raised in a heterogeneous physical and electronic document delivery environment where the life span of a request may be anything between milli-seconds and years. A user should be able to query the status of a request at any stage. Furthermore the request may mesh with the exchange of business messages surrounding the transaction.

Similar impediments exist here. Request requires an agreed way of transmitting request information. There are a number of ways of doing this, none of which is widely deployed. The ILL protocol and Z39.50 item order extended service are being looked at in several contexts. EDI also has a potential role here but does not seem to have been widely deployed in this context. Several proprietary systems are in use by document suppliers, including the British Library Document Supply Centre.

*3.3.1.4 Bringing it together – automating an end-to-end process?* The discussion of serials literature is summarised in Figure 5, which outlines user processes or functions and the data entities that support them.

We can relate this back to the broker discussion by outlining the concrete service scenario to be realised in the Agora prototype (see Figure 6). This will provide a service in autumn 1998 and will provide a view over a social science subset of resources. Several databases will be available for searching for discovery purposes: general purpose union catalogues, specific social science resources, and a library catalogue specialising in social sciences (British Library of Political, Economic and Social Sciences). Several databases will also be available for location purposes, including a union catalogue, a BLDSC resource, and a library catalogue (University of East Anglia). Requests will be sent to the BL, to a group of libraries, and to an individual library.

- *User access.* Agora will explore the construction of personalised landscapes by matching a user profile (with user permissions and interests) against service/collection descriptions. Initially, with a low number of resources available, this may be somewhat artificial and resources may be listed. For discovery, the user interface will need to offer search options based on the capabilities of the resources selected to be searched.

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

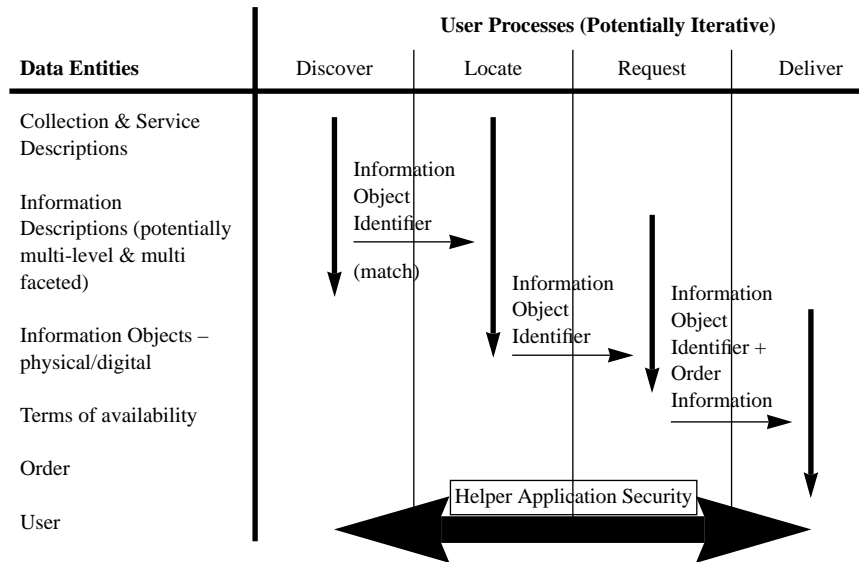


Figure 5. Reference model

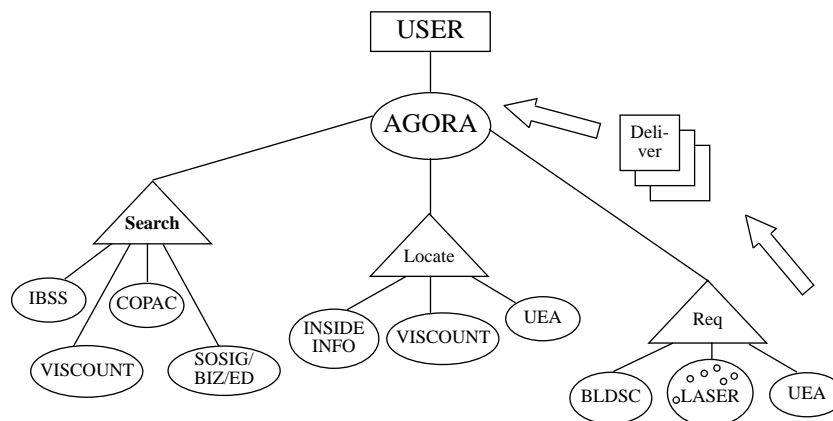


Figure 6. Agora prototype

Decisions will have to be taken about how to present locate and request services which will depend on the business model in operation: end-users may not have access to these services. The aim will be to present services in terms of logical functions: users will not have to worry about underlying mechanics or protocols. As one moves from left to right, the difficulty in business terms of developing the service increases.

- **Applications framework.** Agora will initially support discovery, locate, request and deliver. The system will need to be able to manage intra-function integration within these and inter-function operations between

January 1999

BROKER SERVICES

them. So, for example, in the discovery phase the system will allow searches across the *International Bibliography of Social Sciences* on CD-ROM (SilverPlatter), an Internet subject gateway (SOSIG) and an OPAC. It will have access to collection, service and schema description so as to be able to support the creation of the landscape and have some understanding of the capabilities of the underlying services (e.g. what fields are supported for searching). It will map formats appropriately. The current situation will not allow full integration. Take the link between discovery and location, for example. It is likely that there will need to be human intervention, in the absence of identifier information at appropriate levels of granularity and the unavailability of standard holdings data. Having indicated a desired article, the system may initiate an ISSN search on one of the location tools. The user may have to inspect holdings information.

- *Distributed service access.* Agora will initially support the following interfaces: discovery and locate (Z39.50), request (ILL protocol, the British Library's ARTemail service, and a generic email format which may be customised to comply with vendors' proprietary formats), delivery (supports the GEDI header format, and several carriers). In as much as possible upper levels should be insulated from change at this level. So the user access methods should not change if other protocols or approaches are supported. Other interfaces could also be added: if, for example, a simple web search protocol is developed, an appropriate facility could be added at this level.
- *Control.* Authentication and commerce services will be required at various stages.

As the project develops, the aim is to begin to explore managing a wider range of services in real library service environments.

*3.3.2 The wider information environment.* The above discussion focuses on journal articles. Libraries, and other intermediaries, are exploring how to provide managed access to other types of resource: Internet resources, geospatial data, archives and museums resources, data and text archives, and so on. We are seeing the emergence of a network environment in which there will be an indefinitely large number of repositories. These will be embedded in various technical, service and business environments.

The extent to which the library seeks to offer access to the intellectual record in all its manifestations is one of the issues facing it over the next several years. Whatever the ambition, it will involve partnership with a range of professional and service sectors.

A characteristic of the network environment is that it brings within the same 'organised space' materials which previously were compartmentalised within particular professional or sectoral services. For example, the recently published *New library: the people's network* looks forward to a renovated public library service. It discusses community information, cultural heritage materials, business

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

information, Internet resources, and other information types [39]. Users have cross-domain information needs. That is:

. . . they require access to information about relevant materials irrespective of where, how (e.g. as books, audio tapes, digital objects), or by whom (e.g. librarians, data archivists, museum curators) they are stored, and regardless of the manner in which they are described or catalogued. . . . A university, for example, may wish to enable students and teachers to discover scholarly materials irrespective of whether information about those materials is described and organised differently in separate library, archive, and museum information systems. It may further seek to integrate information to a particular range of externally managed Web-based information resources [40].

A reader interested in the history of Leeds may wish to look through local printed materials, but also search local history image databases, archival resources, as well as Internet search services and a range of other resources, which may be within or without library control. Information about Roman Bath will not be confined to the journal literature. Books and serial articles, demographic and other social data sets, and images are equally relevant to the cultural historian interested in the image of the city in modernist literature. The child doing a school project on butterflies and evolution may wish to look at images in the Natural History Museum, an encyclopedia article, and several journal articles. A teacher may wish to look for instructional materials and articles on a topic. And so on . . .

This at once makes the issue of providing broker services more pressing and more complicated. The addition of new resource types means that additional formats and interfaces may need to be understood, but the approach we have discussed is extensible in that way and the development work we discuss below is drawn from several domains. Without such an approach, the twin problems of user frustration and management overhead will be further exacerbated.

#### 4. THE MODELS INFORMATION ARCHITECTURE

We have already introduced aspects of MIA in considering the construction of broker services. Here we present it more formally.

##### *4.1 The broker model*

MIA is based around the broker model. The broker or 'trading place' is a venue where service requests and service providers come together. Services are 'advertised' in the broker (see Figure 7).

These terms may be familiar from a distributed object environment, but we intend no specialised meaning here. A broker may be a set of annotated web links; it may deploy a more sophisticated apparatus which supports a richer business model or quality of service (see Figure 8).

The broker provides infrastructure for managed access to physically distributed resources. In the examples we discuss below, the broker is provided by a

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

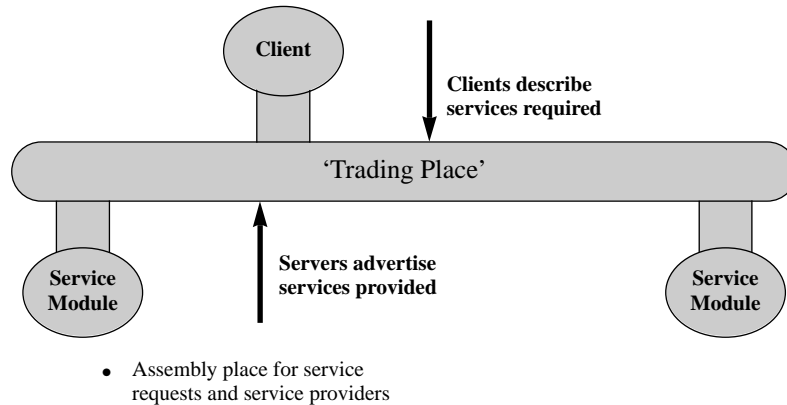


Figure 7. *MODELS brokerage metaphor – vertical integration*

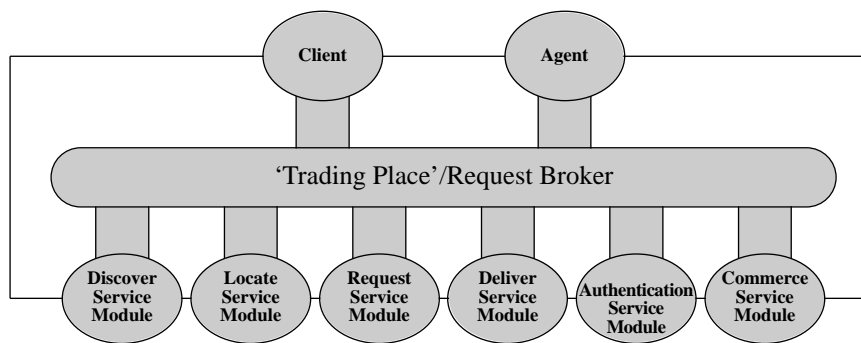


Figure 8. *MODELS brokerage evolution – horizontal integration*

bespoke piece of software which integrates particular capabilities. They follow the 'external mediation' mode discussed above. Other modes may also be supported, where CORBA and DCOM have emerged as means of supporting communication between objects.

#### 4.2 Broker components

There are four broad functional blocks, which we have already introduced:

- *User access.* Functionality which manages interaction with the user. There needs to be support for user profiles and for presenting the 'landscape'.
- *Applications framework.* The business logic and data needed to support operations. We have identified a number of functions here: others could be supported also. The broker needs to have access to knowledge about collections available, about their interfaces, and about data formats.

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

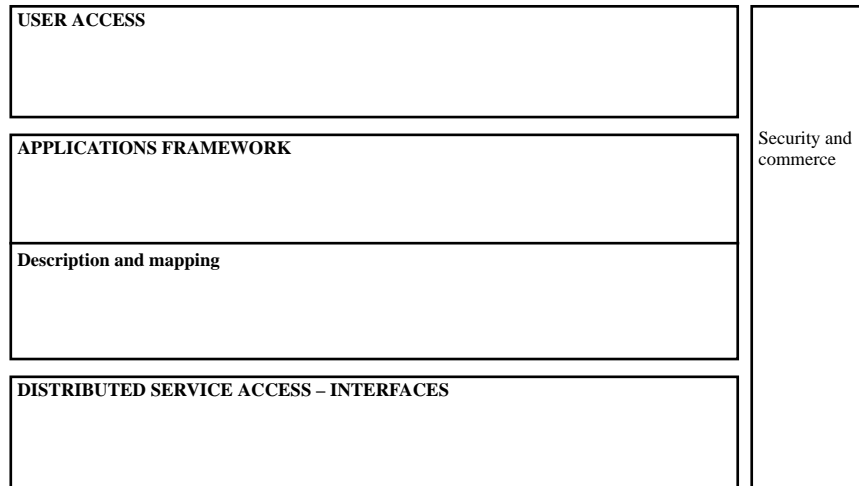


Figure 9. *MIA broker component examples*

**LIBRARY BROKER**

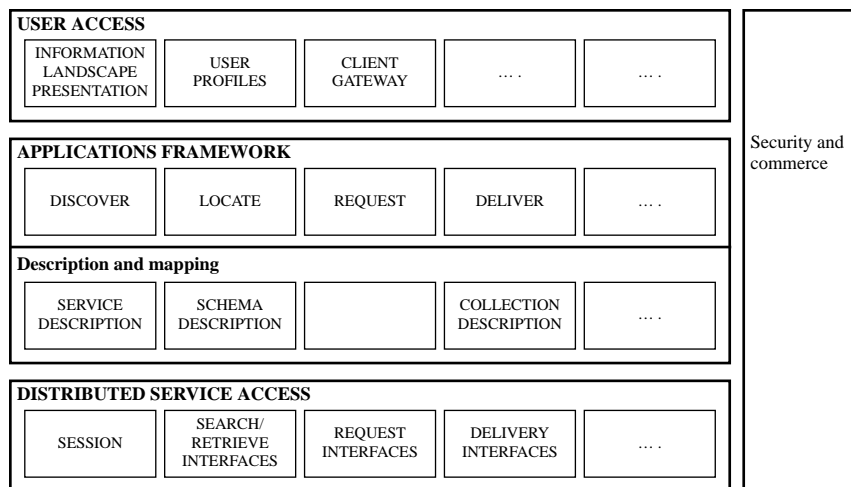


Figure 10. *Detail MIA broker component examples*

- *Distributed service access.* Interfaces to support interaction with physically distributed services. Typically the broker will support client facilities, which communicate with external servers.
- *Security and commerce – control.* It is recognised that authentication, commerce and related services may be required at several places in any set of transactions. We do not discuss these important services further here (see Figures 9 and 10).

January 1999

BROKER SERVICES

#### 4.2.1 Applications framework

4.2.1.1 *Introduction to the identified functions.* The functions we have been working with are:

- *Discovery.* The identification of information resources which satisfy a user requirement. (In our example, the presentation of a set of descriptions of articles of interest to the user.)

Discovery may be quite complex and iterative, involving discovery at different 'levels': one needs to 'discover' the resources which may be further prospected for details about particular items. One also needs to discover the existence of other services: how does one request items from remote sources, for example? Browsing, searching or other discovery avenues may be supported. Some of the metadata requirements for discovery are introduced below.

- *Location.* The identification of instances of required resources. (In our example, the generation of information on where and on what terms a discovered article can be accessed. This might be a holdings statement in a catalogue, or it may be the result of a comparison between references returned in a search and a holdings list or catalogue of some sort.)

When using a catalogue as a discovery tool, location is relatively straightforward, as discovery and location are conflated. Where descriptive and holdings data are in separate databases, the issues are more complicated. However, there are few services which do automated location of results from other databases, leading to duplication of effort. In a wider context, in the future location may increasingly depend on the resolution of identifiers. For example in a networked environment, there is great interest in the use of persistent identifiers, Uniform Resource Names or more recently Digital Object Identifiers. An identifier may be resolved to give several locations.

It is sometimes imagined that the issue of location is one that is peculiar to the print environment; however it will become increasingly important in the digital world also. Resources will be available from different suppliers under different terms and conditions, or they may be mirrored in different locations. It may be sensible to direct users to particular versions of a resource based on decisions about network traffic, special discounts, and so on.

- *Request.* The management of a request/order transaction. (In our example, the successful completion of requests for copies of journal articles from specified users to specified service providers.)

This may be a request to BLDSC or to some other supplier, or it may mean going to the shelf. In a web environment it may be as simple as an HTTP 'get'.

- *Delivery.* The transfer of a resource to the user. (In our example, it might mean actually fetching the relevant journal from a library shelf, it might mean electronic transmission from a publisher's server, it might mean fax or postal delivery. It might mean delivery over HTTP.)

Again, it should be noted that there is nothing definitive or exclusive about this enumeration of functions: it has been found to be useful in discussions and when measured against a range of services and applications. However, different business environments will require different functional support. They may also be refined or extended in various ways: for example, a 'verification' function is introduced in the discussion of Edina below.

*4.2.1.2 Service description and mapping.* The broker needs to have access to various types of metadata to support its operation. This is data about its environment and the resources in it.

It is clear that metadata will be pervasive of distributed information environments [41]. Metadata will be associated with information objects, with applications, with people, with organisations. It will support operations by people and by programs, providing them with advance knowledge of the characteristics of objects of interest and supporting sensible behaviour. We have identified several critical areas:

- *Collection descriptions.* Typically information objects exist in collections, where a collection comprises similar information objects. These collections might be databases, web sites, document supply centres or libraries. Such collections are also, of course, information objects, and collections may contain other collections. Collections will also have different terms and conditions associated with their use. They are embedded in particular organisational and business practices, which may impose additional technical requirements on any networked solution, for charging, for example. Typically collections will be managed by organisations. Information objects may be 'data' or 'metadata'. How to characterise collections is a poorly defined area at the moment, where a variety of approaches exist and for which consensual approaches are urgently required. This is especially so in the service environment we have sketched above where a range of broker services is emerging. Each of these will have to provide ways of describing the collections and services which are available to users of their systems. 'Collection descriptions' provide forward knowledge about the resources that are available to a user, allowing sensible discrimination between them and effective use of resources. These may be intellectually created. There is also some interest in automatically generated representations. For example, centroids are inverted index style representations of database content used to support query routing [42]. They are used to support cross searching of the UK subject gateways.
- *Application (or service, or interface) description.* Collections will be made available through some machine interface which needs to be described. Such descriptions will permit brokers and clients to connect to arbitrary resources. Several approaches exist, which cover such attributes as host name, port number, etc. Other services may also need to be described in this way.



All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

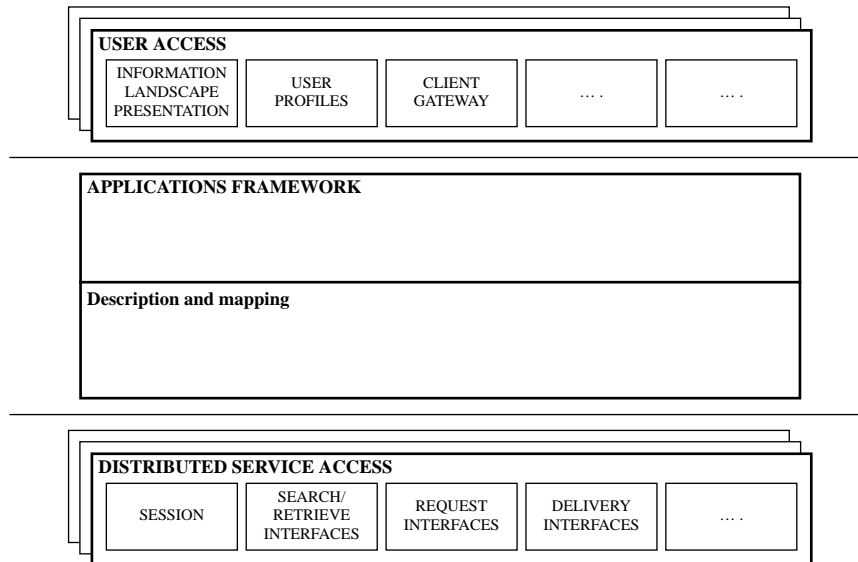


Figure 11. *Broker*

- *Schema descriptions.* If a system is to broker access to heterogeneous metadata collections, to render heterogeneous content in some way, or to map between formats, it needs to have access to schema data which support this activity.
- *User profiles.* We have discussed these above.

In the current environment it is likely that brokers may be configured with these types of data. In due course they will be stored in directory services which the broker queries.

There are various ways to create machine- and human-readable descriptions of collections, applications and user profiles at the moment. None commands universal assent. Approaches may be embedded in particular application and/or professional domains. A review of some current approaches to collection and service descriptions has been prepared as part of the MODELS project [43].

**4.2.2 Variety of brokers.** The user access layer may be realised in different environments: in a web environment, as a push based application, as VRML, and so on. Similarly, there may be several ‘bundles’ of functionality in a service access layer (see Figure 11).

## 5. THE MODEL AND THE REAL WORLD

As noted above, MIA has had a dual function: as a heuristic conceptual tool, and as a framework for development activity. In this section, several development activities are briefly described in relation to MIA. Some of this work is based on

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

MIA, and some directly influenced by it. We also include descriptions of work on the Aquarelle and NESSTARR projects, which have developed in parallel to MODELS but whose insights are now being made available within the project. Other examples could be introduced.

### 5.1 Agora

Agora is a 'hybrid library' project within eLib [44], which is directly based on the MODELS project. It is led by the University of East Anglia, with UKOLN, Fretwell-Downing Informatics and the Centre for Research in Library and Information Management (CERLIM) as partners. It works with associate groups of libraries (who will deploy the solution in the context of their electronic collection management procedures), of service providers (who will interwork with the broker), and of system developers (who will pool technical knowledge). Some of these are listed in Figure 12.

Agora will build on the existing eLib projects EDDIS and NewsAgent and will benefit from the metadata work of UKOLN. It will initially support a limited range of interfaces to remote resources. These will include Z39.50, the ILL Protocol, the British Library's ART interface to the Document Supply Centre, as well as the web protocols. At the centre of the system will be a Hybrid Library Management System (HLMS) which will be a MIA-type broker. Within this context, the project will experiment with creating 'landscapes' which provide integrated access to the types of services listed below. This is given here to outline some of the issues faced.

- Local library catalogue, union catalogues – through Z39.50 service. If the service supports the Explain service then the catalogue will describe its facilities to the HLMS, otherwise the facilities of the local catalogue are entered in the explain proxy database within the HLMS.

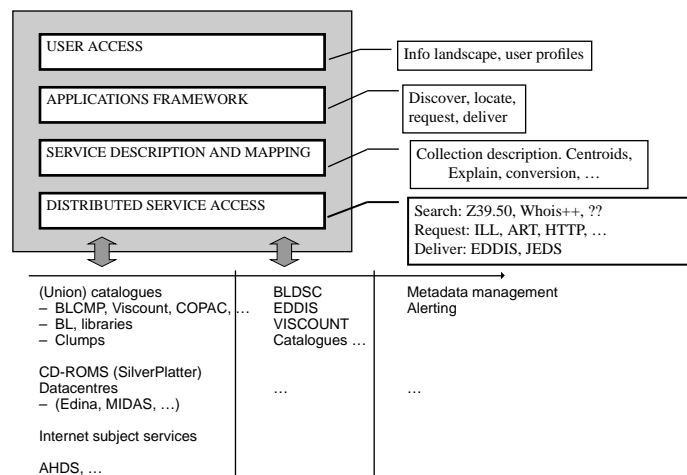


Figure 12. AGORA – Integrated Service example

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

- CD services that support the Z39.50 interface.
- Clump entry points that are visible as Z39.50 servers (i.e. a clump represents a virtual database and therefore should be a Z39.50 server). Alternatively, because the HLMS itself has parallel Z39.50 capabilities and database description facilities it also can provide clump support for its local user population.
- Electronic Document Integration – electronic documents can be integrated into the hybrid library as web servers or as Z39.50 servers.
- Web indexes – any web index that is available as a Z39.50 server can be integrated into the hybrid library system – if the index is Dublin Core enabled it will provide richer facilities. In the absence of such an index the HLMS server itself will provide web-indexing capabilities.
- Subject information gateways. These can be integrated through the Z39.50 to Whois++ gateway which is being developed in the ROADS project. Alternatively, routines are available to bulk load ROADS databases into Z39.50-based servers [45].
- Other information repositories with Z39.50 or web interfaces. For example, Agora will be working with the Arts and Humanities Data Service to ensure interworking.
- ILDRMS (Inter-lending/Document Request Management System) – the HLMS can send and manage document requests to any servers supporting any combination of ISO-ILL, Z39.50 item order, http 'get'. It will also interact with the services of the BLDSC. In this way the HLMS can interoperate with a local ILDRMS system or can provide this service for the local user community.
- News and current awareness – the HLMS can front local alerting systems or can operate an alerting service based on the underlying databases attached as system components. HLMS will interface seamlessly with NewsAgent servers.

In this way, the project will bring together in a controlled environment access to discovery, location, request and delivery applications and services, and provide the library associate group with the tools to project a unified service to their users.

### *5.2 Fretwell-Downing Informatics [46]*

Fretwell-Downing Informatics (FDI) [47] is currently implementing solutions based on the MIA architecture in a range of domains and service settings, from public library resource sharing (with LASER) to cross-domain searching for research (with the Arts and Humanities Data Service).

FDI have been involved in the development of the MIA as consultants to the MODELS project. FDI view MIA as a framework for the design and actualisation of systems which are critical for the potential of networked information to be translated into real services. From the FDI development point of view, MIA has several important features. Firstly, it is designed to accommodate the interoperation of a range of services to meet a business requirement (for example search, locate, authenticate and request for interlending). Secondly MIA services can be

applied to heterogeneous information resources, independent of domain (e.g. libraries and museums) and of format (e.g. MARC and EAD – Encoding Archival Description). These capabilities will be very important in providing implementations that are geared to end user needs – such as those of the researcher or distance learner who needs to access information, resources and services on the basis of research topics rather than by choosing services or by investigating repositories individually by domain or locale. In this context a third key feature of MIA might be emphasised: its provision for ‘forward knowledge’ of resources and services through the mediation of a gateway based information landscape. This will map out who has what where, and on what terms.

The following projects illustrate FDI implementations in a range of environments:

- Community information: the Essex Libraries SEAMLESS project (funded by the British Library) focuses on the distributed information model (which offers local autonomy for resource management) to provide a Z39.50-based service for brokering community and civic information.
- Resource-based learning: the RENAISSANCE project (EU ACTS Programme) has developed a demonstrator for the brokerage of mediated learning which illustrates the use of Z39.50 and document delivery in a distance learning setting.
- Archives: whilst ISAD(G) (International Standard Archival Description – General) provides a common logical framework for metadata, there is huge variety in the physical description of archival resources including SGML (Encoding Archival Descriptions), databases and formatted wordprocessing files. FDI and Index Data have shown in the National Networking Demonstrator Project (NNDP) funded by JISC under the Archives Sub-Committee of the Humanities NFF Committee, the potential of an archival gateway combining advance knowledge with Z39.50 searching (using the Digital Collections profile) as an approach towards addressing these issues.
- Libraries: the Australian Local Interlending & Document Delivery Administration (LIDDA – AVCC) will deploy the FDI Virtual Document Exchange (VDX) library broker to provide an end-to-end search through to delivery service direct to the end-user. The LIDDA service will broker a range of multi-vendor services including the National Bibliographic Utility (Amicus), interlending (FDI VDX) and document delivery (RLG Ariel) in a one-stop solution.

The same framework has already been adopted for key sectoral and national resource sharing initiatives in Australia and New Zealand.

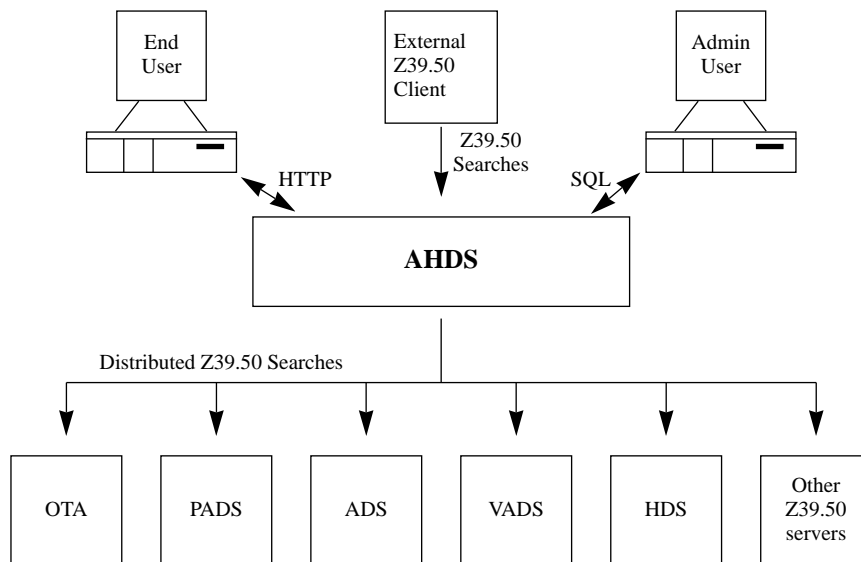
### *5.3 Arts and Humanities Data Service Gateway [48]*

The Arts and Humanities Data Service (AHDS) [49] is a national service funded by JISC to collect, describe, and preserve the electronic resources which result from research and teaching in the humanities. It has five geographically distributed disciplinary service providers: Oxford Text Archive (OTA), Performing Arts

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

Figure 13. *AHDS Gateway*

Data Service (PADS), Archaeology Data Service (ADS), Visual Arts Data Service (VADS), History Data Service (HDS) (see Figure 13).

One of AHDS's key aims is to encourage scholarly use of its collections and make information about them available through an online catalogue. Each of the five AHDS Service Providers is developing its own catalogue with record structures which are most appropriate for their community's resources; e.g. the Archaeology Data Service will adopt an SQL-based database implementation while the History Data Service's catalogue will be SGML-based. However owing to the intrinsic interdisciplinarity of humanities research, the AHDS also needs to allow users to search simultaneously across its distributed, interdisciplinary, and differently catalogued holdings. Issues surrounding cross-domain discovery were explored at the fourth MODELS workshop in December 1996, which prepared the groundwork for the subsequent series of six specialist AHDS workshops [50]. These developed recommendations on metadata to support cross-domain resource discovery. Parallel to this, Fretwell-Downing was commissioned to develop a suite of resource discovery tools. This will provide a 'gateway' based on Z39.50; it will exploit the unifying metadata format identified by the workshops. The development of the AHDS Gateway concept was influenced by the MODELS process and emerged in parallel with MIA.

The Gateway will enable users to query the AHDS's distributed holdings in an integrated way. Standard user access will be through a web interface (although Z39.50 client access will also be possible) and will provide a single point of entry to the online catalogues developed by the five Service Providers. An 'information landscape definition' will provide users with a contextualised map of the information and

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

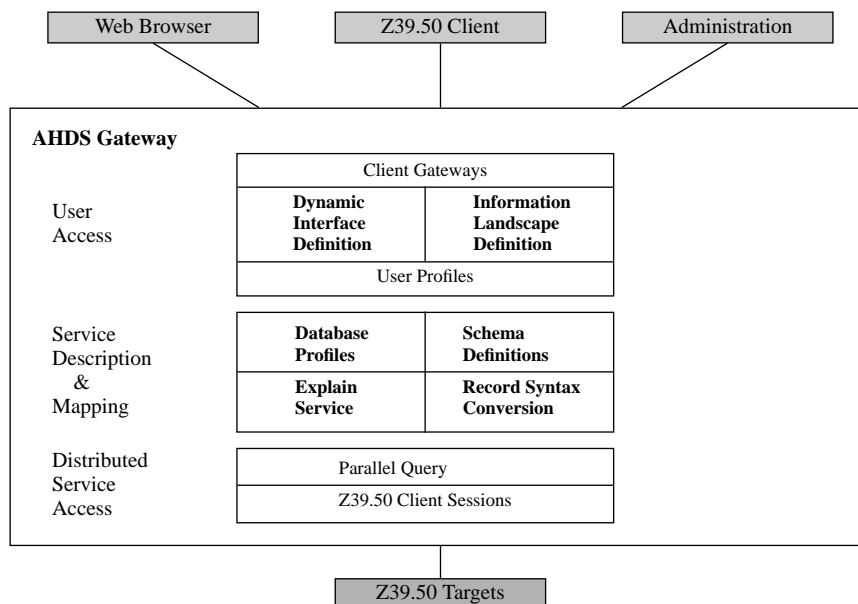


Figure 14. *AHDS Gateway components*

services available from the gateway. 'Dynamic interface definition' ensures that the user interface changes to reflect the services available to users as they move through the landscape. In time the gateway will also enable users to search other network resources of interest to humanities scholars, which are not maintained by AHDS.

The 'Service description layer' will enable the Gateway to smooth out anomalies between underlying databases (see Figure 14). It comprises two principal components: a Z39.50 Explain proxy service and a record syntax conversion facility. The Explain proxy service will store profiles of databases that are known to and thus searchable by the Gateway. A profile will include information about the database's contents and record structure, the kinds of queries it supports, and the format in which results are returned. The Z39.50 Explain function is designed to generate such information about a database whenever it is queried by a Z39.50 client. However there are currently very few Explain implementations so the Explain proxy service is an interim measure until Explain is more widely available.

A record syntax conversion facility will hide the heterogeneity of different databases from the user by converting incoming record structures into a standard internal format which will display single uniform result sets to users.

The system will initially provide access to a small number of scholarly humanities collections, but since its design is aligned with the MIA framework, it is extensible, enabling scholarly communities to take full advantage of network technologies and the proliferation of online information resources.

#### 5.4 EDINA [51]

EDINA (Edinburgh Data and Information Access) [52] was launched as a UK national datacentre early in 1996. In addition to establishing itself as a host for

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

national services, a further task was to find its role in the emerging virtual library for UK higher education, the design for which was being laid out through the MODELS workshops and the eLib Programme.

The MODELS 'user' verbs have been used in EDINA's strategic thinking: discover, locate, request and access. By hosting bibliographic discovery services (such as *Art Abstracts*, *BIOSIS Previews* and *Periodical Contents Index (PCI)*), EDINA already helps to provide part of the response to the first verb (discover). By running and developing SALSER (Scottish Academic Library Serials) [53], the virtual union catalogue of serials holdings in Scotland, the second verb (locate) is also addressed, with specialist emphasis on serials. EDINA is also working with colleagues in Europe, through the CASA project in the EU Telematics Programme, to contribute an extra verb, that of 'verify'. ISSN-based identifiers (such as SICI and DOI) and network access to the ISSN world serials register act as lynchpins to make cost-effective the automatic progression along this chain from discovery of an information object (e.g. an article), to access and use of that object. (The project also involves the development of an international infrastructure for serials services directories.)

EDINA is using OCLC SiteSearch as the software platform to achieve this interoperability. Two bibliographic databases are already in service with version 3.1 of SiteSearch. As the European field tester for version 4.0 (written in Java on the server-side with database object classes to maximise connectivity to heterogeneous databases), EDINA is taking forward several projects which will play an important part in the implementation of the MODELS Information Architecture. The first is to port SALSER from Web/WAIS into SiteSearch, in order to search information on serials holdings across databases using Z39.50. A redesign will also take advantage of the ISSN world serials register, providing a serials authority file for the virtual union catalogue. The second is to set up facilities to allow users of *PCI* and *Art Abstracts* who discover a journal article of interest to have a 'hot link' which will locate the libraries which hold that journal. This redesign of SALSER will be used as an opportunity to demonstrate interoperability, such as cross-database searching, with other discovery facilities as well as the use of SALSER to provide links to other network-accessible information on serials holdings.

EDINA is also interested in helping enact the final verbs for the digital information object of interest: request and access. This might include, for example, the delivery of electronic access to the full text of a journal article, to a research dataset, or to a (digital) map for printing. As a JISC national datacentre, EDINA may be directed to host such information objects, but it is realised that plans must be made to support connectivity to the vast range of specialist objects hosted elsewhere, either in the JISC distributed national electronic resource or in the online services of commercial vendors. EDINA is therefore supportive of the adoption of MODELS thinking, to facilitate interworking between sectors and domains.

### 5.5 *The Aquarelle project [54]*

Aquarelle is an EC Telematics Applications Programme-funded project developing the Information Network on Cultural Heritage. It is a distributed information system, offering access to interrelated multimedia reference documents and primary data that describe the cultural heritage of Europe. It aims to provide a

facility affording uniform access to the varied collections of data held by museums, art galleries, and other cultural organisations throughout Europe.

An Aquarelle user formulates a query at the user interface. This query is passed to the Aquarelle central services module, the Access Server. It is further processed by the Access Server and submitted to the data servers. The responses from the data servers are collated by the Access Server and passed back to the user. As the query passes from the user to the data servers and the results are passed back they undergo a series of transformations as they are encoded in various protocols. The principal protocols are HTTP, AQL, Z39.50 and SGML together with the local protocol used at the data servers.

Aquarelle supports two types of data servers, Archive Servers and Folder Servers. Archive Servers provide information about individual objects or sites, typically returning a record about each object or site. They follow the conventional information retrieval for database access. The Archive Server model is designed so that an existing museum collection management system or data service system could act as an Aquarelle Archive Server using the appropriate interface. Folder servers are SGML documents.

The Aquarelle system can be mapped readily to the MIA model. The Aquarelle Access Server is a broker supporting most of the MIA functions. It maintains a database of registered users for authentication and user profiles. The Aquarelle Directory services maintain a database of collection and service descriptions which support the MIA discover and locate functions. The user interface components present the information landscape in terms of subject domains as well as specific databases; they also provide multilingual thesauri to assist in query formulation. Aquarelle supports the MIA request and delivery functions for Aquarelle folders. In addition Aquarelle offers facilities which are not explicit MIA components: folder publishing, persistent link management and multilingual thesauri for query formulation.

#### *5.6 The NESSTAR project [55]*

The NESSTAR [56] project has brought together several European data archives (the Data Archive, UK (University of Essex), Danish Data Archives and Norwegian Social Science Data Services), as well as data producers and users, to develop distributed yet seamless resource discovery and use systems. It has been funded by Information Engineering, part of the EC Telematics Applications Programme.

The archives typically hold data about society, whether economic or social, which have been collected by many government departments, research institutes and companies. The project will increase the use of these data by developing a set of generic tools that will make it easier to:

- locate multiple data sources across national boundaries;
- browse detailed information about these data;
- tabulate and visualise the data;
- disseminate the data and documentation, in whole or part, in forms suitable for immediate use.

Producers of data will have improved means of structuring and disseminating data in controlled ways. Users of data will have access to improved tools to identify the most appropriate data sources and will be able to analyse the data



All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

effectively, either via the web browser or via downloading to the local site. The effect of these developments will be not only to increase productivity among traditional researchers, but also to lead to greater data use among new, more statistically naive, groups of users.

In order to achieve these goals the project will take advantage of three key technologies. These are:

- SGML/XML to structure both the catalogue records and data dictionaries (codebooks) according to international standards;
- Z39.50 to facilitate interoperability with both the other archives at data dictionary level and other resource providers at the catalogue level (Dublin Core);
- JAVA to create user controlled front ends for interaction with the search engine and data display.

The project has a strong focus on the ease of use of social and economic data, but with a specific aim of making these resources more available by participating in complementary developments in the electronic library world, such as the MODELS Information Architecture (MIA) model.

#### 5.7 ROADS cross searching [57]

The ROADS cross-searching facility was introduced above. This could be discussed in the following terms:

- *User access:* web page offers cgi gateway to a Whois++ client.
- *Applications framework:* Whois++ is a distributed directory service. It supports query routing by means of centroids, index objects which provide inverted index style representations of database content. In effect a centroid is a machine-generated 'collection description', providing 'hints' as to database content by listing words and the template types in which the occur. The Whois++ client may query a local database or receive 'forward knowledge' results from other databases by means of referrals from a query to centroids. If there is a referral, the reference may be displayed, or the record may be retrieved. The Whois++ client has knowledge of several template types or schemas.
- *Distributed access:* remote Whois++ servers are queried using the Whois++ protocol. Forward knowledge is stored as centroids polled from remote servers. They are obtained and dealt with in accordance with the Common Indexing Protocol. CIP provides facilities for centroids (or other index objects) to be moved between a mesh of servers [42].

## 6. CONCLUSIONS

We have described aspects of the current network information environment, and argued that current management approaches are unsustainable: they will not support users adequately, and they will impose an increased management overhead on the library. We have outlined a broker-based approach and have introduced the MODELS Information Architecture, a device for thinking about desirable

components of such approaches and their logical arrangement. We have described some development activity which is aligned with the approach suggested.

MIA separates the presentation of an integrated service to users from interaction with the components which provide that service. In this way, it becomes possible to develop an 'information landscape' which responds to user interests and preferences rather than to particular system constraints. It becomes possible to manage the acquisition or substitution of components while minimising the impact on the user. A framework is suggested which supports intra-function integration so as to protect the user and manager from difference, and inter-function integration so as to support automation of end-to-end processes.

This is not to suggest that there are not many difficulties in a heterogeneous environment, and some of the issues in relation to the serials literature have been explored.

The library will increasingly broker direct user access to resources and services, organising information flows in a managed environment. This will be necessary to support the changing educational and social environments in which they operate. When technologies of control improve (authentication, rights management, commerce), this trend will become more pronounced. The consequence of a lack of integration is that investment in networked information is less efficient than it could be:

- Network resources do not release their full value. There is some evidence, summarised by Bell for example [58], that variety inhibits effective use for casual users who do not have the time or inclination to become familiar with more than a few different approaches, and who do not stray from familiar paths. This variety also presents a management issue for libraries: different technical characteristics may require different approaches, training needs are duplicated, the technology is very much on the surface rather than disappearing into a coherent environment of use. Such case-by-case treatment is expensive.
- Users are not well served. Relevant resources may not be visible, or will be ignored. The learning or social opportunity is diminished.
- Missed opportunity of networking. The opportunity now posed by the developing technologies is not just to automate particular tasks but to automate end-to-end processes in ways that support effective use, and deliver integrated services into working environments.

The approach developed here suggests how libraries can begin to develop a 'coherent context' into which the technology can disappear. We expect the library will continue to organise the assembly places where information users and information products are brought into fruitful contact. The challenge is to help create a new 'utopian place of criticism' where the organised reading space of the world's resources can be opened 'side by side' in the user's workspace.

#### ACKNOWLEDGEMENTS

This work has been supported by the MODELS project. UKOLN is funded by the British Library and the Joint Information Systems Committee. It also receives support from the University of Bath where it is based.

January 1999

BROKER SERVICES

REFERENCES

1. McCullough, M. *Abstracting craft: the practised digital hand*. Cambridge, MA.: The MIT Press, 1996, 251.
2. Manguel, A. *A history of reading*. London: HarperCollins, 1996.
3. Jacob, C. La leçon d'Alexandrie. *Autrement*, 121, Paris, 1993 (as cited in Manguel, A. *A history of reading*. London: HarperCollins, 1996).
4. Heseltine, R. Resource discovery and systemic change: a UK perspective. In: Dempsey, L., Law, D. and Mowat, I., eds. *Networking and the future of libraries 2: managing the intellectual record*. London: Library Association Publishing, 1995, 119–124.
5. JISC. *JISC collections policy: an integrated environment for higher education: developing the distributed national electronic resource (DNER)*. [Bristol]: JISC, 1998. Also available at [http://www.jisc.ac.uk/cei/dner\\_col\\_pol.html](http://www.jisc.ac.uk/cei/dner_col_pol.html) (visited 23 May 1998).
6. Hanson, T. The access catalogue gateway to resources. *Ariadne*, 15, 1998. <http://www.ukoln.ac.uk/ariadne/issue15/main> (visited 2 June 1998).
7. Paepcke, A. et al. Using distributed objects for digital library interoperability. *Computer*, 29(5), 1996, 61–68.
8. Russell, R. Towards new models for managing and accessing serials resources. *Managing Information*, 4(8), October 1997, 37–39.
9. Dempsey, L., Russell, R., Murray, R. and Heseltine, R. Managing access to a distributed library resource. *Program*, 32(3), 1998, 265–281.
10. Lynch, C. A. Identifiers and their role in networked information applications. *ARL Newsletter*, 194, 1997. Available at <http://www.arl.org/newsltr/194/identifier.html> (visited 7 April 1998).
11. Powell, A. Unique identifiers in a digital world. *Ariadne*, 8, March 1997. <http://www.ariadne.ac.uk/issue8/unique-identifiers/> (visited 20 May 1998).
12. Young, A. *Implementation of JANET authentication and encryption services. A report published by the JTAP programme of JISC*. 1997. Available from the JTAP web site: <http://www.jtap.ac.uk/> (visited 20 May 1998).
13. Lynch, C.A., ed. *A White Paper on authentication and access management issues in cross-organizational use of networked information resources*. Revised discussion draft of April 14, 1998. Available from the Coalition for Networked Information web site: <http://www.cni.org/projects/authentication/authentication-wp.html> (visited 19 May 1998).
14. Further information about MODELS and its results can be found at <http://www.ukoln.ac.uk/dlis/models> (visited 20 May 1998).
15. Russell, R. UKOLN MODELS 4: evaluation of cross-domain resource discovery. In: Greenstein, D. and Miller, P., eds. *Discovering online resources across the humanities*. Bath: UKOLN (on behalf of the Arts and Humanities Data Service and the UK Office for Library and information Networking), 1997, 18–21.
16. Further information can be found on the Telematics for Libraries web site: <http://www2.echo.lu/libraries/en/libraries.html> (visited 19 May 1998).
17. Further information about eLib can be found on the eLib web site: <http://www.ukoln.ac.uk/services/elib/> (visited 19 May 1998).

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

18. Further information about the Digital Libraries Initiative can be found at <http://dli.grainger.uiuc.edu/national.htm> (visited 20 May 1998).
19. Dempsey, L., Russell, R. and Murray, R. The emergence of distributed library systems: a European perspective. *Journal of the American Society for Information Science*, 49(10), 1998, 942–951.
20. Mackenzie Owen, J.S. and Wierck, A. *EUR 16905 – Knowledge models for networked library services*. Luxembourg: Office for Official Publications of the European Communities, 1996.
21. Dempsey, L. and Russell, R. ‘Clumps’ – or distributed access to scholarly material. *Program*, 31(3), July 1997, 239–249.
22. Baldonado, M. et al. *The Stanford Digital Library metadata architecture*. [1998] <http://www-diglib.stanford.edu/diglib/pub/delos.html> (visited 6 April 1998).
23. Paepcke, A. et al. Interoperability for digital libraries worldwide. *Communications of the ACM*, 41(4), April 1998, 33–43.
24. Moen, W.E. Accessing distributed cultural heritage information. *Communications of the ACM*, 41(4), April 1998, 45–48.
25. Committee on Earth Observation Satellites. *Interoperable catalogue system: system design document*. Version 1.2. Committee on Earth Observation Satellites, March 1997.
26. Greenstein, D. and Murray, R. Metadata and middleware: a systems architecture for cross domain discovery. In: Greenstein, D. and Miller, P., eds. *Discovering online resources across the humanities*. Bath: UKOLN (on behalf of the Arts and Humanities Data Service and the UK Office for Library and information Networking), 1997, 56–62.
27. Valkenburg, P., ed. *Standards in a distributed indexing architecture*, draft version 1. 24 February 1998. [http://www.terena.nl/projects/chic-pilot/standards\\_v1.html](http://www.terena.nl/projects/chic-pilot/standards_v1.html) (visited 27 May 1998).
28. Lincke, D-M. and Schmid, B. Mediating electronic product catalogs. *Communications of the ACM*, 41(7), 1998, 86–88.
29. Kochikar, V.P. The object-powered web. *IEEE Software*, May/June 1998, 57–62.
30. Web standards can be tracked at the World Wide Web Consortium’s website <http://www.w3.org/> (visited 25 May 1998).
31. Larbey, D. Project EDDIS: an approach to integrating document discovery, locate, request and supply. *Interlending and Document Supply*, 25(3), 1997, 96–102.
32. The ROADS cross searching service is available at <http://www.ukoln.ac.uk/metadata/roads/crossroads/> (visited 20 May 1998).
33. An overview of the Gazebo software is available at <http://emerge.ncsa.uiuc.edu/gazebo.html> (visited 20 May 1998).
34. Russell, R. Standards for serials: building the basis for distributed access. *Serials*, 10(3), November 1997, 357–364.
35. Dempsey, L. Readbites: the scandal of serials holding data. *Catalogue & Index*, number 118, Winter 1995, p. 9.
36. Information about the DOI Foundation and DOIs can be found at <http://www.doi.org/> (visited 19 April 1998).

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

January 1999

BROKER SERVICES

37. Hopkinson, A. *Standards for serial holdings and for serials data in the serials analytic record*. [1997] An eLib supporting study co-ordinated by UKOLN and prepared as a result of MODELS 1. Available from the eLib supporting studies web page <http://www.ukoln.ac.uk/dlis/models/studies/> (visited 12 May 1998).
38. Martin, D. and Bide, M. *Standards for serials metadata and Terms of availability – combined report*. 1997. An eLib supporting study co-ordinated by UKOLN and prepared as a result of MODELS 1. Available from the eLib supporting studies web page <http://www.ukoln.ac.uk/dlis/models/studies/> (visited 12 May 1998).
39. Library and Information Commission. *New library: the people's network*. London: Library and Information Commission, 1997. Text available at <http://www.ukoln.ac.uk/services/lic/newlibrary/> (visited 20 May 1998).
40. Greenstein, D. and Dempsey, L. Crossing the great divide: integrating access to the scholarly record. In: Greenstein, D. and Miller, P., eds. *Discovering online resources across the humanities*. Bath: UKOLN (on behalf of the Arts and Humanities Data Service and the UK Office for Library and information Networking), 1997, 7–10.
41. Dempsey, L. and Heery, R. Metadata: a current view of practice and issues. *Journal of Documentation*, 54(2), 1998, 145–172.
42. Allen, J. and Mealling, M. *The architecture of the Common Indexing Protocol (CIP)*. Request for Comments draft version 1. 1997. Available at <ftp://ftp.ietf.org/internet-drafts/draft-ietf-find-cip-arch-01.txt> (visited 25 May 1998).
43. The MODELS collection description study is available at: <http://www.ukoln.ac.uk/dlis/models/studies/> (visited 8 June 1998).
44. The Agora web-site is at <http://hosted.ukoln.ac.uk/agora/> (visited 3 May 1998).
45. Experimental Z39.50 servers including ROADS databases can be trialled at: <http://roads.ukoln.ac.uk/cgi-bin/egwcgi/egwrtcl/targets.egw> (visited 5 August 1998).
46. This section is adapted from text supplied by David Kay, Fretwell-Downing Informatics.
47. Information about Fretwell-Downing is at: <http://www.fdggroup.co.uk/FDI/> (visited 3 May 1998).
48. This section is based on Greenstein, D. and Murray, R. Metadata and middleware: a systems architecture for cross domain discovery. In: Greenstein, D. and Miller, P., eds. *Discovering online resources across the humanities*. Bath: UKOLN, 1997, 56–62, updated by personal communication from Daniel Greenstein.
49. Information about AHDS is at: <http://www.ahds.ac.uk/> (visited 3 May 1998).
50. Miller, P. and Greenstein, D., eds. *Discovering online resources across the humanities: a practical implementation of the Dublin Core*. Bath: UKOLN, 1997.
51. This section is adapted from text provided by Peter Burnhill, Director EDINA.
52. Information about EDINA is at: <http://edina.ed.ac.uk> (visited 3 May 1998).

All rights reserved. Except as otherwise permitted under the Copyright, Designs and Patents Act 1988, no part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise without the prior written permission of the publisher.

JOURNAL OF DOCUMENTATION

vol. 55, no. 1

53. Information about SALSER is at: <http://edina.ed.ac.uk/salser/> (visited 3 May 1998).
54. This section has been adapted from text supplied by Mike Stapleton, System Simulation Ltd.
55. This section has been adapted from text supplied by Simon Musgrave, Director, The Data Archive, University of Essex.
56. Information about NESSTAR is at: <http://dawww.essex.ac.uk/projects/nesstar/> (visited 3 May 1998).
57. Based on some text supplied by Ian Peacock, UKOLN.
58. Bell, A. The impact of electronic information on the academic research community. *The New Review of Academic Librarianship*, 3, 1997, 1–24.

(Revised version received 4 September 1998)