

Cataloguing the Internet: CATRIONA Feasibility Study

Report To The British Library Research & Development Department

Dennis Nicholson, Mary Steele, Gordon Dunsire and Fred Guy
British Library Research and Development Department, 1995

[Abstract Contents Page Appendices](#)

The British Library Board 1995

The opinions expressed in this report are those of the authors and not necessarily those of the British Library.

RDD/G/260

Abstract

[Back to First Page](#)

The idea of a distributed catalogue of Internet resources integrated with standard Z39.50 library system OPAC interfaces (and hence with retrieval of information on hard copy resources) is already practical at a basic level. Geac's Z39.50 GUI OPAC client, GeoPac, can search remote Z39.50 OPACs, retrieve USMARC records with URLs in 856\$, respond by loading a viewer like Mosaic or Netscape, and utilise it to retrieve and display the remotely held electronic resources on the local workstation. A range of Z39.50 OPACs can be searched server by server, making a basic-level distributed catalogue of Internet resources feasible. At least one other Z39.50 client, Dynix Horizon is close to having similar capabilities.

Significant further development and investigation is nevertheless required. A proposed demonstrator project - based around Scottish University Libraries and the BUBL Subject Tree initiative, but sufficiently 'open' to encompass other sites and approaches - is both feasible and essential, and would provide a focus for Z39.50 developments in the UK.

Z39.50 clients and associated Z39.50 OPACs describing resources could become preferred network navigation tools with other specific NIDR client types (WWW, gopher, WAIS, others) loaded as required. Library involvement is essential to sustainable Internet cataloguing initiatives.

Contents

[Back to First Page](#)

[CATRIONA Feasibility Study: Overview and Conclusions](#)

- [1. Aims of the Study and the Problem in Outline](#)
- [2. The Proposed Solution: Assumptions Underlying the CATRIONA Model](#)
- [3. The Proposed Solution: Outline Description of the CATRIONA Model](#)
- [4. Work Done, Problems Encountered, Limitations of the Project](#)
- [5. Feasibility of the Model and of a Follow-up Demonstrator Project](#)
- [6. Summary of Development Requirements](#)
- [7. CATRIONA Demonstrator: General Points](#)
- [8. CATRIONA Demonstrator: Specific Phase II Proposals](#)

[References](#)

Appendices

[Back to First Page](#)

[A. Glossary of Terms and Acronyms](#)

[B. CATRIONA Model Version 4: Illustrative Description of User Interaction](#)

[C. Methodology and other Background Detail](#)

[D. Feasibility Tests: Background Information, Client Capabilities and Requirements](#)

[E. Related Projects and Developments](#)

[F. Cataloguing and Resource Description Developments](#)

[G. CATRIONA - A Case Study: Napier University](#)

[H. Reports of Meetings on Z39.50 Developments](#)

[I. BUBL Subject Tree Project and Relationship to CATRIONA](#)

[J. List of Project Participants, Advisors and Correspondents](#)

[K. Workshop on Cataloguing Electronic Resources](#)

CATRIONA Feasibility Study: Overview and Conclusions

The CATRIONA feasibility study has shown that the idea of a distributed catalogue of Internet resources integrated with standard Z39.50 library system OPAC interfaces (and hence, with retrieval of information on hard copy resources) is already a practical proposition at its most basic level, and that the proposed next step, a distributed CATRIONA demonstrator project - based on the SCURL group of Scottish University and Research Libraries co-operating to catalogue local electronic resources and selected areas of the BUBL Subject Tree, but also sufficiently 'open' to encompass other sites, projects and approaches - is both feasible and essential.

A Z39.50 GUI OPAC client available from one library system vendor (GeoPac Release 1.23 from Geac) has been observed to be capable of

- *conducting searches of remote Z39.50 OPACs (Brigham Young University in Provo, Utah and Butler University in Indianapolis from Strathclyde University in Glasgow)*
- *retrieving USMARC records enhanced to include URLs in 856 subfield \$u, as recommended by MARBI (MARBI, 1993d)*
- *automatically loading an appropriate viewer (Mosaic, Netscape, a Geac image viewer in different instances) in response to these URLs*
- *utilising the viewer to retrieve and display the remotely held electronic resources on the local workstation.*

Since the client in question can be set up so as to add additional Z39.50 servers to its search list, it would clearly be possible to create a distributed catalogue of Internet resources, to access that catalogue on a server by server basis, and use it to find and display networked electronic resources. Since it is also possible to use the client to search standard Z39.50 OPACs, the process is integrated, as envisaged within the CATRIONA model, with that of retrieval of information on traditional hard copy resources, and with that of finding and accessing locally-held or subscribed-to electronic information resources.

At least one other library system vendor appears to have a Z39.50 client with similar capabilities. The OPAC client in the Dynix Horizon system (Version 3.2) has been shown to have facilities which should enable it to load Netscape in response to a USMARC record containing a URL in the 958 field and of retrieving and displaying an electronic information object (EIO) located at a remote site, although this has not actually been observed working within the CATRIONA project. Dynix do not yet utilise the 856\$u subfield, but are expected to make changes in the near future. Since the client is capable of conducting a search of remote Z39.50 OPACs and retrieving records with URLs in 856\$u, it should have similar capabilities to the Geac client once the field utilised for the URL has been changed. A useful feature of the Dynix client not at present found in GeoPac is its use of

a .ini file to specify the EIO viewers. This should allow a wide range of different 'viewers' to be specified for a range of different EIOs and is therefore more flexible and potentially more powerful than the present GeoPac approach.

It seems possible that a client of this kind - a Z39.50 client able to load a range of different NIDR (Network Information Discovery and Retrieval) clients in response to specific types of URL, and backed up by a range of Z39.50 catalogues describing electronic resources - would be a serious candidate to become a key network navigation tool, perhaps in preference to other specific NIDR client types (WWW or WAIS or Gopher clients) although such clients might well be further developed to encompass Z39.50 compliance.

Although it would be entirely possible to implement a basic demonstrator project without it, further software development is required to create a more sophisticated user-friendly environment for searching a distributed catalogue of the Internet and retrieving/viewing electronic resources envisaged in the full CATRIONA model. An initial list of required features and facilities has been compiled, although it is probably the case that it will only be possible to clearly identify and describe some of the details relating to the refinement of the model itself, and to associated software development requirements, within the context of an actual demonstrator project. Another aim of the demonstrator project would be to test whether the creation of a libraries-based world-wide distributed catalogue of Internet resources constructed from regional building blocks, 'Union' OPACs linked to cooperative cataloguing projects, subject OPACs, and other non-library elements is a practical proposition. Such a project would also help provide a focus for Z39.50 development, implementation and experimentation in the UK.

There is good reason to suppose that a demonstrator based on the full model is feasible. A range of Z39.50 clients have been examined within the project, and whilst none of these has all of the facilities required for the full model, many of these facilities have been seen to work in one client or another. Other requirements for implementing a demonstrator based on the full model are also in place, or will be in future. For example, on the co-operative cataloguing/'Union' catalogue front, it is now clear that both OCLC and RLG will be creating USMARC records describing EIOs (Electronic Information Objects), and it seems probable that co-operative cataloguing of Internet resources will follow these initiatives; indeed the OCLC project entails a co-operative element as do the RLG plans (OCLC, 1994), (Washburn, 1995).

The requirements for a Phase II demonstrator project have been sketched out and costed, although it is true to say that a range of Phase II options exist and that the project could, within certain limits, be tailored to fit a desired level of funding. The position of the commercial companies involved is subject to negotiation regarding the nature of the follow up project. The estimated Phase II cost of 470,000 pounds over 3 years is provided for guidance only, subject to further discussion of the exact nature, shape and timescale of the demonstrator project.

A Phase II project would aim to develop a Scottish distributed network resources OPAC, integrated with existing traditional OPACs (i.e. records in the same catalogue) and accessible through a series of enhanced traditional OPAC clients produced by a number of different suppliers. The end result will be an embryonic but stable operational (non-experimental) system based on the CATRIONA Phase I model - i.e. a system and associated infrastructure capable of supporting a sustainable, distributed and scalable approach to network resource discovery, description, retrieval, and presentation/utilisation, based upon existing library procedures, practices and standards (e.g. Z39.50, MARC). The system would also incorporate library system based access control mechanisms. The system would be 'open' in that other libraries, organisations and suppliers could join at any time, simply by adopting the same, or a similar, model, and it is by this means that more comprehensive library-related catalogues of networked resources might eventually emerge. The possibility that some of the proposed FIGIT subject services should also be based on MARC and Z39.50 should be considered.

The current version of the CATRIONA model (Version 4) is described in Section 3 of the report, with further illustrative detail in Appendix B.

Further information on the CATRIONA feasibility study, and on other related information and projects, can be found on the BUBL Information Service at URLs:

[CATRIONA Web Site](#)
[CATRIONA Gopher Site](#)

[Return to Contents Page](#)

Cataloguing the Internet:

CATRIONA Feasibility Study Report

Chapter 1. Aims of the Study and the Problem in Outline

The stated aim of the CATRIONA (Cataloguing and Retrieval of Information over Networks Applications) feasibility study was to investigate the technical, organisational and financial requirements for the development of applications programs and procedures to enable the cataloguing, classification, and retrieval of documents and other resources over networks, and to explore the feasibility of a library system supplier led collaborative project to develop such applications and procedures and integrate them with one or more existing library housekeeping systems and associated OPAC (Online Public Access Catalogue) interfaces. The full proposal is available at URL: <http://www.bubl.bath.ac.uk/BUBL/catriona.html>

The CATRIONA feasibility study proposal arose out of a desire to tackle, from a library service perspective, four related elements of what is essentially the same problem:

1.1. Providing library users with a reliable and user-friendly means of finding, retrieving or otherwise utilising the growing range of 'Electronic Information Objects' (EIOs) - electronic texts and journals, images, sounds, interactive services, multimedia documents, programs, video, human expertise, and so on - available over the Internet. One that utilises library- standard finding tools and library-standard descriptions and which preferably also encompasses access to information on traditional (i.e. non-electronic) resources.

1.2. The need identified by systems and reference librarians to provide library and campus network users with seamless access, integrated within a single workstation (ideally, of various types: PC, MAC, X-Windows), and mediated through the library OPAC, to a range of electronic information resources held, or subscribed-to, locally, including:

- the library OPAC itself
- CDs and other services on the library CD network
- remote subscribed-to services like BIDS and Embase (See Glossary)

This ideally to extend beyond these elements to encompass both Internet access as described at 1.1 above and, potentially, access to local campus-based electronic resources and to the wider 'virtual library' as outlined at 1.3 and 1.4 below.

1.3. The question of how best to ensure that the growing range of e-resources available at an institution is exploited to its fullest both across campus at the institution itself and, where so desired for income-related or promotional reasons, beyond the local institution via network access - the suggestion being that the best way of doing this may be to catalogue these resources in the network-accessible library OPAC.

1.4. The need to find a practical mechanism for finding, retrieving and 'viewing' or utilising EIOs in the much-heralded distributed 'virtual library', which will simply extend and encompass what is already available over the Internet and which, like the Internet itself, can ultimately only function if it has at its heart a system and associated infrastructure capable of supporting a sustainable, distributed and scalable approach to networked resource discovery, description, retrieval, and presentation.

[Return to Contents Page](#)

Chapter 2. The Proposed Solution: Assumptions Underlying the CATRIONA Model

In proposing the CATRIONA model (described in Section 3 and Appendix B of this report) as a means of dealing with this problem, a number of assumptions have been made about the kind of solution required. These are presented here for three reasons:

1. Because they are essential to an understanding of the reasoning behind the proposed model.
2. Because they give an understanding of both the general and the library-specific problems that any alternative model would have to address.
3. In order that they may be critically examined. If some or all of them are incorrect then this will have a bearing on the future shape of the model.

Many of the assumptions were identified or brought into clearer focus during the project itself.

Assumptions:

2.1. Any solution proposed must be practical, sustainable, scalable and appropriate to the problems faced by library and campus-wide users in any given local context. That is, it must:

- be technically and organisationally feasible and workable
- have a sound maintainable infrastructure underpinning it
- be workable in essentially the same way at the local level, at the Internet-wide level and at any intermediate levels
- be an appropriate and integrated library finding, retrieval and electronic resource presentation tool
- be adaptable to local institutional requirements

2.2. A searchable catalogue (or integrated group of catalogues) of Internet resources containing library-standard resource descriptions is what is required. Finding tools which rely on general subject headings, hierarchical menus, and webs of hypertext links of world-wide scope, and indexes based on menu-entry descriptions are no substitute for full descriptive cataloguing and library-OPAC level search engines. They would not be considered adequate in a single library and certainly cannot be regarded as an adequate answer to dealing with the world-wide resources of the Internet. This is not to suggest that non-library descriptive cataloguing projects based on e.g. the Text Encoding Initiative (TEI) Project (Sperberg- McQueen and Burnard, 1994) do not also have a significant role to play.

2.3. It cannot and will not be done on the scale, or to the level of reliability, descriptive detail, and sustainability required, by volunteers (e.g. BUBL and other subject trees), resource creators, individual services, or publishing companies, although such efforts will continue to play a role in the overall process. This is true in the world of printed resources and there is no evidence to suggest that the world of electronic resources will be significantly different in this respect. Even TEI headers, descriptive metadata for SGML (Standard Generalised Markup Language) documents which, it may be assumed, will be created by authors as an integral part of the electronic document itself, (Giordano, 1994a) are recognised as being limited and likely to require expansion and extension to make them suitable to (local) library purposes. Description of electronic resources held, created, or utilised locally must become a recognised part of the cataloguer's official function and be carried out on a world-wide scale. None of the other methods mentioned above will be sufficient to solve the problem, nor does there seem to be any evidence at present to suggest that automated methods based on resource content can replace the need for descriptive cataloguing.

2.4. The library community has the skills, experience, traditions and organisation required to enable them to solve the various aspects of this problem. They have already begun to respond to the need to catalogue local electronic resources such as CD- ROMs and it is possible that, for purely local reasons, most Internet EIOs important to library users will be catalogued locally. At a basic level, this is all that is required to ensure that the whole Internet is catalogued to a level sufficient for library purposes, and the other types of effort mentioned above should cover most of what is not covered by this means.

2.5. The established practice in the library community of using and contributing to co-operative cataloguing ventures can play a significant role in reducing the cost and burden of cross- Internet cataloguing.

2.6. One big OPAC for the Internet is not a practical proposition for a whole range of organisational, financial, cross-national, and language-related reasons. Apart from problems of scale and of keeping up with demand and

publishing effort and of deciding what should be catalogued and in what order of priority, there is also the question of who would fund it and run it. Would RLG (Research Libraries Group) let OCLC (Online Computer Library Center) go it alone for example, or would they set up in competition? Would one country allow another to be the only host? A joint approach by all organisations and countries might be a possibility, but how many years would it take them to reach an agreement? And what about the problems caused by different languages? Given these kinds of problems, would a centrally funded joint project ever become a reality?

The idea of a single central OPAC (or even several) also has other drawbacks. It could not, by its very nature, be a catalogue of local electronic resources, which means that it would not solve the aspects of the total problem identified at 1.2, 1.3 and 1.4 above, nor would it cover Internet resources authored locally and not more widely publicised. A single central OPAC must also solve the problem of the rapid identification and updating of changed URLs based at remote sites. Although the Internet community is presently investigating ways of dealing with this problem, it is arguably much easier to solve at present if both the catalogue record maintenance and the URL (Uniform Resource Locator) of the resource itself are under local control.

For all of these reasons, it is assumed that the best solution to the problem is not one or several large central OPACs, but a distributed catalogue of Internet resources based on a range of regional or other building blocks, and that the way forward is to design a mechanism and set of protocols and procedures which will allow any library (or other group) to participate and which will allow the whole catalogue to be built up gradually and ultimately be as comprehensive as possible.

2.7. A distributed search of every library and all other OPACs on the Internet is not efficient either in terms of network use or in terms of user (and library workstation) time, so 'Union' catalogues created through co-operative cataloguing ventures should be part of any efficient system - especially since, if these catalogues exist anyway, users will use them if they provide faster resource discovery.

2.8. A distributed catalogue able to be built up gradually as new catalogues 'join' the system MUST be based on standards, especially Z39.50 but also accepted record formats such as MARC (Machine Readable Cataloguing). It must also, of necessity, be an 'open' system, which is to say hospitable to the products of all systems suppliers and, hence, potentially, to all possible library and other participants. (See 3.1.1 for reasons for choosing MARC).

2.9. Even after the proposed distributed catalogue is well-developed and established, other types of Internet metadata - IAFA (Internet Anonymous FTP Archives) templates (Deutsch et al, 1994), TEI headers (Sperberg-McQueen and Burnard, 1994) and searching tools - e.g. Veronica, ICE - will nevertheless continue to be important and necessary for some considerable time, perhaps indefinitely. Any system must therefore encompass the use of such tools. There will be an associated need for reliable regional or national indexing services - Veronica, Archie etc. - presumably funded in the UK by the Joint Information Systems Committee of the Higher Education Funding Councils for England, Scotland and Wales (JISC).

2.10. Since paper-based resources will continue to be important for some time to come, there is a need to integrate searching for electronic resources with hard-copy searches and related paper-based ILL and document delivery services, hence the need to integrate with local OPACs and involve library system suppliers.

2.11. Libraries will wish to offer users access to most Internet and local network resources but they will wish to manage access for a number of reasons (See 3.4 below). Any system must therefore encompass access control mechanisms.

2.12. Access control can be achieved via library system user classification and verification plus control over which remote catalogues, other sources of metadata and so on are catalogued locally

2.13. A client-server model is essential if resources such as images, sounds, video etc. are to be retrieved and displayed or 'played', but VTxxx access is also necessary despite limitations (i.e. text-only) because of limitations on technical capabilities at individual sites and on individual desks.

2.14. The need to cope with resources only accessible through different NIDR clients, such as WWW (World Wide Web) or WAIS (Wide Area Information Servers) clients and with other as yet undeveloped methods of dealing with Internet resources requires that the CATRIONA client must be able to load a range of these and to distinguish between them as appropriate.

2.15. In the longer term, the library community world-wide will build up an information and communication network which will enable them to keep in touch with new publishing developments on the Internet in rather the same fashion as they now do with hardcopy resources. Until this network is established, however, projects like the BUBL Subject Tree Project, with its associated new resource finding and monitoring efforts, will continue to be important if previously unknown EIOs and new editions of remotely produced resources held and catalogued locally are to be identified and catalogued (See Appendix I).

[Return to Contents Page](#)

Chapter 3. The Proposed Solution: Outline Description of the CATRIONA Model

The original CATRIONA model as specified in the funding application documentation (available at URL: <http://www.bubl.bath.ac.uk/BUBL/catriona.html>) was developed and clarified as the study progressed. What is described here encompasses these changes. Some of the details were added after the results of the tests of the feasibility of the model described in Section 5 of the report were known. It is expected that further development and clarification will take place, particularly in the context of the proposed demonstrator project, which is seen as an essential pre-requisite of refining the model in a practical context, but also as a result of continuing discussions with various interested parties and of continuing investigations of related projects.

The following is a list of the main features of the present version of the CATRIONA model:

3.1. Central Elements: Z39.50 Servers and Clients

At the heart of the model are two essential elements:

3.1.1. A distributed catalogue of Internet resources based on regional and other building blocks of the type proposed for CATRIONA Phase II and comprising a series of Z39.50 OPACs sited all over the world. These OPACs would be of two types:

- Catalogues of electronic resources created by institutions, organisations, companies and services. These describe electronic resources which are either held at the local site (purchases, free items downloaded from other Internet sites, or items created/authored locally) or subscribed-to/utilised at the local site (e.g. BIDS, other Z39.50 OPACs describing electronic resources).
- 'Union' catalogues of such resources (e.g. OCLC, RLG) created through co-operative cataloguing and covering a wider range of electronic resources than the average local OPAC. Again, descriptions of hard copy resources will usually be included in the same catalogue.

In both cases, items catalogued will include a wide range of resource types - electronic text, multi-media documents, images, sounds, programs, videos, interactive services, other OPACs etc. In both cases, the records describing the resources utilise the MARC standard enhanced as recommended by MARBI (MARBI, 1993a, 1993d, 1994a). **(Note: These enhancements were developed initially for USMARC and implementation in other MARC formats is limited as yet. See Appendix E - Dutch InfoServices Project).** This means, amongst other things, that they hold in 856\$a a URL or URLs for each electronic resource catalogued, information that can be utilised by clients such as Mosaic or Netscape to locate, retrieve and display or 'play' the resource on a local workstation.

The use of MARC is proposed for a number of reasons:

- adherence to accepted library standards is essential in a distributed 'open' model such as this.
- the Z39.50 standard is essential to the working of a distributed open system and existing Z39.50 clients and servers are designed to handle MARC (not true of, for example, IAFA templates).
- the requirement to integrate the retrieval of EIOs with the retrieval of information on non-electronic resources

means that it is important to choose an accepted library standard for non-electronic resource description, which MARC is.

- the requirement that any system based on the model be sustainable implies that organisations like libraries be involved in the cataloguing side of the model, and this in turn implies that any standards adopted be library standards.

3.1.2. A series of enhanced Z39.50 clients capable of:

- Conducting parallel distributed searches of all of the Z39.50 OPACs which make up the core of the distributed catalogue of the Internet resources described at 3.1.1 above.
- Retrieving enhanced MARC records containing URLs in 856\$.u.
- Responding to these URLs by offering the user the option of loading an appropriate NIDR client - e.g. WWW, Gopher, WAIS clients - and of retrieving and displaying or playing the electronic resource on the local workstation.

Since the 'playing' of such things as images, sounds, video and programs requires facilities only available on a microcomputer or workstation as opposed to a terminal, a client-server model is assumed, with at least the presentation or display level under the full control of a GUI (Graphical User Interface) client on a microcomputer or workstation. However, it is recognised that sites and individuals with more limited technical facilities must also be catered for and that a VTxxx client capability with more limited facilities for retrieving and displaying electronic resources (i.e. text-only) is also essential.

3.2 Local Control of URL Changes and URL Updates

The common Internet problem of a changed or 'broken' URL is dealt with by ensuring that all resources catalogued are essentially under the local control of the site doing the cataloguing. In the CATRIONA model, the updating of changed URLs in catalogue records will be managed as follows:

- If URLs for major services like BIDS change, this will be widely known and local cataloguers will update the appropriate records as a result.
- Local catalogues will only contain records for major BIDS type services and for locally held electronic resources. Changes of URLs will therefore either be widely known or under local control, allowing rapid and reliable updating of records.
- These processes will be backed up by automated processes for identifying and reporting on 'broken' URLs in the local system.

The model also recognises the need for a mechanism to enable URLs to be updated in 'Union' catalogues - OCLC, CURL (Consortium of University Research Libraries) etc.

One advisor to the project has suggested that this approach does not deal adequately with the situation where the original of the resource held locally is changed at the remote site and that the URN (Uniform Resource Name) to URL resolution service proposed by the Internet Engineering Task Force (IETF) is needed to resolve this problem (Daniel, 1994b), (Mitra et al, 1994). Most experts in this area also appear to take the view that the URN to URL resolution service will offer the basis of a better solution to the problem of broken URLs than the one proposed at present in the CATRIONA Project. For these reasons, it is felt to be essential that the demonstrator project be so designed as to be adaptable to developments in this area. Nevertheless, the following points are felt to be worth noting:

1. It is not at present clear when a stable and reliable resolution service will be available, or, indeed that such a service ever will be available, although work is progressing in this area (Wieder, 1995). This being so, the method of handling the URLs problem proposed above is the best that can be managed in the short term.
2. Even if the resolution service does come into being, the question of how URLs are kept up to date, who will

be responsible for doing this, and what mechanisms will be employed, will be crucial if the service is to be as reliable, consistent and comprehensive as it must be. Arguably, it is not impossible that libraries and library cataloguers will have a major role to play in this process and that the mechanism will be local control and local URL updating, together with some kind of automatic method of updating library URLs in the proposed resolution service. Although an increasing number of electronic resources will be offered by libraries providing users with appropriate access mechanisms, the idea that there will be few if any electronic publications held locally seems to ignore both the need to provide local users with fast and reliable delivery appropriately tailored to local needs and the possibility that many institutions may be responsible for 'authoring' large numbers of electronic resources. It is therefore possible that library catalogues will increasingly contain records describing EIOs created, purchased or downloaded and held at the local institution, and that resolution service URL updates must inevitably involve some mechanism for logging changes in library URLs, so that library cataloguers will almost inevitably have some role to play in this process. Arguably, this will in time be a major role, since it is possible in this world view that significant electronic resources will be held and catalogued at some library somewhere in the world (including, perhaps, in commercial publishers' own catalogues) and there is good reason to suppose that library records of URLs will be reliable and consistent. In this scenario, therefore, it is not impossible that the resolution service might not depend, or partially depend, upon the kind of URL updating process proposed in the CATRIONA model.

3. A URN to URL resolution service would probably be essential to the long term development of the CATRIONA project if only because it seems to offer a practical solution to the problem of electronic documents which reference other electronic documents. It is inconceivable that cataloguers could keep such electronic links up to date if they were referenced by URLs. If, on the other hand, they quoted URNs, viewers such as Netscape could be developed in such a way as to 'collect' associated URLs from the resolution service and locate and display the resource when 'clicked on' by the user.
4. Arguably, a URN to URL resolution service is not an essential pre-requisite for solving the problem of locally held copies of electronic resources where a new edition of the source document has been published. Libraries have their own means of identifying and purchasing new editions of works that they hold and these will also presumably develop in the electronic world through publicity material provided by publishers and services like BUBL. The new edition will be purchased and catalogued. Possibly the old edition will continue to be held rather than replaced since there may well be reasons for providing access to both. If the library did not hold the original locally and referenced it through a URN pointing ultimately at a remote site, access to both versions might not be guaranteed. Where a resolution service might be essential is for referencing relatively minor updates to an edition. The original might contain a URN which, via the resolution service, enabled access to an updates file reliably maintained by the publisher until a full new edition was produced.
5. It is assumed that if parallel searches are conducted, a mechanism for identifying duplicates will be required for the convenience of the user. This is a potential role for URNs if and when they become stable. In the shorter term, however, a pseudo-URN may have to be employed for this purpose.

3.3 Cataloguing OPACs

The assumption behind the CATRIONA project is that the Internet will be catalogued in a distributed fashion. In order to carry out a search of this whole distributed catalogue, therefore, it is necessary for the OPAC client being utilised to be able to locate and search all of the individual OPACs that make up the distributed catalogue. In existing clients this information is programmed into the client itself by the system administrator. However, the assumption in the CATRIONA model is that this information will initially be located on the local OPAC which will contain catalogue records for other OPACs selected by the library and uploaded to the client as required. It is further assumed that any given OPAC will normally only contain a small subset of all of the available OPACs and that much of the information about the rest of the distributed catalogue of the Internet will come from other OPACs and, in particular, the 'Union' catalogues. It is likely that Internet OPACs will be categorised according to their subject strengths and that users will be able to search their own local catalogues for records of other OPACs strong on a particular subject category. The user will then be able to conduct a distributed search of a selection of these OPACs as required. This assumption will have to be critically examined in the context of a demonstrator project. This will imply looking at the concept itself and its practical implementation, but also a comparison with any other solutions to this problem proposed by others (e.g. the WHOIS++ directory service approach proposed in ROADS (ROADS, 1995)). One advantage of the CATRIONA approach is that other OPACs can be catalogued, classified and subject-indexed to take account of local guidance requirements.

3.4 Access to Other Internet Resources

The model assumes that for some considerable time to come, and perhaps indefinitely, there will be many sources of information about the existence, location, and type of important Internet resources that are not part of the core of the CATRIONA model. These will include things like:

- catalogues (possibly Z39.50, but probably not) utilising record formats other than MARC (IAFA templates, TEI headers and others - see Appendix F).
- cross-Internet indexing tools like Veronica, Archie, WWW (World-Wide Web Worm) (See Appendix E and McBryan, 1994).

The model therefore includes the possibility that a CATRIONA client will also be able to offer the user the option in appropriate circumstances of utilising such non-standard services to search for, retrieve and display Internet resources. In the absence of Z39.50, such searches would have to be done on a service by service basis (i.e. not via a parallel/distributed search) and would entail the CATRIONA client invoking an appropriate NIDR client (WWW or Gopher or WAIS or whatever) and allowing the user to utilise this client to conduct the search as required.

It is assumed that reliable and important services of this kind will be catalogued in the local and other OPACs on the Internet in the same way as described at 3.3 above and that information on them and how to access them will be 'served up' to the user whenever the circumstances of the search make them appropriate. It is further assumed there will be a role for JISC- funded services like BUBL, NISS (National Information on Services and Systems) and others in the provision of stable and reliable access to services like Veronica, Archie and WWW.

3.5 Access Control

Access control is regarded as an essential element of the model. This is partly to ensure that only those with the right to do so - either because they are local as opposed to remote users, or because they are one of a particular category of local users - are allowed to access certain services and resources. However, it is also to enable the library both to manage how, when, and in what circumstances, certain resources are accessed and to monitor such access. For example, it is probable that libraries will wish to offer users access to most Internet and local network resources but they will wish to control access for a number of reasons:

- to encourage efficient use of the network by directing to geographically close sites first
- to direct attention to commercial resources with special deals before more expensive alternative commercial resources
- to direct attention to local institutional sites with special agreements before sites further afield
- to save user time through expert guidance and knowledge of local requirements
- to monitor use so as to obtain management information on what users are accessing and under what circumstances

It is assumed that access control can be managed via a combination of user classifications on the local library housekeeping system and selective cataloguing and judicious classification/subject indexing of other Internet catalogues and other Internet indexing services. More work is required in the practical context of a demonstrator project to clarify exact requirements and identify feasible solutions. It is possible that there will be a requirement to develop functionality based on some aspects of Z39.50 Version 2 not yet addressed by library systems suppliers (Access Control and Resource Control).

The model has changed and developed throughout the project. A description of Version 4 of the model illustrating how it would interact with the user in various situations is presented in Appendix B, together with the development requirements appropriate to each stage.

[Return to Contents Page](#)

Chapter 4. Work Done, Problems Encountered, Limitations of the Project

It is important to note the limitations of the study described in this report. As is clear from the summary of work done presented below, the topic being addressed encompasses many wide-ranging concerns. In addition, it is also the subject of much current discussion and is undergoing constant change. In contrast, the time available for investigation was short, partly because the project as proposed was of short duration (23 weeks in total, with a significant proportion required for report preparation), partly because of a number of additional unpredicted factors (See 4.1 to 4.5 below). Some of these factors meant that it was sensible to alter some aspects of the planned course of the project in order to best fulfil its aims (See Section 1), others simply meant that some of the predicted work took longer than expected. Both types had the effect of further reducing the total time available.

These limitations on the time available for the project have meant that, whilst all or most of the various topics germane to the enquiry have been addressed to some extent, only core concerns relating to the feasibility of the model have been investigated in any detail. Another consequence is that the report is necessarily a snapshot of progress at a particular point in time. More work needs to be done on the wider field of topics and projects and this is one of several reasons to consider a follow-up demonstrator project.

Unpredicted Factors:

4.1. Obtaining feedback from the library community, the wider Internet community, and, to a lesser extent, from system suppliers and similar organisations, proved more difficult than anticipated. Some of this was identified as arising from difficulty in understanding the model and a great deal of unplanned work was done to clarify the written description both through re-writing it and through addressing the question at various meetings.

4.2. When the project became public knowledge a large number of library system suppliers not originally involved in the project indicated that they felt they should have been involved. For a number of reasons, it seemed wise to involve them, and doing so proved useful on a number of counts, but it also inevitably increased the work that had to be carried out within the project.

4.3. The need to install and look at an associated number of Z39.50 clients, and the inevitable extension of technical difficulties that this caused, also increased the work that had to be carried out, but also helped identify and assess the feasibility of CATRIONA Phase II development requirements.

4.4. A key assumption - that the client required for CATRIONA would be created by enhancing the internal functionality of existing OPAC clients proved to be misconceived. It became clear that the best approach to this was for such clients to be able to call up an appropriate NIDR client (e.g. Mosaic, or a WAIS client), that some suppliers were already moving in that direction and others planned to do so. This meant that it became sensible to concentrate on the features of existing Z39.50 clients rather than on those in NIDR clients as originally planned. It also meant that suppliers felt no need to discuss the subject of NIDR client capabilities with NIDR client developers as originally planned.

4.5. The discovery of a Z39.50 client able to act on a URL in a USMARC record meant it was sensible to attempt to find a Z39.50 site with records of this kind in their database and to set up the client so that it could inter-operate, again a time-consuming process. A message requesting information about Z39.50 servers with URLs in 856\$u was sent to the USMARC, Z3950IW and GILS lists (See References). Only one of the responses contained information which was sufficient to make a Z39.50 connection to the server and access the records. Each of the other responses required further follow up for further details of site address, search type required and encoding of the 856\$u subfield.

Summary Of Work Carried Out:

Survey and examination of hard-copy and electronic literature relating to various aspects of the project:

- network navigation issues
- cataloguing - developments in USMARC, TEI headers, IAFA templates, URC etc. (See Appendix F)
- Uniform Resource Names, Uniform Resource Locators, Uniform Resource Characteristics (URNs, URLs, URCs)
- Internet cataloguing projects
- Z39.50 protocol (See Glossary) and implementation
- Z39.50 clients and servers

- NIDR client developments

Creation of a wide range of hypertext references on the BUBL WWW and Gopher servers:
Available at the following URLs:

<http://www.bubl.bath.ac.uk/BUBL/maincatriona.html>
<gopher://bubl.bath.ac.uk:7070/11/Link/Catriona>

Development of CATRIONA model Versions 2, 3 and 4 (Version 1 in original bid. Available at <http://www.bubl.bath.ac.uk/BUBL/catriona.html>).

Discussion of model over e-mail and at face-to-face meetings with systems suppliers: Geac, Dynix, FDG, SLS, SCG, MDIS. And with cataloguing services: CURL, OCLC, RLG.

Dissemination of information relating to the project through BUBL and e-mail discussion lists. (LIS-LINK, LIS-BAILER, Z3950PIG etc.).

Presentations of the model and cataloguing issues as well as panel discussion at the Cataloguing and Indexing Group Scotland meeting (See Appendix K).

Discussion of Z39.50 issues and attendance at a number of related meetings (See Appendix H).

Survey, installation and examination of a number of existing Z39.50 OPAC clients: GeoPac Release 1.23, WinPAC Beta 1.0, Horizon 3.2, VIZION, Telnet access to Libertas 6.3 and Innopac clients. (Telnet provides a connection to a remote computer over the Internet or a remote login). Observation of Oracle Libraries client and MDIS Lion prototype client during visits.

Survey, installation and examination of a number of NIDR clients: Cello, MS windows Mosaic, Netscape, Hgopher, WinWais and others.

Testing the feasibility of the CATRIONA model at a basic level.

Design of proposed demonstrator project.

Formulation of development requirements for proposed demonstrator project.

Discussion of likely costs of CATRIONA Phase II with selected suppliers.

Preparation of report

Further information is provided in Appendix C under the above headings.

[Return to Contents Page](#)

Chapter 5. Feasibility of the Model and of a Follow-up Demonstrator Project

The stated aim of the CATRIONA feasibility study was to **investigate the technical, organisational and financial requirements for the development of applications programs and procedures to enable the cataloguing, classification, and retrieval of documents and other resources over networks, and to explore the feasibility of a library system supplier led collaborative project to develop such applications and procedures and integrate them with one or more existing library housekeeping systems and associated OPAC interfaces.** A particular model for cataloguing network resources over the whole Internet was proposed (For Version 1 of this model see URL: <http://www.bubl.bath.ac.uk/BUBL/catriona.html>). This assumed a distributed catalogue of Internet resources with distributed cataloguing, local control of record updates and EIO location, and a key role for co-

operative cataloguing and 'Union' catalogues through organisations such as OCLC, CURL and BLCMP. Initially, it had been assumed that the feasibility of the model and the proposed demonstrator project would be examined through critical analysis and discussion of the model and of technical, cataloguing, and software development requirements with library system suppliers, NIDR client developers, systems and reference librarians and other groups with interests in this area. However, as the study progressed it became evident:

1. That USMARC cataloguing standards had already been enhanced (MARBI, 1993a, 1994a) to enable descriptive and, more importantly in the CATRIONA context, location information to be recorded for electronic resources. In particular, the 856\$u subfield had been earmarked for recording a resource's Uniform Resource Locator or URL, a method utilised by WWW clients like Mosaic and Cello to locate, retrieve and display or 'play' an electronic resource. (MARBI, 1993d).
2. That at least one existing Z39.50 OPAC client (an unreleased version of Geac's GeoPac client) already had the capability to conduct Z39.50 searches of remote databases and to respond to a URL in the 856\$u subfield by automatically offering the option to load a NIDR client, pass it the recorded URL and thus enable it to locate, retrieve and display or 'play' a resource stored at the URL in question. The client also had the capability to open a Telnet session to a remote service in response to a 'Telnet' URL.

This raised the possibility that a more practical test of the feasibility of a demonstrator project based on the model could be conducted. GeoPac was installed on an IBM compatible PC running a Winsock compliant version of PCNFS 5 and shown initially to be capable of retrieving a catalogue record on a database on the same machine, responding to a local URL by offering the option to link to the image described in the record, then loading an image viewer to display the locally held image. The next step was to install Mosaic for MS Windows on the machine and set up GeoPac so that it could locate and load Mosaic as required. This done, it was necessary to identify one or more remote Z39.50 OPACs describing electronic resources and containing URLs. Messages were sent to a number of electronic mail discussion lists (USMARC, Z3950IW and GILS) and three OPACs - at Brigham Young University, Butler University and North Carolina State University - containing URLs were identified. Z39.50 connection details were obtained and GeoPac was set up so as to be able to access the OPACs. Details of the records themselves and how to search for them were also obtained and an attempt was made to connect to each of the databases in turn to retrieve the enhanced USMARC records and get GeoPac to load Mosaic and retrieve and display the described EIOs.

The first database searched was the BYU server at Brigham Young University, in which there were two records that have been enhanced to include URLs in 856\$u. On retrieval of the records, the client's 'link' button was highlighted and clicking on this had the effect of making the OPAC client load a suitable 'viewer', and pass it the URL so that it could locate, retrieve and display the EIO described by the MARC record. The two records from the BYU server were retrieved in turn and the 'link' button 'clicked on' as described above. GeoPac duly retrieved the EIO described, loading an appropriate viewer to display it. In the first case, the viewer loaded was the WWW client Mosaic and the EIO was a full text article on a gopher server in Minneapolis. In the second case the viewer loaded was a .gif image viewer and the EIO was a .gif file located at Brigham Young University. The second database searched was at Butler University Library which has at least 3 records with URLs in the 856\$u subfield. The viewer loaded was Mosaic and the EIO was a full-text HTML format file located at Indianapolis. The third database searched was at North Carolina State University Library, where there are records with URLs in the 856 field, but not in the \$u subfield. In this case the EIO could not be retrieved.

Subsequent tests showed that GeoPac could also load Netscape in preference to Mosaic, although in its present incarnation it can be configured only to run one NIDR client at a time. (Note: it can also open a Telnet session in addition to this and utilise an image viewer if required). The client was subsequently installed on another PC which utilised Beame and Whiteside's Winsock compliant TCP/IP in preference to PCNFS V.5. The tests were repeated with similar results.

These tests and subsequent investigations show that the idea of a distributed catalogue of Internet resources integrated with standard Z39.50 library system OPAC interfaces is already a practical proposition at its most basic level. Since GeoPac can be set up so as to add additional Z39.50 servers to its search list, it would clearly be possible to create a distributed catalogue of Internet resources, to access that catalogue on a server by server basis, and use it to find and display networked electronic resources. Since it is also possible to use the client to search standard Z39.50 OPACs, the process is integrated, as envisaged within the CATRIONA model, with that of retrieval of information on traditional hard copy resources, and with that of finding and accessing locally-held or subscribed-to electronic information resources.

At least one other library system vendor appears to have a Z39.50 client with similar capabilities. The OPAC client in the Dynix Horizon system has been shown to have facilities which should enable it to load Netscape in response to a USMARC record containing a URL in the 958 field and to retrieve and display an EIO located at a remote site. Dynix do not yet utilise the 856u subfield, but plan to make changes in the near future. Since the client is capable of conducting a search of remote Z39.50 OPACs and retrieving records with URLs in 856u, it should have similar capabilities to the Geac client once the field utilised for the URL has been changed. A useful feature of the Dynix client not at present found in GeoPac is its use of a .ini file to specify the EIO viewers. This should allow a wide range of different 'viewers' to be specified for a range of different EIOs and is therefore more flexible and potentially more powerful than the present GeoPac approach.

Although it would be entirely possible to implement a basic demonstrator project without it, further software development is required to create a more sophisticated user-friendly environment for searching a distributed catalogue of the Internet and retrieving/viewing electronic resources as envisaged in the full CATRIONA model (See Section 6 and Appendix B). On the basis of the results of this study there is good reason to suppose that a demonstrator based on the full model is feasible. A range of Z39.50 clients have been examined within the project, and whilst none of these has all of the facilities required for the full model, most of these facilities have been seen to work in one client or another. Other requirements for implementing a demonstrator based on the full model are also in place, or will be in future. For example, on the co-operative cataloguing/'Union' catalogue front, it is now clear that both OCLC and RLG will be creating USMARC records describing electronic information objects (EIOs), and it seems probable that co-operative cataloguing of Internet resources will follow these initiatives; indeed the OCLC project entails a co-operative element, as do the RLG plans for cataloguing electronic resources (OCLC, 1994) (Washburn, 1995).

Note: Appendix D provides further information on the Z39.50 OPACs described above and of the USMARC records retrieved, together with details of the functionality of the Z39.50 clients examined at the time of the study.

[Return to Contents Page](#)

Chapter 6. Summary of Development Requirements

Development here is taken in the widest sense and includes such things as skills development as well as software development.

N.B. The client and server requirements list at 6.1 below arose from a consideration of Version 3 of the CATRIONA model. These development requirements were utilised in the examination of Z39.50 client functionality and in the exercise to produce an estimate of costs in the proposed follow-up demonstrator project. The additional requirements listed at 6.2 below arose out of Version 4 of the model. This is described in outline in Section 3 and in detail in Appendix B and is an enhancement of Version 3 of the model (which is not otherwise produced in this report). Appendix B shows how requirements relate to specific aspects of the model.

6.1 Client and Server Requirements Identified from CATRIONA Model Version 3

Clients

1. PC client (plus, ideally, MAC and X-Windows and VTxxx).
2. Data for accessing other Z39.50 catalogues not on menu can be uploaded from catalogue records.
3. Distributed parallel searching.
4. Group of databases can be selected by user for parallel searching.
5. Number of hits given before records are received.
6. Numbers of hits from several servers are given separately.
7. Warning if result sets are too large.
8. Stop button (does this need to be individual to a server session?).
9. Option to merge results from several servers.
10. Option to sort results (URNs or pseudo URNs?).
11. Filing duplicates together (URNs or pseudo URNs?).
12. Related works search (i.e. more like this title).

13. Ability to call up one of several NIDR clients as appropriate (e.g. WWW, Gopher, WAIS) (use .ini file as in Mosaic).
14. Save search strategy.
15. Access control and charging mechanisms.
16. Servers to appear in catalogue.
17. Ability to conduct distributed (server by server) search for other non-Z39.50 Internet OPACs.
18. Ability to do a Z39.50 OPAC EIO/Internet search only (i.e. to search only for electronic resources instead of hard-copy resources as well).
19. Ability to handle URLs in SUTRS/GRS (Simple Unstructured Record Syntax/Generic Record Syntax - see Glossary) i.e. not in USMARC 856 (client and server requirement).
20. Close down NIDR client session button in OPAC client.
21. Ability to be loaded by e.g. Mosaic (e.g. if Mosaic finds a Z39.50 URL).
22. Ability to offer 'other subject catalogues' option after a subject search.
23. Ability to respond to failed known-item search by offering to search other Internet resources.
24. Ability to automatically search other OPACs in turn for known item and offer to stop searching when hit found.
25. Gateway to search ERL-compliant databases.
26. Ability to load networked CD ROM titles and to handle associated memory requirements.

Servers

1. Ability to support new client functionality as described above.
2. Ability to hold, search for, 'serve up', records e.g. of other Internet OPACs.
3. Any server problems related to distributed searching?
4. Duplicates: can servers support 'serve up' pseudo- URNs?
5. Access control beyond initialisation level.
6. Ability to respond to failed known-item search by serving up all known OPACs to client.

6.2 New Development Requirements from Version 4 of the CATRIONA Model

Client Requirements

- (a) Ability to offer Telnet access to service if and only if user permissions permit.
- (b) Ability to auto logon to remote service using username and passwords.
- (c) Ability to suppress display of username and password stored in (MARC) record except at system administrator permissions level.
- (d) Ability to suppress retrieval/display of records of commercial services when user is not local.
- (e) Ability to display subject category of catalogued remote OPACs.
- (f) Ability to handle multiple URLs.
- (g) Ability to make an intelligent choice between alternative URLs.
- (h) Ability to conduct a subject search for records of other Z39.50 OPACs.
- (i) Ability to copy retrieved records containing URLs from a remote non-Z39.50 OPAC to a user file.
- (j) Ability to convert document containing URLs to HTML (Hypertext Markup Language) in order to allow retrieval of EIOs by NIDR client.
- (k) Ability to load appropriate client to conduct search of Internet indexing services such as Archie or Veronica.
- (l) Ability of server to use a thesaurus to identify general subject terms appropriate to more specific terms entered by user, and to use the general terms to identify appropriate subject catalogues to offer the user as a means of expanding the search.
- (m) Ability to conduct a search for items utilising a URN.
- (n) Ability to copy URN to CATRIONA client and utilise it to conduct a URN search.

Cataloguing Requirements

- (a) Ability to catalogue any and every possible type of Internet resource and service.
- (b) Ability to record username and password of commercial services in suppressible (MARC) field.

- (c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.
- (d) Records must be in a form that can be searched and retrieved using Z39.50.
- (e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.
- (f) Ability to store several alternative URLs in catalogue record.
- (g) Ability to record Z39.50 information needed to set up client-server dialogues with non-MARC Z39.50 OPACs.
- (h) Records must be in a form that is compatible with developing Internet standard.
- (i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.
(a specific example of cataloguing requirement 1).
- (j) Ability to catalogue CD-ROM titles and record access information for them
(a specific example of cataloguing requirement 1).

URL Requirements

- (a) URLs for local electronic resources and remote major services must be kept up to date. It is assumed that this will be handled as described in Section 3.2.
- (b) URLs for remote electronic resources must be kept up to date. It is assumed that this will be handled as described in Section 3.2. Progress towards creating a URN to URL resolution service will be relevant here.

URN Requirements

- (a) Filing duplicates together may require URNs or pseudo-URNs.

Other Requirements

(a) Union Internet Catalogues

The existence of 'Union' Internet catalogues is a requirement. OCLC are currently compiling a catalogue of Internet resources in their project 'Building a Catalog of Internet Resources' (OCLC, 1994). This catalogue will be made publicly available. RLG are planning to add Internet resources to their Z39.50 Zephyr catalogue at the beginning of 1995 (Washburn, 1995). Similar records may be available on the CURL database in future. Given that such cooperative 'Union' catalogues will exist, it is reasonable to assume that they will be utilised as a more efficient way of searching the whole Internet than searching every single local OPAC. In addition, BUBL is currently investigating the installation of the 'Alex' catalogue of Internet resources at Bath. If it is successfully installed, the intention is that it would entail a cooperative element.

(b) Updating URLs in Union Catalogue

The method for updating URLs in Union catalogues is not yet established. An automatic solution may only develop as these catalogues are established. There are already robot-type programs which can check for broken links and some automatic mechanism for updating the links may develop. In the CATRIONA model at present each site would have to send the Union catalogues an updated record when the local URL changes. This is something that a demonstrator project would help to clarify.

If a reliable URN to URL mapping service proves feasible (Daniel, 1994b), this would cease to be a problem, although the problem of how URLs are to be reliably kept up to date in a mapping service still has to be tackled. The possibility that a URN to URL resolution service might utilise CATRIONA-type library catalogues as a stable source of URN to URL mappings is worth considering.

- (c) In order to inter-work with Z39.50 servers containing IAFA templates (say), other projects need to be identified which utilise Z39.50 and IAFA or which can be persuaded to do so. IAFA templates would have to be registered as Z39.50 record format and clients and servers built to handle them.

- (d) Stable national Veronica, Archie and WWW indexing services will be required, possibly at duplicate sites (e.g. BUBL and NISS).

(e) Librarians must have access to information services which enable them to keep up to date with new publications in the electronic world, just as they now have this for the hard-copy world. These would include publishers' catalogues, online and printed, and services like BUBL.

6.3 Additional Notes And Other Requirements

Access Control

Requirements here remain vague, largely due to lack of time. Z39.50 Version 2 is understood to include provision for access control at a more specific level than that usually applied at the initialisation stage (e.g. at the level of an individual database or an individual resource or catalogue record rather than at service access level) but suppliers appear not to have implemented this aspect of version 2 as yet due to lack of an identifiable need. It is likely that a CATRIONA demonstrator would eventually require this more specific level of access control (ANSI/NISO, 1994, Section 3.2.5).

Cataloguing

All of the enhancements required for the CATRIONA model to function at its most basic level are already in place. A particular requirement was the addition of a new USMARC field 856 'Electronic Location and Access'. This was accepted in January 1993 and is now formally part of the USMARC format. The essential subfield for CATRIONA purposes is the \$u subfield for Uniform Resource Locator (URL). A range of other requirements are also already met and others are being discussed by MARBI, the Library of Congress and OCLC. It would obviously be important for a CATRIONA Phase II project to liaise closely with groups discussing electronic resource related enhancements to MARC. Similar liaison with groups dealing with IAFA templates, TEI headers and other record formats would also be important. (See Appendix F).

Z39.50

Some CATRIONA requirements imply the implementation of some aspects of Version 3 (e.g. sorting, SUTRS) and of some as yet un-implemented aspects of Version 2 (e.g. access control, resource control). Many suppliers have already begun to implement aspects of Version 3 but others are adopting a 'wait and see' approach.

In terms of moving the demonstrator beyond the experimental stage and into the embryonic operational system stage, there is a requirement for the installation of a number of Z39.50 catalogues at UK sites. Only one or two such catalogues exist at present, but a large number of libraries are either currently out to tender for new systems or will be within the proposed span of the project and it is almost certain that this requirement will begin to be met within the next six to nine months.

There is also a need to consider whether the proposed FIGIT (Follett Implementation Group for Information Technology) subject services should not also be Z39.50 based to enable them to be integrated fully within a CATRIONA-type distributed catalogue. Ideally, they should also be MARC based so that additional client and server development would not be required to allow existing clients and servers to handle their records. Failing this, the record-types utilised should be registered for use with Z39.50 and costs for the development of client and server functionality for handling these record formats should be built into the CATRIONA project.

(Note: it is known that the initial FIGIT subject services will be based on IAFA, WWW and WHOIS++ technology. In which case integration with CATRIONA may rely upon Z39.50/WWW and Z39.50/WHOIS++ gateways).

Mechanism For Updating URLs

See Section 3.2 of this report.

URNs or Pseudo URNs

It is possible that stable and widely agreed and implemented proposals on Uniform Resource Names (URNs) (Mitra et al, 1994) are a requirement for CATRIONA. One area in which they may be an early requirement is that of the automated identification of duplicates within the proposed parallel and distributed searching context. For this limited purpose, however, a pseudo-URN scheme could easily be substituted. In the limited context of a demonstrator (e.g. Scottish libraries only) this should be relatively easy to administer.

Training

Training and development efforts relating to the cataloguing of electronic resources and to the location of resources to be catalogued (as in, for example, the BUBL Subject Tree project) are a requirement.

Stable Internet Indexing Services

There is likely to be a requirement for the development and maintenance of stable and reliable (and perhaps duplicated) Internet indexing services such as Veronica, Archie and similar WWW services (e.g. ICE, WWW). This would be best approached through JISC who might either fund existing services such as BUBL or NISS to offer such services or introduce new services at other locations.

LIS Communication And Information Network

In the longer term, the Library and Information Science (LIS) community world-wide will build up an information and communication network which will enable them to keep in touch with new publishing developments on the Internet in rather the same fashion as they now do with hardcopy resources. Until this network is established, however, projects like the BUBL Subject Tree Project, with its associated new resource finding and monitoring efforts will continue to be important if previously unknown EIOs and new editions of remotely produced resources held and catalogued locally are to be identified and catalogued.

Organisational Framework For Project Co-ordination

Clearly, any co-operative project of the kind proposed will require co-operation and discussion at the managerial and library systems levels. Fortunately, the proposed participant libraries have an organisational framework in place which will facilitate managerial and systems level co-ordination of a demonstrator project. All are members of the Scottish Confederation of University and Research Libraries (SCURL) group which has several years experience of managing co-operative activity across Scotland, with the experience of the Scottish Committee on Library Automation Requirements (SCOLAR) committee - all chief librarians and systems librarians - being of particular importance.

[Return to Contents Page](#)

Chapter 7. CATRIONA Demonstrator: General Points

7.1 CATRIONA Demonstrator: A Necessary Step

There are a number of compelling reasons for moving beyond this feasibility study to a CATRIONA demonstrator project. It could, of course, be argued that companies are moving in this direction anyway, so why not just wait for them to pay for the development, and that libraries are also moving in this direction and will help drive the companies along, however it should be recognised that:

- this kind of 'natural' development, if it happens at all, won't happen quickly
- a solution to the Internet aspect of the problem is needed soon before it becomes too chaotic to be of value to libraries and their users
- the much-heralded distributed 'virtual library' can ultimately only function if it has at its heart a system an associated infrastructure capable of supporting a sustainable, distributed, and scalable approach to networked resource discovery, description, retrieval, and presentation - again this is needed sooner rather than later
- libraries and other institutions also need an early solution to the other aspects of the problem noted in Section 1

and that, perhaps more crucially:

- many aspects of the model and of associated development and infrastructure problems can only be clarified in a demonstrator so it is quite likely that development will not occur in the absence of a demonstrator

- there is a need to develop UK-based expertise in this area both to back up the libraries who wish to progress in this direction, and to inform and influence suppliers, and as a means of influencing the world at large (e.g. via liaison with Veronica developers, IETF working groups and others)
- that there is a significant requirement for training and awareness development efforts relating to the cataloguing of Internet resources and that the lack of such efforts may well hold back development of a CATRIONA-type operational system
- a clear vision of the future as regards the 'virtual library' is needed to drive and, more importantly, focus effort and development in other areas (Z39.50, JISC policies etc.)
- there is a need to encourage libraries to commit resources in these areas
- as indicated earlier in this report, additional work is needed on a wide range of topics related to network resource description and retrieval
- it is necessary to encourage libraries to support resource discovery efforts of the kind seen in the BUBL subject tree efforts
- a Z39.50-based demonstrator which integrates with hard-copy descriptive efforts is required to focus Z39.50 implementation development

7.2 CATRIONA Demonstrator: An Open System

The CATRIONA model is 'open' in three senses:

7.2.1 Because the Z39.50 clients and server programs utilised in the model can and, it is assumed, will come from a range of different suppliers, mainly library system suppliers (Geac, Dynix, SLS etc.), and because any organisation, institution, company or service can become part of the distributed catalogue simply by utilising Z39.50 protocols and the enhanced MARC standard and by letting the world know how to access their catalogue.

7.2.2 Because the model allows other solutions (IAFA template catalogues, Veronica, Archie and so on) to be incorporated and utilised where necessary or appropriate, albeit in the limited ways described in Appendix B, stages 6 and 7. Full integration requires the full integration of the services into the Z39.50 universe (which is proposed here) or special development of Z39.50 client functionality outwith Z39.50 (which is not at this stage proposed here). It is not known to what extent gateways will allow full integration into the Z39.50 universe (e.g. the Z39.50/WHOIS++ gateway alluded to as a possible future development in the ROADS FIGIT proposal) (ROADS, 1995) but the model will also be open to such possibilities.

7.2.3 Because the latest version of the model described in Appendix B allows for any NIDR client, past, present or future to be incorporated and utilised in the network navigation/resource discovery process

Any demonstrator project and, indeed, any working system based on the model would therefore necessarily be 'open' also. Other libraries, organisations, services and suppliers could join at any time, simply by adopting the same, or a similar, model, and it is by this means that a more comprehensive catalogue of Internet resources might eventually emerge.

7.3 CATRIONA Demonstrator: An Adaptable Approach

Both the model as it has developed within the context of the project and the proposals for a Phase II demonstrator outlined below assume that the context within which they will operate will remain roughly as they are at present. It is obviously possible, therefore, that the model and the proposed demonstrator may have to be adjusted to take account of relevant new developments as they arise. For example, if a Z39.50/MARC 'catalogue of catalogues' were to be created by an organisation such as OCLC this might well affect how the CATRIONA model/demonstrator is designed to handle access to other OPACs. Developments in the area of URLs, URNs and URCs might also have a bearing on the detailed working of the model and the demonstrator. It is recognised that an adaptable approach to the development of the model and of the demonstrator is essential.

[Return to Contents Page](#)

Chapter 8. CATRIONA Demonstrator: Specific Phase II Proposals

8.1 Overview

CATRIONA Phase II would seek to implement the basic core of a small-scale working system based on the CATRIONA model. Development, implementation, cataloguing and assessment would take place at a number of local and central sites and would in due course entail all of the following:

- local cataloguing of locally-held electronic resources, incorporating URLs and other descriptive details within the enhanced USMARC standard. Pseudo-URNs might also be incorporated. Items catalogued would include free Internet resources significant to the local institution and also in time digitised items entailed in the SCURL/Stirling University FIGIT project. Cataloguing would ultimately occur at different sites utilising different library systems and specialising in different subject areas as regards coverage of resources not created locally and would be backed up by subject-based Internet resource discovery efforts, based around support activities related to the BUBL subject Tree. The value of enhancing records through the use of contents pages would also be investigated.
- the enhancement of a range of standard OPAC clients to give them the facilities described in Section 6 above, including (where necessary) facilities to enable the location, retrieval and display of electronic resources (at least one client and possibly two already has these facilities).
- the utilisation of distributed searching (e.g. of all Scottish Z39.50 OPACs and any FIGIT subject services utilising Z39.50 and MARC) and of cooperative cataloguing into larger central databases (e.g. OCLC, RLG, CURL) as the mechanism for (eventually) allowing all significant Internet resources to be found and accessed by the library OPAC user
- the cataloguing of a locally determined set of other catalogue servers as the mechanism for facilitating the distributed searching model
- investigations of access control utilising user classifications and control of which resources are catalogued
- user and librarian tests of the resulting system, with these feeding back into development
- liaison and discussion with external organisations as necessary and appropriate
- development of mapping programs to enable records from non-Z39.50 FIGIT subject servers to be brought into the Z39.50 universe

The aim would be to move, over a three year period, towards an embryonic but stable operational (i.e. non-experimental) system based upon the CATRIONA model or some adaptation of it - i.e. a system and associated infrastructure capable of supporting a sustainable, distributed and extendible approach to networked resource discovery, description, retrieval, and presentation, based upon existing library procedures, practices, and standards (e.g. Z39.50, MARC) and incorporating library system based access control mechanisms.

It is not envisaged, however, that it would be possible to move directly to the installation of the fully operational system immediately. Rather, a three year programme is envisaged, as indicated below.

8.2 Phase II: Year 1

The aim in Year 1 would be to pursue and develop the various threads outlined below with a view to establishing an initial platform from which the move to the distributed model envisaged could be more easily managed:

8.2.1. Identification of a representative mix of network resources for cataloguing, utilising the BUBL subject tree, local resource identification at Scottish libraries, liaison with FIGIT projects creating new resources. (Including, in time, the Stirling On-Demand Publishing project). If implemented at BUBL, items from the Alex catalogue might also play a part.

8.2.2. Subject-area based division of resource discovery duties between Scottish libraries and other external groups (e.g. SOSIG, OMNI) based on the BUBL/NISS LIS-subjects approach, in order to help identify network resources for cataloguing and build an infrastructure for future co-operative activity in this area. Liaison, co-operation and, if

possible, shared cataloguing agreements with organisations such as OCLC and RLG.

8.2.3. The storage of e-resources identified for cataloguing and not already located either at the site of a participating library or on BUBL on a centrally located experimental server.

8.2.4. Creation of network-accessible MARC-based catalogue of network resources so identified on the same centrally located experimental server. This to be Z39.50 compliant but also to allow, via a link to BUBL, more basic x.29 and Telnet access to support the needs of the widest possible range of HE institutions (Telnet from BUBL to a VT session on supplier Z39.50 server, for example). Access via WWW would also be investigated.

8.2.5. Development of skills and experience in project assistants and transfer of these to staff in participating libraries (cataloguing, systems, reference) by means of on-site training programmes.

8.2.6. Working in conjunction with a range of library, library systems supplier, library bibliographic utility organisation partnerships, establish technical and other requirements for developing facilities and skills at individual participating libraries to the level required for active participation in the proposed distributed system in years 2 or 3 (or earlier if this is possible).

8.2.7 Further investigation at two or more sites of existing client capabilities and software development requirements with non- parallel distributed searching being utilised and examined, although at this stage only one central CATRIONA Z39.50 OPAC would exist, other sites will exist elsewhere (Brigham Young and perhaps OCLC and RLG) and some Scottish universities (e.g. Strathclyde) will be implementing Z39.50 OPACs. In addition, there may be FIGIT subject services utilising Z39.50 and MARC if the recommendations of this report are heeded.

8.2.8 If appropriate, initial development of client functionality in the areas described in Section 6 of the report (not yet prioritised). However, development requirements may have to be more clearly defined in Year 1 in consultation with suppliers.

8.3 Phase II: Years 2 and 3

8.3.1 Move towards distributed system based on the CATRIONA model as quickly as is practical in the case of each individual library, supplier and utility grouping (in some cases this may be possible in Year 1).

8.3.2 As far as is appropriate, distribute centrally-held resources and associated enhanced records to local sites from central server (essential to model that sites only hold records of items stored locally, but a reliable mechanism for informing sites of URL changes at BUBL (or indeed other JISC services such as NISS) might be a workable alternative, in which case only catalogue records would be distributed in many cases.

8.3.3 Contribute all centrally held records to various participating co-operative cataloguing utilities to establish the basis of a working system based on the model.

8.3.4 Train local staff to participate in the use and further development and growth of the system.

8.3.5 Turn what remains of the BUBL-based central catalogue and document server into a subject service covering LIS and associated areas such as general reference tools (Internet-wide directories, electronic glossaries, dictionaries etc.)

8.3.6 Identify Z39.50 OPACs outwith the project which might usefully be encompassed within the distributed catalogue.

8.4 Proposed Deliverables (over 3 years)

8.4.1. A sustainable, distributed, and scalable approach to network resource discovery, description, retrieval and presentation, based upon existing library procedures, practices and standards, and incorporating library system-based access control and monitoring mechanisms.

8.4.2. Enhancement of existing local and 'Union' catalogues to incorporate descriptions of existing and FIGIT-

created network resources and data to enable their retrieval and display on campus and library workstations.

8.4.3. Refinement and/or enhancement of existing library system Z39.50 OPAC clients to enable the integration of the retrieval of traditional resources with that of the retrieval of electronic resources.

8.4.4. An infrastructure for the creation and maintenance of a distributed OPAC system for electronic resources in the UK and, ultimately, the wider world.

8.4.5. Increased skills and awareness in library staff of the requirements of networked resource discovery, description and retrieval

8.4.6. Improved service to the communities served by UK academic libraries.

8.4.7. Improved understanding of management issues relating to co-operative approaches to network resource access

8.4.8. Help to establish a world-wide UK library influence in areas relating to cataloguing networked electronic resources and finding, retrieving and displaying networked electronic resources.

8.4.9. Advise JISC on related requirements with regard to the network navigation support efforts required from centrally- funded JISC services (e.g. stable and duplicated Veronicas, Archies and similar WWW indexing services)

[Return to Contents Page](#)

Chapter 9. CATRIONA Phase II Requirements: Skills, Personnel, Organisation

9.1 CATRIONA Phase II Requirements: Skills, Personnel, Organisation

Experience and Expertise

What has emerged from CATRIONA Phase I is that most of the expertise and experience likely to be required in a CATRIONA Phase II demonstrator project is already available within the group of individuals, libraries, organisations and, in particular, system suppliers involved in Phase I. The companies and organisations involved with the group include a number of library system suppliers with experience and expertise in the development of Z39.50 based clients with distributed searching capabilities and NIDR client inter-working, together with a number of organisations involved in the provision of co-operative cataloguing utilities. Clearly, any co-operative project of the kind proposed will require co-operation and discussion at the managerial and library systems levels. Fortunately, the proposed participant libraries have an organisational framework in place which will facilitate managerial and systems level co-ordination of a demonstrator project. All are members of the SCURL group which has several years experience of managing co-operative activity across Scotland, with the experience of the SCOLAR committee (all chief librarians and systems librarians) of particular importance. In addition, Strathclyde University Library has specific expertise and experience arising out of its management of the BUBL Information Service, the BUBL subject tree, and the CATRIONA Phase I study. Napier University Library and National Library of Scotland were also closely involved in CATRIONA Phase I.

Potential Partners And Proposed Roles

The companies and organisations involved in CATRIONA Phase I are all potential partners, although some have indicated minimal levels of involvement and others more significant levels. The participants are listed in Appendix J.

9.2 CATRIONA Phase II Requirements: Estimating Costs

Problems Estimating Phase II Costs

Software Development

It has not been possible within the context of the feasibility study to establish the likely costs of the software development required for Phase II with any certainty. From exploratory investigations carried out to date it is clear that the cost will be dependent upon a number of factors and that it will only be possible to determine it with any certainty within the context of negotiations between the funding body or bodies, the companies interested in participating, and the CATRIONA Phase II proposers. It may also be necessary to utilise Year 1 to specify development requirements more precisely. In this event, the companies involved might wish to charge consultancy fees in Year 1, rather than software development fees.

Factors likely to affect the cost of software development:

- Obtaining an exact specification of requirements, something which may only be obtainable through closer examination in the context of Year 1 of the demonstrator project.
- The number of suppliers involved and the present state of development of their client and server software.
- The extent to which required developments are already 'in the pipeline' or are regarded as being likely or required developments from a commercial viewpoint. In the latter case companies will charge less or nothing depending upon required timescales.
- The timescale within which the developments are required to be made (if the developments are spread out and do not require continuous commitment then the costs will be less).
- Decisions as to how much of the identified development is regarded as essential and how much is regarded as peripheral.
- The position taken by the funding body or bodies on the level of contribution to costs likely to be required from participating commercial companies, together with the view taken by participating companies of the importance or otherwise of being involved in the project.

To a lesser extent, there is a degree of uncertainty about some of the other costs involved, with the cost of travel and peripheral equipment likely to vary according to the exact shape and form agreed for the follow-up project and the position of funding bodies and libraries on funding relating to peripheral equipment such as microcomputers.

The estimates detailed below are therefore only a rough guide to the likely costs of a demonstrator project. They assume company consultancy fees in Year 1 and software development costs in years 2 and 3.

[Return to Contents Page](#)

Chapter 10. CATRIONA and the BUBL Subject Tree Project: Further Information

Further information on the CATRIONA feasibility study, on other related information and projects, and on the BUBL Subject Tree will be found on the BUBL Information Service at URLs:

<http://www.bubl.bath.ac.uk/BUBL/maincatriona.html>

<gopher://bubl.bath.ac.uk:7070/11/Link/Catriona>

<http://www.bubl.bath.ac.uk/BUBL/Tree.html>

<gopher://www.bubl.bath.ac.uk:7070/11/Link/Tree>

This information will be added to and enhanced by the BUBL Information Service as information vital to the library

community and its users regardless of whether or not the CATRIONA project continues into Phase II and beyond.

More information on the Subject Tree project and its relationship to the CATRIONA project is provided in Appendix I.

[Return to Contents Page](#)

References

ANIR (1994). The Working Group on Access to Networked Information Resources of the Information Services Sub-Committee of the UK Higher Education Funding Councils. Report.

ANSI/NISO (1994). ANSI/NISO Z39.50-1994, Information retrieval application service definition and protocol specification. Available from <ftp://ftp.loc.gov/pub/z3950>

Bowman C M et al (1994). Harvest: a scalable, customisable discovery and access system. Technical Report CU-CS-732-94, Department of Computer Science, University of Colorado, Boulder. Available from <ftp://ftp.cs.colorado.edu/pub/cs/techreports/schwartz/Harvest.FullTR.ps.Z>

Christian E (1994) The Government Information Locator Service (GILS): report to the Information Infrastructure Task Force. May 2, 1994. Available from <ftp://info.er.usgs.gov/public/gils/gils.txt>

Daniel R Jnr (1994). URC scenarios and requirements. Internet draft. November 21, 1994 (Work in progress). Available from <ftp://nic.nordu.net/internet-drafts/draft-ietf-uri-urc-req-00.txt>

December J (1994) New spiders roam the Web. Computer-Mediated Communication Magazine 1: 3. Available from <http://www.rpi.edu/~decemj/cmc/mag/1994/sep/spiders.html>

Dempsey L (1994) Network resource discovery: a European library perspective. In Libraries, networks and Europe: a European networking study. N Smith (ed) London: British Library Research and Development Department. Available from <gopher://ukoln.bath.ac.uk/00/Publications/ResDes/europe.rtf>

Deutsch P et al (1994). Publishing information on the Internet with anonymous FTP. Internet draft. September 1994 (Work in progress). Available from <ftp://nic.nordu.net/internet-drafts/draft-ietf-iiir-publishing-02.txt>

Dillon, M et al (1993). Assessing information on the Internet: toward providing library services for computer mediated communication. Dublin, Ohio: OCLC.

Gaynor E (1994). Cataloging electronic texts: the University of Virginia Library experience. Library Resources and Technical Services 38: 403-413.

Giordano R (1994a). Notes on operations: the documentation of electronic texts using Text Encoding Initiative headers: an introduction. Library Resources and Technical Services 38: 389-401.

Giordano R (1994b). URLs, etc. TEI discussion list. 29 December, 1994. Available email: TEI-L@UICVM.EARN.

Herr-Hoyman D (1994). Re: Library standards and URIs. URI discussion list. 29 December, 1994. Available email: URI@BUNYIP.COM.

Koster M (1994). Re: URCs and IAFA templates. URI discussion list. 13 October, 1994. Available email: URI@BUNYIP.COM.

Hazelden L (1994). SilverPlatter Information Ltd. Personal communication, 19 December, 1994.

MARBI (1991a). Discussion paper no. 49: dictionary of data elements for online information resources. Library of

Congress. May 1, 1991. Revised July 15, 1991.

MARBI (1991b). Discussion paper no. 54: providing access to online information resources. Library of Congress. November 22, 1991. Revised April 13, 1992.

MARBI (1993a). Proposal no. 93-4: changes to the USMARC bibliographic format (computer files) to accommodate online information resources. Library of Congress and OCLC Internet Resources Project. November 20, 1992. Revised March 29, 1993. Available from <gopher://marvel.loc.gov:701/00/.listarch/usmarc/93-4.doc>

MARBI (1993b). Discussion paper no. 69: accommodating online systems and services in USMARC. Library of Congress and OCLC Internet Resources Project. April, 30 1993. Revised August 4, 1993.

MARBI (1993c). Proposal No. 94-2: addition of subfields \$g and \$3 to field 856 (electronic location and access) in the USMARC holdings/bibliographic formats. Library of Congress. December 6, 1993. Available from <gopher://marvel.loc.gov:701/00/.listarch/usmarc/94-2.doc>

MARBI (1993d). Proposal no. 94-3: addition of subfield \$u (Uniform Resource Locator) to field 856 in the USMARC Bibliographic/Holdings Formats. Library of Congress. December 6, 1993. Available from <gopher://marvel.loc.gov:701/00/.listarch/usmarc/94-3.doc>

MARBI (1994a). Proposal no. 94-9: Changes to the USMARC Bibliographic Format to Accommodate Online Systems and Services. Library of Congress and OCLC Internet Resources Project. May 6, 1994. Revised July 20, 1994. Available from <gopher://marvel.loc.gov:701/00/.listarch/usmarc/94-9.doc>

MARBI (1994b). Proposal no. 95-1: Changes to Field 856 (Electronic Location and Access) in the USMARC Bibliographic Format. Library of Congress and Federal Geographic Data Committee. December 2, 1994. Available from <gopher://marvel.loc.gov:701/00/.listarch/usmarc/95-1.doc>

McBryan O (1994) GENVYL and WWW: tools for taming the Web. In: First International World-Wide Web Conference. Advance proceedings. May 25-27. Geneva.

Mealling M (1994). Encoding and use of Uniform Resource Characteristics. Internet draft. July 8, 1994 (Work in progress) Available from <http://www.acl.lanl.gov/URI/archive/uri-94q3.messages/21.html>

Mitra, Weider C & Mealling M, (1994) Uniform Resource Names. Internet draft. October 20, 1994. (Work in progress). Available from <ftp://nic.nordu.net/internet-drafts/draft-ietf-uri-resource-names-03.txt>

Neuss, C and Hofling, S (1994) Lost in hyperspace? Free text searches in the Web. In: First International World-Wide Web Conference. Advance proceedings. May 25-27. Geneva.

OCLC (1994). Building a Catalog of Internet Resources project. Available from <http://www.oclc.org/oclc/man/catproj/catcall.html>

ROADS (1995) Private document supplied by Lorcan Dempsey on behalf of the ROADS consortium, 9 March, 1995.

SilverPlatter Information (1994). An Introduction to SilverPlatter's ERL technology.

Sperberg-McQueen C M and Burnard L eds.(1994) TEI P3: Guidelines for electronic text encoding and interchange. Oxford and Chicago: The Text Encoding Initiative. Available from <http://etext.virginia.edu/TEI.html>

van der Werf T (1994). InfoServices: cooperation between the national research network service and the National Library in the Netherlands. *Journal of Information Networking* 2: 13-22.

Washburn B (1995). Research Libraries Group. Personal communication, 10 January, 1995.

Weibel S (1994). OCLC. Outline notes for address to IETF Meeting in San Jose, December 1994. Personal communication and subsequently posted on URI list, 12 December 1994. Available email: URI@BUNYIP.COM.

Weider C & Deutsch P (1994) A vision of an integrated internet information service. Internet draft. July 31, 1994. (Work in progress). Available by email from ds.internic.net Send the following message: document-by-name RFC1727

Weider C (1994) The Internet anonymous FTP archive templates: towards an Internet resource location system. Journal of Information Networking 1: 256-260.

Weider C (1995) Bunyip Information Systems Inc. Personal communication, 3 February 1995.

VINE (1994) VINE:Theme issue: Z39.50 and SR. 97.

Discussion Lists

Lists and List Addresses

CATRIONA: private list, USMARC: listproc@loc.gov, INTERCAT: listserv@oclc.org, Autocat: listserv@ubvm.cc.buffalo.edu, E-Media: Emedia-Request@vax1.elon.edu, LIS-CIGS: mailbase@mailbase.ac.uk, TEI: listserv@uicvm.uic.edu, IAFA: iafa-request@bunyip.com, URI: uri-request@bunyip.com, Z3950IW: listserv@nervm.nerdc.ufl.edu, GILS: listproc@cni.org, PACS-L: listserv@uhupvm1.uh.edu, LIS-LINK: mailbase@mailbase.ac.uk, LIS-BAILER: mailbase@mailbase.ac.uk, web4lib: listserv@library.berkeley.edu, UNITE: mailbase@mailbase.ac.uk, UPTURN: mailserver@rare.nl

[Return to Contents Page](#)

Appendices

Appendix A. Glossary of Terms and Acronyms

Included in the following list are terms and acronyms used in the report and appendices with the exception of some of the acronyms used in Appendix E: Related Projects, which are given in full in Appendix E.

AACR-2	Anglo-American Cataloguing Rules, second edition
ALIWEB	Archie Like Indexing in the Web (See Appendix E)
ANSI	American National Standards Institute
Archie	An information system for searching the contents of anonymous FTP archives.
BIDS	Bath Information Data Service
BL	British Library
BUBL	BUBL Information Service (formerly: Bulletin Board for Libraries)
CURL	Consortium of University Research Libraries
CWIS	Campus Wide Information System
EIO	Electronic Information Object
EMBASE	BIDS Excerpta Medica Database
FAQ	

	Frequently Asked Questions
FIGIT	Follett Implementation Group for Information Technology, a sub-committee of JISC
FTP	File Transfer Protocol. A protocol designed to allow the transfer of files from one machine to another
Gopher	Client/server Internet publishing and navigation software
GRS	Generic Record Syntax. (Z39.50 record syntax)
GUI	Graphical User Interface
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
IAFA	Internet Anonymous FTP Archive
ICE	A search engine which allows free text searching on a WWW archive (See Appendix E)
IETF	Internet Engineering Task Force
Internet	A world-wide network of networks.
ISBD	International Standard Bibliographic Description
ISO	International Standards Organisation.
ISSC	Information Services Sub-Committee, a sub- committee of JISC
JANET	Joint Academic Network.
JISC	Joint Information Systems Committee of the Higher Education Funding Councils for England Scotland and Wales
LIS	Library and Information Science
Mailbase	The Mailbase service supports electronic mail based discussion lists for special interest groups (UK)
MARBI	American Library Association Committee on Representation in Machine Readable form of Bibliographic Information. MARBI creates and revises standards for the representation of bibliographic information in machine-readable form.
MARC	Machine Readable Cataloguing
NIDR client	Network Information Discovery and Retrieval client (e.g. WWW client such as Mosaic, WAIS client such as WinWais)
NISO	The National Information Standards Organisation. NISO is accredited by ANSI to develop voluntary technical standards for the library, information sciences, and publishing communities.
NISS	National Information on Services and Systems
OCLC	Online Computer Library Center
OMNI	Organising Medical Networked Information (See Appendix E)
OPAC	Online Public Access Catalogue
OSI	

	Open Systems Interconnection.
PIN	Personal Identification Number
ROADS	Resource Organisation and Discovery for Subject-based services
RLG	Research Libraries Group
SALSER	Scottish Academic Libraries Serials
SCOLAR	Scottish Committee on Library Automation Requirements
SCURL	Scottish Confederation of University and Research Libraries
SGML	Standard Generalised Markup Language
SOSIG	Social Sciences Information Gateway (See Appendix E)
SR	Search and Retrieve (ISO 10162/3). An International Standard which is a functionally compatible subset of the US standard Z39.50-1992
SUTRS	Simple Unstructured Text Record Syntax. (Z39.50 record syntax)
TCP/IP	Transport Control Protocol/Internet Protocol
Telnet	Provides a connection to a remote computer over the Internet or a remote login.
TEI	Text Encoding Initiative
UKOLN	UKOLN: the UK Office for Library and Information Networking
URC	Uniform Resource Characteristic. A method of encoding information about a given network resource using attribute/value pairs under development by the URI Group of the IETF
URI	Uniform Resource Identifier. The URI Group of the IETF is developing a URI architecture that is concerned with network resources and their discovery, identification and retrieval (URCs, URNs and URLs)
URL	Uniform Resource Locator. A compact string representation for a resource available via the Internet.
URN	Uniform Resource Name. A globally unique, persistent identifier used both for recognition of and access to characteristics of the resource or access to the resource itself
Veronica	A keyword search utility for gophers world wide
WAIS	Wide Area Information Servers
WWW	World-Wide Web
WWW	World-Wide Web Worm. A resource location tool which locates WWW-addressable resources on the Internet and provides a user interface to these resources (See Appendix E)
X.400	The CCITT recommendation for a message handling system protocol.
X.500	The CCITT recommendation for a Directory Service.
X-Windows	Unix-based Windows system
Z39.50	Z39.50 is a US standard (ANSI/NISO). It is one of a set of standards produced to facilitate the interconnection of computer systems and is concerned in particular with the search and retrieval of

information in databases. Z39.50-1994 has been proposed and is being discussed. Earlier versions of the standard are Z39.50-1988 and Z39.50-1992.

[Return to Contents Page](#)

Appendix B

CATRIONA Model Version 4:

Illustrative Description of User Interaction

During the course of the project, a number of attempts have been made to describe different incarnations of the CATRIONA model for project participants and others. Judging by the evidence of later discussions of the description with some of the participants, none of these attempts has been entirely successful. It is believed that this has been due partly to an attempt to describe all of the model all at once rather than in piecemeal fashion, and partly to a failure to put the description in a real context. What is presented here, therefore, in Version 4, is an attempt to build a complete picture of the model gradually, and to do so in the context of a real user, in this case a student, utilising a system based on the model. (Version 3 of the model description was also based on this approach).

It is worth reiterating that whilst it has been possible to refine and improve the model as a result of research carried out under the project, it is assumed that further adjustments will be necessary in the light of applying the model in a real working situation in a demonstrator project.

The model description is built up in stages by following the options that would be available to a user of a CATRIONA-type OPAC client. There are 14 stages in this illustrative description, each of which is based upon different kinds of circumstances which the user might face:

STAGES 1-8: USER-CONTROLLED ESCALATION OF SEARCH

1. Conduct search of local OPAC only
2. Distributed search but limited to Scottish (Z39.50) library catalogues and other subject catalogues
3. Option 2 plus Union catalogues (OCLC, RLG, CURL, for example)
4. Search of other (MARC) Z39.50 OPACs served up by Union and other catalogues
5. Search of (non-MARC) Z39.50 OPACs containing records in the form of IAFA templates/ TEI headers.
6. Search of non-Z39.50 catalogues
7. Search Internet indexing services e.g. Veronica
8. Search networked CD-ROMs or other databases

STAGES 9-10: SYSTEM-INFLUENCED ESCALATION

9. Failed known item search
10. System response to subject search

STAGES 11-14: OTHER POSSIBILITIES

11. New editions
12. Known URN search
13. Electronic references
14. Z39.50 URLs

Stages 1-8 address a situation in which decisions as to when and how to escalate a search beyond the local catalogue are taken by the user. Stages 9-10 address the situation in which these 'decisions' are controlled or influenced by the system itself. It was felt that the best way of describing the model was to present the user-controlled escalation options first. This should not be taken to imply either that this is the first choice that would be 'offered' to the user or that there is any reason to believe that this would normally be the user's preferred choice. It may well be that most users would prefer to let the system handle such 'decisions'. Stages 11-14 cover some additional situations which are

considered pertinent.

What is envisaged, then, is that the CATRIONA-type OPAC client (which would come in many different supplier-specific forms) would present the user with a number of the usual search options and also an option to 'search the whole Internet' or something of that kind. In steps 1-8 of the description of the model the assumption is that the user has chosen this latter option. As a result she is presented with a list of individual cataloguing and indexing services, one of which is the local OPAC, but most of which are remote. She may choose to search each individually or to mark some or all for parallel or distributed searching.

The services will fall roughly into the categories listed at 1-8 above. Whether or not they will be categorised in this way for the user has yet to be decided. In the illustration, the user escalates the search gradually from 1 to 8. However, she will not necessarily proceed in this way in a real situation. She may well jump to step 3 or 7 on the basis either of her own experience of her current type of problem or on the basis of information provided by the system help.

B.1. STAGE 1: LOCAL OPAC SEARCH

In this illustration, the user chooses to perform a subject search in only the local OPAC. She performs her search in the usual way and retrieves a number of hits, three of which are relevant to her information problem.

The first record describes a hard copy book. She notes the classmark in the usual way, aiming to go to the shelf and retrieve it later.

The second record describes an electronic version of an article by a lecturer at her institution. The MARC record she has retrieved has a URL in the 856 field (subfield \$u) and the CATRIONA client responds to this automatically by placing a 'link' button on the screen. The user points to the 'link' button and clicks the mouse button. The CATRIONA client responds by automatically loading a WWW client such as Windows Mosaic and feeding it the URL for the article. Mosaic locates the article on a local server and transfers it to the user's PC, loading an appropriate viewer to display it if necessary.

The third record describes the BIDS Embase service (the user used a very general subject heading). The user indicates a wish to search this service, at which point the system asks her to identify herself by her barcode or PIN (Personal Identification Number). If she has access rights, the CATRIONA client loads a communications package from the client PC and makes a Telnet connection to the Embase service allowing the user to perform a new search on Embase.

Stage 1 Identified Development Requirements

Client / Server Requirements

- (1) PC client (plus, ideally, Mac and X Windows and VTxxx).
- (13) Ability to call up one of several NIDR clients as appropriate (e.g. WWW, Gopher, WAIS) (use .ini file as in Mosaic).
 - (a) Ability to offer Telnet access to service if and only if user permissions permit.
 - (b) Ability to auto logon to remote service using username and passwords.
 - (c) Ability to suppress display of username and password stored in (MARC) record except at system administrator permissions level.
 - (d) Ability to suppress retrieval/display of records of commercial services when user is not local.

Cataloguing Requirements

- (a) Ability to catalogue any and every possible type of Internet resource and service.
- (b) Ability to record username and password of commercial services in suppressible (MARC) field.
- (c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

URL Requirements

(a) URLs for local electronic resources and remote major services must be kept up to date. It is assumed that this will be handled as described in Section 3.2 of the main report.

B.2. STAGE 2: DISTRIBUTED OR PARALLEL SEARCH OF OTHER 'ORDINARY' OPACS AND/OR CATALOGUES OF SUBJECT- SPECIFIC RESOURCES

Having exhausted the possibilities of the local OPAC, the user chooses to perform a distributed search of a number of remote Z39.50 OPACS catalogued in the local OPAC and offered in response to the user clicking the "search other catalogues" button. In the proposed CATRIONA Phase II project, these would include a number of Scottish Z39.50 OPACS with subject specialisations, together with any available Z39.50 catalogues of subject-specific services. The local library might have a special arrangement with other libraries in a given geographical area (e.g. an agreement to share the work of Internet resource discovery and cataloguing on a subject basis or to purchase or create electronic resources co-operatively on a subject basis) and the CATRIONA client would offer a choice based on such considerations.

The user would 'mark' the catalogues then enter a request which the CATRIONA OPAC client would send to all of the Z39.50 OPACS in parallel. The OPACS would return the results showing the number of hits. The user might then choose to examine results from individual servers. Ideally these will be returned sorted for ease of use (this requires Z39.50 Version 3). Another option that might be available to the user is the option to combine some or all of the result sets into a single sorted set with duplicates filed together. Depending upon the hardware limitations this option may only be available if the result sets fall within certain size limits. Eventually the user would choose a number of records that describe resources relevant to her enquiry. One of these may be a hardcopy resource which in some cases she may be able to 'reserve' or order through Inter- Library Loan. Another may be an electronic resource which will again be located at a URL specified in the MARC record (or, perhaps, ultimately at a URL or URLs specified via a URN). Once again, the 'link' button will appear and if the user clicks on it a suitable client/viewer combination will be loaded by the OPAC client and the resource will be retrieved from the appropriate site and displayed for the user.

Stage 2 Identified Development Requirements

Client / Server Requirements

- (2) Data for accessing other Z39.50 OPACS not on menu can be uploaded from catalogue records.
- (3) Distributed parallel searching.
- (4) Group of databases can be selected by user for parallel searching.
- (5) Number of hits given before records are received.
- (6) Numbers of hits from several servers are given separately.
- (7) Warning if result sets are too large.
- (8) Stop button (does this need to be individual to a server session?)
- (9) Option to merge results from several servers.
- (16) Servers to appear in catalogue.
- (10) Option to sort results (URNs or pseudo URNs?)
- (11) Filing duplicates together (URNs or pseudo URNs?)
- (e) Ability to display subject category of catalogued remote OPACS.

Cataloguing Requirements

- (d) Records must be in a form that can be searched and retrieved using standard protocols such as Z39.50.
- (e) Ability to catalogue and subject index other Z39.50 OPACS and record Z39.50 information needed to set up client-server dialogues.
- (f) Ability to store several alternative URLs in catalogue record.

URL Requirements

- (b) URLs for remote electronic resources must be kept up to date. It is assumed that this will be handled as described in Section 3.1.2.3 of the main report. Progress towards creating a URN to URL resolution service will be relevant here.

URN Requirements

(a) Filing duplicates together may require URNs or pseudo-URNs.

B.3. STAGE 3: OPTION 2 PLUS UNION CATALOGUES

This search is similar to the previous search (See B.2), but in this case one or more Union catalogues (e.g. OCLC, RLG, CURL) are also marked for searching, having been loaded with the other Z39.50 OPACs catalogued locally in response to the user choosing the "search other catalogues" option. Perhaps the user will choose to limit her search to EIOs only. Note, however, this is just an option, she may, equally, decide not to limit it.

EIOs catalogued on these OPACs will usually be held in many different places on the Internet (they will be held at the sites which have catalogued the EIO themselves). However, the EIO retrieval mechanism will be the same as at B.2 above. The user will choose a retrieval record which will contain at least one, but possibly several, URLs (or, perhaps, ultimately, a single URN which will 'point at' several URLs through a resolution service). If there is one, the client will load an appropriate NIDR client (such as Mosaic or Cello), pass it the URL or the URN, and the EIO will be retrieved and displayed/played. If there are several, a user choice will have to be made. Hopefully, this will be an informed choice based on system-provided information. Ultimately, it may be that the system will make the choice intelligently for the user on the basis of local information on network conditions - i.e. it will choose which URL to try first, which to try next and so on. The exact process utilised will depend upon the extent to which a practical URN to URL resolution service is ultimately developed.

Stage 3 Identified Development Requirements

Client / Server Requirements

(18) Ability to do a Z39.50 OPAC EIO/Internet search only (i.e. to search only for electronic resources instead of hard-copy resources as well).

(f) Ability to handle multiple URLs.

(g) Ability to make an intelligent choice between alternative URLs.

Other Requirements

(a) Union Internet catalogues . The existence of 'Union' Internet catalogues is a requirement here. OCLC are currently compiling a catalogue of Internet resources in their project 'Building a Catalog of Internet Resources'. This catalogue will be made publicly available. RLG are planning to add Internet resources to their Z39.50 Zephyr catalogue at the beginning of 1995 (Washburn, 1995). Similar records may be available on the CURL database in future. Given that such co-operative 'Union' catalogues will exist, it is reasonable to assume that they will be utilised as a more efficient way of searching the whole Internet than searching every single local OPAC.

(b) Updating URLs The method for updating URLs is not yet fully established. An automatic solution may only develop as these catalogues are established. There are already robot-type programs which can check for broken links and some automatic mechanism for updating the links may develop. In the CATRIONA model at present each site would have to send in an updated record to the Union catalogue when the local URL changes.

If a reliable URN to URL mapping service proves feasible this would cease to be a problem, although the problem of how the URLs are to be kept up to date in the mapping service would still have to be tackled.

URL Requirements

(See above - Updating URLs in Union Catalogues)

B.4. STAGE 4: SEARCH OF OTHER (MARC) Z39.50 OPACS SERVED UP BY UNION AND OTHER CATALOGUES

In this case the user conducts a distributed search as in B.2 and B.3 above but either gets no hits, or no useful hits, or insufficient hits for her purposes. At this point she might have the option of searching for records of additional Internet OPACs which might, in turn, satisfy her enquiry. It is assumed that her search would be subject-based and that the catalogue records describing other Internet OPACs would also have general subject classifications. The result of the user's search would be a new list of "other Internet catalogues" being offered by the CATRIONA client for her to mark for distributed searching. The search, retrieve and display/view process would then proceed as described in B.2 and B.3. In some cases the document or resource found in one or other of the catalogues might be charged for and a means of handling this would be built into the client-server dialogue.

The assumption behind the CATRIONA project is that the Internet will be catalogued in a distributed fashion. In order to carry out a search of this whole distributed catalogue, therefore, it is necessary for the OPAC client being utilised to be able to locate and search all of the individual OPACs that make up the distributed catalogue. In existing clients, this information is programmed into the client by the system administrator. However, the assumption in the CATRIONA model is that this information will initially be located on the local OPAC which will contain catalogue records for other OPACs selected by the library and uploaded to the client as required. It is further assumed that any given local OPAC will only contain a small subset of all of the available OPACs and that much of the information about the rest of the distributed catalogue of the Internet will come from other OPACs and, in particular, the Union catalogues. It is likely that Internet OPACs will be categorised according to their subject strengths and that users will be able to search and locate their catalogues for records of other OPACs strong on a particular subject category. The user will then be able to conduct a distributed search of a selection of these OPACs in a similar fashion to that described at B.1 to B.3 above.

Stage 4 Identified Development Requirements

Client / Server Requirements

- (15) Access control and charging mechanisms.
- (h) Ability to conduct a subject search for records of other Z39.50 OPACs

B.5. STAGE 5: SEARCH OF Z39.50 OPACS CONTAINING NON-MARC RECORDS (e.g. IAFA TEMPLATES AND TEI HEADERS)

At some point in her distributed parallel searching of other Internet catalogues, it is envisaged that the user might encounter Z39.50 OPACs which are not MARC-based. There may, for example, be library OPACs which do not use MARC or other non-library services using IAFA templates, TEI headers, Z39.50 SUTRS or GRS.

It is assumed that for some time to come, and perhaps indefinitely, valuable OPACs will exist on the Internet which do not utilise MARC and that these will also be catalogued in local and Union OPACs and potentially able to be part of any distributed search conducted by the user. It is unlikely that the user will know or care whether the records she retrieves are MARC records or IAFA templates or TEI headers or GRS or SUTRS or whatever and there may be nothing in the list of 'other services' the user is presented with to distinguish these catalogues from some of the others she will be presented with. They may be categorised by subject area but it is unlikely that they will be categorised as non-MARC. They will simply be one of a number of OPACs that the user may or may not choose to include in her distributed search. When records of Internet resources are retrieved from these OPACs the client will call up Mosaic (or whatever) and utilise the URL in a similar way to that described earlier, although this may require client development in order that the client should be able to identify a URL that is not in a MARC 856\$u subfield and client and server development to handle the record formats used. The record formats used (e.g. IAFA templates) would also have to be registered as accepted Z39.50 record formats.

Stage 5 Identified Development Requirements

Client / Server Requirements

- (19) Ability to handle URLs in SUTRS/GRS i.e. not in MARC 856 (client and server requirement).

Cataloguing Requirements

- (g) Ability to record Z39.50 information needed to set up client-server dialogues with non-MARC Z39.50 OPACs.
- (h) Records must be in a form that is compatible with developing Internet standards.

Other Requirements

- (c) In order to inter-work with Z39.50 servers containing IAFA templates (say), other projects need to be identified which utilise Z39.50 and IAFA or which can be persuaded to do so. IAFA templates would have to be registered as Z39.50 record format and clients and servers built to handle them.
-

B.6. STAGE 6: SEARCH OF NON- Z39.50 CATALOGUES

Some of the "other sources" presented to the user as options for a remote search of other Internet catalogues will be presented under a separate heading which will indicate that these are usually to be utilised after other options have been exhausted and also that they are non-Z39.50 and must be searched individually using whatever search language they employ.

If the user chooses to search one of these catalogues, the CATRIONA client will automatically load a Telnet program or Web or WAIS client depending on how the non-Z39.50 service has been described in the local OPAC (i.e. what is in the 856s subfield). The user will be logged on to the remote site and will conduct a search using whatever interface it provides (the ability to save, edit and resubmit search strategies may be of value here. In cases where the search client needed to search the remote catalogue is incapable of using a URL to retrieve a resource, it will be possible to copy retrieved records containing URLs to a file on the user's workstation and to convert them to HTML so that the user can then use Mosaic to retrieve and display the EIOs found.

The assumption here is that a large number of catalogues at present and for the foreseeable future will not only not be MARC, they will not be Z39.50 either. They will be offered to the user as important sources of Internet metadata but it will probably not be possible either to include them in any distributed search of the Internet or to utilise the same search strategies in them as were applicable to Z39.50 OPACs.

Stage 6 Identified Development Requirements

Client / Server Requirements

- (14) Save search strategy.
 - (17) Ability to conduct distributed (server by server) search for other non-Z39.50 Internet OPACs.
 - (20) Close down NIDR client session button in OPAC client.
 - (i) Ability to copy retrieved records containing URLs from a remote non-Z39.50 OPAC to a user file.
 - (j) Ability to convert document containing URLs to HTML in order to allow retrieval of EIOs by NIDR client.
-

B.7. STAGE 7: SEARCH INTERNET INDEXING SERVICES

Some of the kinds of 'other sources' dealt with in B.6 will not be non-Z39.50 OPACs but Internet indexing services such as Veronica or Archie or a WAIS site. Acting on the information found on the local catalogue record regarding the service concerned, the CATRIONA client would load the appropriate client for the service concerned, make the connection, and allow the user to conduct a search which would usually result in the eventual retrieval and display of an appropriate EIO. In the case of Archie there may be a facility to capture screens and HTML-ise the results so that the WWW client can start an actual FTP (File Transfer Protocol - see Glossary) session. The possibility of utilising Internet indexes such as Veronica and Archie is likely to continue to play a part in any Internet- wide search for the foreseeable future. These will be options on the menu of other services and, once again, an appropriate client will be called up to enable access. As with category 6, they will be accessed individually and will require a service-specific search strategy.

Stage 7 Identified Development Requirements

Client / Server Requirements

(k) Ability to load appropriate client to conduct search of Internet indexing services such as Archie or Veronica.

Cataloguing Requirements

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc. (A specific example of cataloguing requirement 1).

Other Requirements

(d) Stable national Veronica, Archie and WWW indexing services will be required possibly at duplicate sites (e.g. BUBL and NISS).

B.8. STAGE 8: SEARCH NETWORKED CD-ROMS OR OTHER DATABASES

The user may retrieve records locally describing locally held or remote CD-ROM titles accessible over the network from her workstation. Ideally the client would seamlessly connect the user to the preferred service and allow her to use it.

Stage 8 Identified Development Requirements

Client / Server Requirements

(25) Gateway to search ERL-compliant databases.

(26) Ability to load networked CD-ROM titles and to handle associated memory requirements.

Cataloguing Requirements

(j) Ability to catalogue CD-ROM titles and record access information for them. (a specific example of cataloguing requirement 1).

Note: ERL

SilverPlatter is using ERL (Electronic Reference Library) as the means SilverPlatter to provide Wide Area Network (WAN) access to large databases using TCP/IP. ERL is made up of two software components: the ERL clients (retrieval interface) and the ERL server (search engine). ERL runs over DXP (Data eXchange Protocol) which was designed by SilverPlatter to facilitate client/server interaction while keeping the functionality of stand-alone retrieval systems. Thus the DXP specification includes Boolean searching, thesauri and hot links as well as multi-database searching. Although several databases available on one server can be searched, at present the databases are not searched in parallel and the results are not sorted. This development is planned. In future DXP will also specify meta-thesauri for multiple database searching. Secure access to the databases is also provided as only authorised users can log in.

ERL is media independent so that it can accommodate databases in either CD-ROM or hard disk format. With ERL the databases can be copied to hard disk and a hybrid CD-ROM/hard disk solution can be created.

Z39.50

In ERL, the DXP co-exists with various standards including Z39.50 for libraries. Library clients written to Z39.50 can access all ERL compliant databases through a gateway. Ameritech is developing a gateway to allow their retrieval clients (WinPAC, ProPAC, TermPAC and MacPAC) to search ERL-compliant databases from the same Z39.50 client interface and SilverPlatter has programs to share technology with vendors who want to develop gateways from their clients to ERL compliant databases.

This would operate in a way similar to the other Z39.50 searches described above as the gateway maps the DXP

protocol to Z39.50.

Limiting and paying for remote access

Remote access can be arranged and controlled as ERL administrators control the security for multiple server access over the Internet. Servers are configured to indicate the machines and/or users who have permission to use one or more of the databases located on that server, through either user names and passwords or TCP/IP address ranges. Authorised users are presented with a single list of available databases, which may reside on a number of ERL servers' (SilverPlatter Information, 1994). Payment for remote access would be arranged with the individual information providers through SilverPlatter (Hazelden, 1994).

STAGES 9-10: SYSTEM-INFLUENCED ESCALATION

This section of the description of the model deals with the possibility of system-influenced escalation of searches beyond the local site as a result of failed local OPAC searches.

B.9. STAGE 9: FAILED KNOWN-ITEM SEARCH

This part of the illustration deals with a situation where the user conducts a known item search of the local OPAC. The search fails. In addition to presenting the user with 'near hits' from the local OPAC, the system also suggests that the user try a search of other Internet OPACs offering the user a choice of 'search none', search all', 'search selectively' (user choice). In the event of the user choosing 'search all', the system sends the search to each of the Z39.50 catalogues it 'knows' about. If it finds a hit it informs the user and offers to stop at that point. If all Z39.50 searches fail, it offers a user-driven option to search the other catalogues and indexes it 'knows' about (of course, if the user knows the URN, any URN to URL resolution service would be brought into play here).

If the user chooses 'search selectively' the system loads details of all of the catalogues catalogued in the local OPAC into 'other searches' and allows the user to mark those she wishes to search, then proceeds as above.

Stage 9 Identified Development Requirements

Client / Server Requirements

- (23) Ability to respond to a failed known item search by offering to search other Internet resources
 - (24) Ability to automatically search other OPACs in turn for known item and offer to stop searching when hit is found.
-

B.10. STAGE 10: SYSTEM RESPONSE TO SUBJECT SEARCH

This part of the illustration deals with a situation where the user performs a subject search in the local catalogue. Since it is possible either that a subject search will produce no hits or that it will produce hits which are either inappropriate or insufficient for the user's purposes, the system would always respond by offering to search other Internet sources. In the longer term, it might be possible that systems will respond intelligently to this situation, utilising the original search terms, together with a thesaurus, to automatically search for and identify appropriate subject catalogues on the Internet for the user to search (e.g. a search for 'quarks' might throw up Physics-strong OPACs as other sources). In the shorter term, a more likely approach would be to offer the user the chance to choose from a list of subject catalogues, those most appropriate to her search.

Stage 10 Identified Development Requirements

Client / Server Requirements

- (22) Ability to offer 'other subject catalogues' option after a subject search.
- (1) Ability of server to use a thesaurus to identify general subject terms appropriate to more specific terms entered by

user, and to use the general terms to identify appropriate subject catalogues to offer the user as a means of expanding the search.

STAGES 11-14: OTHER POSSIBILITIES

This section of the description of the model deals with the possibility of system-influenced escalation of searches beyond the local site as a result of failed local OPAC searches.

B.11. STAGE 11: NEW EDITIONS

Having found and scanned two electronic documents via the local OPAC, the user wants to find out if there are new editions of the document available elsewhere on the Internet. There would be an option to do a 'find others like this but more up to date' option. Having found a later edition of one of the works, the user asks the librarian whether he can find out if there is a later edition of the other. After some research, discussion with colleagues, and a search of recent BUBL updates bulletins, the librarian states that he is fairly confident that there isn't.

Stage 11 Identified Development Requirements

Client / Server Requirements

(12) Related works search (i.e. more like this title)

Other Requirements

(e) Librarians must have access to information services which enable them to keep up to date with new publications in the electronic world, just as they now have this for the hard-copy world. These would include publishers' catalogues, online and printed, and services like BUBL.

B.12. STAGE 12: KNOWN URN SEARCH

Having been given a hard copy article which makes reference to a number of EIOs by quoting, amongst other things, their URNs, the user utilises her OPAC client to conduct a distributed URN search of the Internet. Depending on how things have developed in the interim, the CATRIONA OPAC client would either search the other Internet OPACs it 'knows about' or would send the query to the URN to URL resolution service. In either case a number of alternative URLs would be presented to the user and the CATRIONA client would act on one or other of these by loading a NIDR client and retrieving and displaying the EIO.

Stage 12 Identified Development Requirements

Client / Server Requirements

(m) Ability to conduct a search for items utilising a URN.

Other Requirements

(f) A URN to URL resolution service perhaps utilising library catalogues as a stable source of URN to URL mappings.

B.13. STAGE 13: ELECTRONIC REFERENCES

Having retrieved and displayed a multimedia HTML document, the user decides to access some of the references at the end of the document. These include a URN for each of the EIO references in question. Depending on how things have developed in the interim, the user would either: a) 'click on' the URN and the Web client she is using would

access the URN to URL resolution service and attempt to retrieve and display the EIO utilising first one, then, if necessary, another of the associated URLs recorded in it. (or) b) 'copy' the URN back to the CATRIONA client and utilise it to conduct a distributed URN search of Internet OPACs.

Stage 13 Identified Development Requirements

Client / Server Requirements

(n) Ability to copy URN to CATRIONA client and utilise it to conduct a URN search.

B.14. STAGE 14: Z39.50 URL

Having accessed an EIO utilising a WWW client loaded by a CATRIONA client, the user finds a reference to a Z39.50 OPAC in the document and wants to access and search that OPAC. The web client 'recognises' the Z39.50 URL and automatically loads a CATRIONA client to enable searching of the OPAC to take place.

Stage 14 Identified Development Requirements

Client / Server Requirements

(21) Ability to be loaded by e.g. Mosaic (e.g. if Mosaic finds a Z39.50 URL).

Note:

This version of the model does not take account of the possibility of finding other subject-based OPACs by conducting a WHOIS++ directory service search as proposed in the ROADS initiative (ROADS, 1995). This will be taken into account in later versions of the model.

[Return to Contents Page](#)

Appendix C. Methodology and other Background Detail

C.1 Information Gathering

The survey and examination of hard-copy and electronic literature relating to various aspects of the subject area was carried out throughout the project:

Network navigation issues

Material identifying the problems of location and retrieval of electronic resources was gathered from various articles and reports, the most important of which are listed in the References section.

Investigation of developments in cataloguing standards and resource description

Much of what is happening in these areas can only be readily found by means of discussion lists and archived files. Information on MARC developments was monitored through the USMARC discussion list and archive. Other lists such as Autocat were monitored as they occasionally mentioned the cataloguing of Internet resources. The TEI list, as well as being a forum for the discussion of TEI development issues, has an archive of TEI documents. The Internet draft archive and archive of the IAFA discussion list provides information on IAFA template development, although the IAFA discussion list itself has been discontinued. (See Lists in References section)

Uniform Resource Names, Resource Locators, Resource Characteristics (URNs, URLs, URCs)

The URI list and Internet draft archives were the main source of material on the rapid developments in these Internet standards. Most of the discussion on the URI list focuses on specific technical areas of the various standards, while the Internet drafts provide more general information. (See Lists in References section)

Internet cataloguing projects

The aim was to find out about projects that might have a bearing on the CATRIONA project. Information was gathered from discussion lists, asking questions on the lists, keyword searches and searches of hard-copy literature, as well as enquiries received from those involved in related projects when information about the CATRIONA Project was posted up on the BUBL WWW server and on lists. The main related project is the second OCLC Project: 'Building a Catalog of Internet Resources' which began in October 1994 (OCLC, 1994), although there has not been any feedback from the project as yet. (See Appendix E for a full list of related projects identified).

Z39.50 protocol and implementation

Because this protocol is developing so quickly (Version 3 is currently under discussion), it was found that most of material that was available on the subject referred to the previous versions of the standard and was therefore out of date. Up to date information was obtained from the Z3950IW list, at meetings held by the Z39.50 Pre-Implementation Group, LAITG (Central Scotland) and CIGS (joint meeting), and the SCOLAR Technical Sub-Committee (See Appendix H) and by reading the proposed Version 3 standard itself. (ANSI/NISO, 1994). Another useful source was the pointer page of Z39.50 resources available at URL:

<http://ds.internic.net/z3950/z3950.html>

It has been recognised that there is a lack of easily accessible background information to the new version of the Z39.50 standard and Robert Waldstein from the Z3950IW discussion list is currently compiling a new FAQ (list of Frequently Asked Questions with answers).

Z39.50 Servers and Clients

News about developments in NIDR clients was gathered from BUBL, various discussion lists and journals.

C.2 Creation of a wide range of hypertext references on the BUBL WWW and Gopher Servers

A list of hypertext references to some of the electronic resource materials used in the project was created. Topics covered include cataloguing, navigation tools and issues, projects, URN/URL/URC (Uniform Resource Characteristic - see Glossary) documents and Z39.50.

Available from <http://www.bubl.bath.ac.uk/BUBL/catlinks.html>

C.3 Development of CATRIONA Model Versions 1, 2 and 3

The CATRIONA model developed from the initial proposal at URL:

<http://www.bubl.bath.ac.uk/BUBL/catriona.html>

to a model described in a series of 26 points at URL:

gopher://ukoln.bath.ac.uk:7070/0/Link/Catriona/Catriona_Model_Clarification

which was presented over the CATRIONA discussion list to systems suppliers in September 1994 for feedback. The model was then revised to produce Version 3 as more information was gained on the issues involved. Version 4 of the model (See Appendix B) was developed late in the project. Version 4 is an enhanced and extended version of Version 3. Version 3 was used to identify the initial development requirements utilised in the costing for Phase II (See also Section 6 of the main report).

C.4 Discussion of model over e-mail and at face-to-face meetings with systems suppliers and cataloguing services

Systems suppliers and cataloguing services sent in responses to the model (Geac, Dynix, FDG, SLS, MDIS, OCLC, RLG) and these formed the basis for further meetings and email discussions with systems suppliers and cataloguing services. Meetings to discuss CATRIONA were held with Geac (October 1994), FDG and CURL (November 1994), Dynix and MDIS (November 1994). Discussions with other suppliers including SLS took place at meetings of the Z39.50 Pre-Implementation Group.

C.5 Dissemination of information relating to the project through BUBL and e-mail discussion list

Information about the CATRIONA Project was disseminated via the BUBL WWW and gopher servers. The main documents relating to the project: the initial proposal, the CATRIONA model, and the proposal for a second phase, were posted.

(See: gopher://ukoln.bath.ac.uk:7070/0/Link/Catriona/Catriona_Phase II)

The model document was also sent to several LIS email discussion lists including (e.g. PACS-L, LIS-LINK, LIS-BAILER, Z3950PIG) for feedback.

C.6 Presentations of the model and cataloguing issues

An opportunity was provided to present the CATRIONA model and to discuss cataloguing issues with cataloguers and systems staff at a meeting of the Cataloguing and Indexing Group Scotland (November 1994) which focused on the accessing and cataloguing of Internet resources and the CATRIONA model. Aspects of the model relating to Z39.50 were discussed at the Z39.50 discussion meeting held by the SCOLAR Technical Sub-Committee and Ameritech (See Appendix H).

C.7 Z39.50 Issues

(See Appendix H)

C.8 Survey, installation and examination of a number of existing Z39.50 OPAC clients

Several systems suppliers provided Z39.50 OPAC clients for test evaluation. Geac provided GeoPac Release 1.23, Dynix provided WinPAC Beta 1.0, and Horizon 3.2, and SIRSI provided VIZION. Access to the SLS Libertas 6.3 client and the Innovative Interfaces Innopac Z39.50 client was carried out over the Internet using Telnet. Demonstrations of the Fretwell- Downing client and MDIS prototype client were provided during visits to the companies. WinSpirs (the SilverPlatter ERL client for access to its ERL servers was also provided for examination). Data Research Associates also provided their Z39.50 client but because of time constraints it could not be included in the evaluation.

C.9 Survey, installation and examination of a number of NIDR clients

A range of NIDR clients were installed and evaluated including Cello, MS Windows Mosaic, Netscape, Hgopher and WinWais.

C.10 Testing the Feasibility of the Model

(See Appendix D)

C.11 Design of a Proposed Demonstrator Project

A demonstrator project based on the CATRIONA model was designed with a view to submitting an 'expression of interest' under the FIGIT initiative. An updated version of this, amended to take account of later feasibility study developments, is outlined in Section 8 of the main report. Alternative designs for a demonstrator project were not considered, although it is the case that a range of options exist and that a demonstrator project could, within certain limitations, be tailored to fit a desired level of funding.

C.12 Formulation of development requirements for proposed demonstrator project

A list of development requirements for the proposed demonstrator model was drawn up and amended during

discussions with suppliers. This consisted of a list of 26 client- related requirements and 6 server-related requirements (See Section 6) drawn from an analysis of Version 3 of the model. The list was then distributed to the systems suppliers during December with a request for feedback in the first half of January. The systems suppliers were asked to provide information on which of the client requirements were already being developed, which were planned and which were not yet planned, with approximate time and cost estimates for these categories. (See Appendix D for results).

C.13 Discussion of Likely Costs of Phase II with Selected Suppliers

An attempt was made to gauge the likely costs of Phase II. Because of time limitations, only selected suppliers were utilised in this exercise. Results are summarised in Appendix D and in Section 9 of the main report. However, they were difficult to specify exactly at this stage.

C.14 Preparation of report

The report was prepared with a view not only to presenting the results of the project but also with the intention of creating a summary of information and contacts gathered to date.

C.15 Limitations of the Feasibility Study

Some areas which required investigation could not be addressed in any detail because of time constraints. These included the following:

- URNs, URLs, URCs
- Future Z39.50 developments
- Links with /conversion to and from IAFA templates and TEI headers
- Access control issues in Z39.50
- Improvements to existing tools like Veronica and Archie
- Web client Z39.50 gateways and autoload of appropriate OPAC client

[Return to Contents Page](#)

Appendix D. Feasibility Tests: Background Information, Client Capabilities and Requirements

This appendix focuses on the feasibility of setting up the CATRIONA model at a basic level, describes how this was done, examines the current capabilities of the various OPAC clients, and briefly describes each client.

Note: This appendix should be read as an addendum to Section 5 of the report.

D.1 Feasibility Tests: Background Information

GeoPac provided the main test of the feasibility of setting up the CATRIONA model at a basic level. It proved to be possible to set it up so as to be able to access three remote Z39.50 OPACs (Brigham Young University, Butler University and North Carolina State University).

Connection details provided:
Brigham Young University
Host=adm4.byu.edu
Port=20003
Database Name=OPAC
AttributeSet=1.2.840.10003.3.1

Butler University
Host=ruth.butler.edu
Port=210
sDBNames=Marion

North Carolina State University
Host=library.ncsu.edu
Port=210
Database Name=MARION

The connection information above was used to add the two databases to the existing range of databases available in GeoPac as originally installed. Brigham Young University also provided details of the attribute set (e.g. 1 = Personal Name, 2 = Corporate Name etc.) supported by the server and this was set up in the client so that the client search was conducted under the appropriate parameters. Details of the records were also provided. These records had been enhanced to include URLs in 856\$u (as recommended by MARBI in Proposal 93-4). On retrieval of the records the client's 'link' button was highlighted and clicking on this had the effect of making the OPAC client load a suitable 'viewer', and pass it the URL so that it could locate, retrieve and display the electronic information object (EIO) described by the MARC record.

D.2 Feasibility Tests: Background Information

The first database searched was the BYU server at Brigham Young University, in which there were two records with URLs in the 856\$u subfield.

Brigham Young University Library Server (BYU)

BYU Record 1

The viewer loaded was the WWW client Mosaic and the EIO was a full text article on a gopher located at Minneapolis:

The Internet Gopher Protocol; a distributed document search and retrieval protocol: Alberti, Bob

Marc record type USmarc
Logical record length 590
Record Status n
Type of record a
Bibliographic level m
Base address of data 109
Encoding level b
Descriptive cataloging form a
Linked record requirement b

```
008 940628s1992bbbbbmnubbbbbbrrbbbb00000bengb
035 bb$a1234567
100 1b$aAlberti, Bob
245 14$aThe Internet Gopher Protocol;$ba distributed document search and retrieval protocol $c/ Bob Alberti,
Farhad Anklesaria, Paul Lindner, et al.
260 bb$aMinneapolis;$bUniversity of Minnesota Microcomputer and Workstation Networks Center
650 b0$aInternet (Computer network)
856 0b$agopher.med.umich.edu;$fInternet Gopher Protocol $ugopher://gopher.med.umich.edu:70/00/Information%
20About%20Gopher/Internet%20Gopher%20Protocol
```

BYU Record 2

The viewer loaded was a .gif image viewer and the EIO was a .gif file located at Brigham Young University:

Photograph of Joseph F. Smith: Johnson, Charles Ellis

Marc record type USmarc
Logical record length 298
Record Status n
Type of record g
Bibliographic level m
Base address of data 97
Encoding level b
Descriptive cataloging form ?
Linked record requirement b
007 kh?buz 008 941102b????????unk????????????bbb???eng?d 035 bb#a1234568 100 1b#aJohnson, Charles Ellis.
245 10#aPhotograph of Joseph F. Smith 856
0b#aLibrary.byu.edu#fPhotograph#uhttp://library.byuu/multimedia/jfsmith.gif

Butler University Library Server

Butler University Library has at least 3 records with URLs in the 856\$u subfield.
The first record was a full-text HTML format file located at Indianapolis.

Butler Record 1

Mainstreaming our online catalogs: Kambitsch, Timothy G.

Marc record type USmarc
Logical record length 1131
Record Status c
Type of record m
Bibliographic level m
Base address of data 181
Encoding level I
Descriptive cataloging form a
Linked record requirement b

001 ABR-9659
008 941007s1994bbbbbinnubbbbfbbzbbsbbbbbengbd
099 bb#ahttp://www.butler.edu/www/library/papers/lotf.html
100 1b#aKambitsch, Timothy G.
245 10#aMainstreaming our online catalogs#h[computer
file] /# by Tim Kambitsch
(kambitsch@butler.edu)
260 bb#aIndianapolis,#c1994
300 bb#a1 computer file in HTML format; anchors to 5 GIF
illustrations
520 bb#aThis paper outlines a model of providing
intellectual access to internet resources through the use of the
World Wide Web. The author advocates that library, system vendors and
database providers make databases available from WWW saavy servers or
through gateways. Librarians can participate in providing
access to specific electronic resources via web-aware online catalogs.
650 bb#aLibraries#xAutomation
650 bb#aInformation retrieval
650 bb#aLibraries and electronic publishing
856 7b#uhttp://www.butler.edu/www/library/papers/lotf.html
#awww.butler.edu#2World We idWeb
857 7b#uURL:http://www.butler.edu/~kambitsch/kambitsch.
html#2World Wide Web

North Carolina State University Library Server (NCSU)

NCSU Record 1

**ALA Washington Office newsline ALAWON : an electronic publication of the American Library Association
Washington Office.**

Marc record type USmarc
Logical record length 1540
Record Status c
Type of record a
Bibliographic level s
Base address of data 373
Encoding level b
Descriptive cataloging form a
Linked record requirement b

001 AFW-1740
003 OCoLC
005 19940616115148.0
008 920721c19929999dcubx1bbbbbb0bbba0engbd
010 bbçasn93004037bç26226155
040 bbçaVPIçcVPIçdNSD
012 bbçla
022 0bça1069-7799
042 bbçansdpçalcd
082 10ça025ç212
090 bbçaZ673.A5çbA42
049 bbçaNRCC
210 0bçaALA Wash. Office newslne
212 1bçaAmerican Library Association Washington Office newslne
222 b0çaALA Washington Office newslne
245 00çaALA Washington Office newslneç[computer file]:çbALAWON :
an electronic publication of the American Library Association Washington
Office.
246 10çaALAWON
260 bbçaWashington, DC :çbThe Office,çc[1992-
265 bbçaAmerican Library Association Washington Office,
110 Maryland Ave., .NE, Washington, DC 20002-5675
310 bbçaIrregular
362 0bçaVol. 1, no. 1 (July 9, 1992)-
500 bbçaMode of access: Electronic mail on BITNET;listserv@uicvm;
SUBSCRIBE ALA-WO First Name Last Name
500 bbçaTitle from title screen.
650 b0çaLibrariesçzUnited StatesçxPeriodicals.
650 b0çaInformation servicesçzUnited StatesçxPeriodicals.
610 20çaAmerican Library Association.çbWashingtonOfficeçxPeriodicals.
710 20çaAmerican Library Association.çbWashington
Office
936 bbçaVol. 2, no. 18 (May 10, 1993) LIC
856 bbça
<http://dewey.lib.ncsu.edu/stacks/alawon-index.html>

D.3 Current Client Capabilities

The functionality of the 6 Z39.50 clients that were made available or demonstrated was compared. Systems developers were also asked to indicate any features which were under development or were planned. The MDIS 'Lion' client was observed but is not included as it is under development. A request was also made for the VTLS client and the TRW SmartSearch client but access to these was not provided. The focus of the assessment was in relation to the features that would be required in a CATRIONA type client and so the list of features follows the sequence of stages in the user's search described in the model (Appendix B).

Main capabilities required

1. PC client (plus, ideally, MAC and X-Windows and VTxxx)
2. Data for accessing other Z39.50 catalogues not on menu can be uploaded from catalogue records
3. Distributed parallel searching
4. Group of databases can be selected by user for parallel searching
5. Number of hits given before records are received
6. Numbers of hits from several servers are given separately
7. Warning if result sets are too large
8. Stop button (does this need to be individual to a server session?)

9. Option to merge results from several servers
10. Option to sort results
11. Filing duplicates together (URNs or pseudo URNs)
12. Related works search (i.e. more like this title)
13. Ability to call up one of several NIDR clients as appropriate (e.g. WWW, Gopher, WAIS) (use .ini file as in Mosaic)
14. Save search strategy
15. Access control and charging mechanisms
16. Servers to appear in catalogue
17. Ability to conduct distributed (server by server) search for other non-Z39.50 Internet OPACs
18. Ability to do a Z39.50 OPAC EIO/Internet search only (i.e. to search only for electronic resources instead of hard-copy resources as well)
19. Ability to handle URLs in SUTRS/GRS (Simple Unstructured Record Syntax/Generic Record Syntax) (i.e. not in MARC 856)
20. Close down client session button
21. Ability to be loaded by e.g. Mosaic (e.g. if Mosaic finds a Z39.50 URL)
22. Ability to offer 'other subject catalogues' option after a subject search
23. Ability to respond to failed known-item search by offering to search other Internet resources
24. Ability to automatically search other OPACs in turn for known item and offer to stop searching when hit found
25. Gateway to search ERL-compliant databases
26. Ability to load networked CD ROM titles and to handle associated memory requirements

Notes on features and table entries

As can be seen from the table below, each of the features required for the model to function at a basic level are available in at least one of the clients. In some cases an alternative feature was provided:

5. Number of hits given before records are received

If the number of hits is very small, e.g. 1-4, in some cases (e.g. Libertas 6.3) the full records are retrieved.

7. Warning if result sets are too large

Rather than providing a warning if the result set is too large, most of the developers used the alternative method of setting up the size of result set required in advance.

14. Save search strategy

This only applies to one search and does not mean "save search strategy from search a and then combine with search b".

16. Servers to appear in catalogue

The currently available method of offering a choice of servers is by providing a list for the user. This is simple to use, but does not give any indication of which server to select for the kind of search, although VIZION gave a basic searchable subject listing for the various servers. One possibility is to create searchable catalogue records for servers with a subject description. Another possibility is to use some form of directory service.

20. Close down client session button

This refers to the NIDR client which the user may wish to close down, and not to the Z39.50 client which it is assumed can be closed down. Some suppliers required clarification on this feature.

Table of Z39.50 Client Capabilities

Features	WinPAC Beta 1.0	Horizon 3.2	GeoPac Release 1.23	VIZION	Oracle Libraries Client	Libertas 6.3
1. PC client	Yes	Yes	Yes	Yes	Yes	No. Definitely planned.
2. Data for accessing other catalogues not on menu can be uploaded from	No. Under review.	No. Facility for adding other catalogues under development.	No. But other catalogues can be added if user has data.	No. But other catalogues can be added if user has	No	No

catalogue				data.		
3. Distributed (parallel) searching	No	No	No. Under development.	No. Planned.	Yes	Yes. (Serial searching).
4. End-user can select group of servers	No. (Single server).	No. Under development.	No. (Single server).	No. (Single server). Planned.	Yes	Yes
5. Number of hits given before records received	Yes	Yes	Yes	Yes	Yes	Yes
6. Number of hits from servers given separately	No	No	No	No	Yes	Yes
7. Warning if result set is too large	Yes	Yes	No. Size can be set in advance.	No	No	No. Size can be set in advance.
8. Stop button for each server	No. Can't stop search.	No. One stop button.	No. One stop button.	No. One stop button. Planned.	Yes. And overall stop button.	No
9. Option to merge results from several servers	No	No. Possible for user to do locally and manually.	No	No. Planned.	Yes	No
10. Sorting of results (Z39.50 Version 3)	No. Version 3 server function. Client can do with Version 2.	Yes. (Single server).	Yes. (Single server).	No. Planned. Though best done on servers.	No. Planned.	No. Planned.
11. Filing duplicates together	No	No	No	No. Planned.	No. Possible.	No
12. Related works search	Yes. (One server).	Yes. (One server).	Yes. (One server).	No	No. Possible.	No
13. Ability to call up NIDR clients (via URL in catalogue record)	No. Called up via WinGOPHER	Yes. (Not in MARC 856\$u).	Yes. (In MARC 856\$u).	No. NIDR clients in menu.	No. Near future.	No. Planned.
14. Save search strategy/ results	Yes/Yes. (Single server).	Yes/Yes. (Single server).	Yes/Yes. (Single server).	Yes/Yes. (Single server).	No/No. Near future.	No. Can search same group of servers./No.
15. Levels of access/ charging	No. Planned.	No. Planned.	No. Planned.	No	No. Near future.	Yes
16. Servers to appear in catalogue	No	No. Menu provided.	No. Menu provided.	No. Menu provided.	No. Will be in DALI.	No. Menu provided.
17. Ability to search for a non Z39.50 Internet OPAC	WinPAC has Telnet option.	Can do using SQL.	Yes. Telnet access via a defined URL.	Yes. (Not via URL in catalogue record).	No. Planned.	No.
18. Ability to search for only electronic resources	No	Could if server supports it.	No	No	No	No
19. Ability to handle URLs in					No. Near	

other Z39.50 record syntaxes e.g. GRS	No. Planned.	No	No	No	future.	No
20. Close down NIDR client session button	No	No	Yes. But error message given.	No	No	No. Planned.
21. Ability to be loaded by e.g. Mosaic	No	No	Further investigation required.	No	No	No
22. Ability to offer "other subject catalogues" option after a subject search	No. (Subject list for each OPAC).	No. (Subject and title keyword list).	No	No. (FILTER provides subject search of OPACs)	No	No
23. Offer to search other Internet resources if known item search fails	No	Others can be searched and system would prompt user.	No	No	No	Quick Search option does this.
24. Ability to automatically search other OPACs in turn for known item and stop searching when found	Under review.	No	No	No	No	No
25. Gateway to search ERL-compliant databases	No	Under development.	No	No	No	No

D.4 Additional Notes on Clients

WinPAC Beta 1.0

WinPAC has an attractive interface although there are a few features that limit the value of the WinPAC version evaluated. One of these is that in order to select a server the user has to exit from WinPAC and go into WinGOPHER.

HORIZON 3.2

A demonstrator version of Horizon 3.2 was provided for testing. This client provided most of the key features and several additional features. A number of developments are planned or already underway, such as remote distributed searching of more than one Z39.50 site, and the facility to add servers for searching that are not on the menu originally provided. The resource can be accessed and downloaded as well as the catalogue record for the resource, but this is not yet available according to established standards (i.e. the URL is not in MARC 856\$u).

GeoPac Release 1.23

GeoPac provided most of the key features required as well as a user-friendly interface and was successfully used to access and download electronic resources in remote locations using a standard method of resource description. Additional servers not on the menu can be added and searched if the user knows the connection details. Distributed searching is currently under review.

VIZION

A test version of VIZION was provided for evaluation. The interface is fairly user-friendly although the user has to work out which choice to make first rather than being guided. VIZION offers a searchable list of destinations including Z39.50 sites, Gopher sites and WAIS sites. The user chooses and the client uses the appropriate protocol to communicate with the server. Servers can be added if the user knows the connection details.

ORACLE LIBRARIES CLIENT

The Oracle Libraries client is capable of distributed searching and this was demonstrated. The Oracle Libraries client has already been up and running for some time (See Appendix E - Projects for information on the IRIS project). The facility for distributed searching was the main advantage of the Oracle Libraries client, although the distributed searching was not carried out by the client itself but by a Unix 'collator' which sent off the requests to the 5 Irish servers, merged the results and sent them back to the client. The resources themselves could not be accessed or downloaded but only the records.

Libertas 6.3

This client differed considerably from the other clients in that it is not a PC client, although SLS plans to bring out a PC client in 1995. The other main difference was that this client is aimed more at library staff rather than the end user and offers a wide variety of facilities for downloading records in various formats and modifying records. It is understood that the PC client will be aimed more at the end-user. Again the main advantage of the SLS client was the facility to do distributed searching. The servers were searched serially which made the response time longer, especially for the US servers. Again the resources themselves could not be downloaded - only the records.

[Return to Contents Page](#)

Appendix E. Related Projects and Developments

E.1 Template/Record based Projects

Alex: A Catalogue of Electronic Texts on the Internet

The Alex catalogue is a central Internet OPAC for the location and retrieval of the full text of documents on the Internet. It currently indexes over 700 books and shorter texts by author and title, including texts from Project Gutenberg etc. At the moment it does not have serials. Access details: **<gopher://rsl.ox.ac.uk:70/11/lib-corn/hunter>**
Alex has recently been converted to MARC format which is available for downloading. Access details: **<ftp://ftp.lib.ncsu.edu/pub/stacks/alex/alex-950224-marc.txt>**

ALIWEB - Archie-Like Indexing in the Web

The ALIWEB framework is a method of providing distributed indexing in the Web and thus reducing the load on network resources. Server administrators complete index files in the form of IAFA templates (See Appendix F), which are regularly retrieved and compiled into a database. The database can be searched using a simple WWW search interface or via other databases. Access details: **<http://web.nexor.co.uk/public/aliweb/aliweb.html>**

CNI TopNode

CNI TopNode is an Telnet-accessible database of directory level and individual resources in the Internet and is produced by Merit Networking and Indiana University. The topnode database has been developed using BRS/Search software on CNI's server. Individuals and organisations offering a resource are required to complete registration templates developed for different types of resources. Access details: **<gopher://gopher.cni.org:70/11/cniftp/projects/topnode>**

Harvest

Harvest is a system for efficiently gathering and distributing indexing information. It supports the construction of various types of indexes which can be customised to suit the each information collection; and it provides caching and replication to prevent network congestion. Harvest is made up of several subsystems including the Information Gatherer and Information Broker - which can be searched using author, keyword, title or URL - as well as subsystems for indexing/searching, replicating and caching. The indexing format used by Harvest Summary Object Interchange Format (SOIF) is similar in some respects to IAFA templates (Bowman et al, 1994). Access details: **<http://rd.cs.colorado.edu/harvest/>**

INFOSERVICES

INFOSERVICES is a collaborative project between SURFnet, the Dutch Academic Network and the National Library of the Netherlands to improve access to networked information resources. Suppliers of information are required to provide elementary descriptions in an accompanying INDEX file which is based on the IAFA template

standard (See ALIWEB and Appendix F). The IAFA template material is then enhanced to form catalogue records for electronic resources. Cataloguing enhancements recommended by MARBI, such as the addition of USMARC field 856 were adopted by the project and implemented in the Dutch Pica cataloguing environment which allows materials to be accessed through the Dutch OPACs as well as through the combined central Union catalogue and ILL system (van der Werf, 1994). In association with OCLC, PICA have recently set up an experimental Z39.50 interface link to OCLC's FirstSearch service. This is currently undergoing evaluation in six Dutch libraries.

NISS Information Gateway

The NISS Information Gateway brings together all of the existing NISS services and employs an enhanced, easy to use WWW interface. Resources are classified according to a subject tree approach and described by means of templates. The templates contain object information, such as information name, information type, UDC number and free text description; and implementation information, such as URL, service type, network address and currency. Access details: <http://www.niss.ac.uk/>

OCLC: Building a Catalog of Internet-Accessible Materials

In the new OCLC project - Building a Catalog of Internet- Accessible Materials, libraries and representatives from their host institutions will identify, select and catalogue EIOs accessible through the Internet. The records produced in the project will be made accessible over the Internet. The main aim of the project is to test and evaluate the effectiveness of using the USMARC format for bibliographic records, and field 856 in particular, as a means of providing description, location and access information for EIOS. The project duration is October 1, 1994 to March 31, 1996. Access details: <http://www.oclc.org/oclc/man/catproj/catcall.htm>

OMNI - Organizing Medical Networked Information

OMNI is a project to build a gateway for access to quality health and biomedical information resources to the UK higher education and research community. The gateway will be a WWW interface to a catalogue of biomedical and health- related information. Resources will be filtered, classified, subject indexed and catalogued using a modification of the NISS/SOSIG template, which underlies the search mechanism. Access details: <http://www.nimr.mrc.ac.uk/OMNI/>

SOSIG - Social Science Information Gateway

SOSIG is an ESRC-funded project to provide a centralised means of access to social science resources available over the Internet. All of the resources on the SOSIG server are catalogued using a standard template (See NISS Information Gateway). Access details: <http://sosig.esrc.bris.ac.uk>

E.2 Other IR Projects

ICE

The ICE indexing server extension enables sophisticated free text searches on a Web archive. The indexer provides relevance feedback, use of a thesaurus to make conceptual searches for e.g. image, forms searching and boolean searching. As well as being used on individual archives the ICE software can be used as a data gathering script to provide and index file forarchie like indexers such as ALIWEB. (Neuss and Hofling, 1994) Access details: <http://www.dsi.unimi.it/cgi-bin>

LYCOS

LYCOS currently provides one of the largest indexes to the Web. The huge index is built by a Web crawler using a system that combines random choices with certain preferences. Information gathered includes title, headings and subheadings, weighted list of 100 terms, first 20 lines, size and number of works. The index can be searched using document title, headings, links and keywords and an extract can be examined which allows the user to evaluate the document before retrieval (December, 1994). Access details: <http://lycos.cs.cmu.edu>

Subject Trees

Subject Trees and subject directories are used for organising subject-related resources in a hierarchical manner. Examples of these include the following: the BUBL Subject Tree (See Appendix I. Access details: <http://www.bubl.bath.ac.uk/BUBL/Tree.html> or <http://www.bubl.bath.ac.uk/BUBL/Treealphabet.html>); EINet Galaxy (Access details: <http://www.einet.net>); the WWW virtual library (Access details: <http://info.cern.ch/hypertext/DataSources/bySubject/Overview.html>); GENBB (Access details: http://www.cs.colorado.edu/homes/mcbryan/public_html/bb/su mmary.html); YAHOO (Access details: <http://akebono.stanford.edu/yahoo/>). Users can contribute to these systems by supplying

the appropriate information about their resources.

WebAnts

Foreseeing that single indexes of the Web can only grow at the expense of coverage, the aim of the WebAnts project is to develop a cooperative Web explorer which shares results with other explorers and avoids duplication of effort. The distributed approach which involves smaller cooperating search engines means that sites do not have to have large resources available, and so more sites can be involved. It is also envisaged that a cooperative method can be used to provide index information to the end-user. Access details: <http://thule.mt.cs.cmu.edu:8001/jrrl-space/webants.html>

WWW (World-Wide Web Worm)

The aim of WWW, released in March 1994, is to locate all of the WWW addressable resources on the Internet and provide a powerful user interface to the resources. The WWW program scours the Internet locating all web resources. The search engine is accessed on the WWW home page and uses forms and searches, including keyword searches, can be carried out on titles, reference hypertext or within the URL name strings of documents (McBryan, 1994). Access details: <http://www.cs.colorado.edu/home/mcbryan/WWW.html>

E.3. European Z39.50/SR Library-based Projects

British Library/ONE

The British Library are developing a target Z39.50 server based on the NLC code which is due to be available in August 1995. A range applications of Z39.50 are also being investigated under the Initiatives for Access projects. The British Library is one of the 9 partners in ONE, an EU project which includes partners from Finland, Norway, Sweden and Germany. One of its key aims are to produce a public domain origin and target kernel code running over Internet and OSI telecommunication protocol stacks.

DALI (Document and Library Integration)

The DALI (Document and Library Integration) project led by Fretwell-Downing will build on the IRIS project to develop, pilot and test a multimedia document delivery service in a distributed environment over Z39.50. One of the key aims is to develop a modular software package which will incorporate SR and Z39.50, X.400 and MIME. Other features include a current awareness service, directory service (X.500) and charging mechanisms.

DECOMATE (Delivery of Copyright Material to End-Users)

The aim of this project is to develop software which will provide links between bibliographic records and full text articles, allowing the user to view the article or have it delivered.

DVB-OSI II

DVB-OSI II aims to achieve transparent unified access to a range of resources most of which are Union catalogues, including the Pica Union catalogue, but also including abstracting and indexing services.

EUROPAGATE

EUROPAGATE is a gateway project between SR and Z39.50 with Danish, Irish and Portuguese partners.

ION

The aim of the ION project, completed at the end of 1994, was to develop interconnection between three computerised ILL networks: LASER in the UK, PICA in the Netherlands and PEB in France - in order to support and develop international interlending and messaging services; and to demonstrate the capabilities of the ILL and SR communication protocols.

IRIS (Irish Networked Database Service)

IRIS is a document delivery service providing access via a Z39.50 client developed by Fretwell-Downing to six Irish library servers developed by Dynix, MDIS, BLCMP and Fretwell-Downing. The user can select a group of library servers to for distributed parallel searching. The Fretwell-Downing Z39.50 client (Oracle Libraries client) is evaluated above in Appendix D.

NORDIC - SR-Net

The aim of the early NORDIC - SR-Net project was to establish a network based on the SR protocol between library systems in the 5 Nordic countries.

SOCKER

The objective of SOCKER is to demonstrate the viability of the ISO SR protocol within various environments: CD-ROM workstation, library system and network access point.

Sources: Lorcan Dempsey (1994) and see also Appendix H; Fretwell Downing; (VINE, 1994) and information available at the above URLs.

[Return to Contents Page](#)

Appendix F. Cataloguing and Resource Description Developments

F.1 Introduction

If it is to be possible to find and retrieve EIOs, whether they are documents, images or online catalogues, they must be adequately described and their electronic location must also be specified in such a way as to facilitate the retrieval of the EIO. In other words, the gathering and encoding of meta-information (information that is not found in the resource itself (e.g. source, revisions) is required.

This is a problem that has not been addressed by those involved in resource description until relatively recently. However, a number of initiatives are now underway to accommodate the cataloguing of EIOs, conducted both by those traditionally concerned with cataloguing standards, MARBI in particular, and by those newer to the field such as the URI (Uniform Resource Identifier - See Glossary) and TEI groups.

In this section we consider some of these initiatives from the perspective of what is required for the CATRIONA model to operate successfully, looking both at what has been achieved and what remains to be done.

It should be stressed at the outset that almost all of the enhancements to resource description required for the CATRIONA model to function at a basic level are already in place in one of the cataloguing schemes (USMARC). The other three schemes seem to be moving in this direction with IAFA templates coming closer to fulfilling the CATRIONA requirements than URCs or TEI headers, although all three schemes are presently inadequate for CATRIONA since, at present, they are not able to be handled by Z39.50 clients and servers.

The basic level requirements of a cataloguing scheme needed for the CATRIONA model to function are listed below as (a) (- to a certain level), (d) and (f). The remaining requirements are not necessary for the demonstrator model to function, but would be necessary for a fully functional, integrated implementation of the CATRIONA proposals.

F.2 Cataloguing Requirements for the CATRIONA Model

The cataloguing requirements for the CATRIONA model discussed below correspond to cataloguing requirements (a) to (j) given at the end of each stage of the model description in Appendix B. They were identified through an analysis of Version 4 of the model.

BASIC REQUIREMENTS

(a) Ability to catalogue any and every possible type of Internet resource and service

Resource descriptions are required that cover all of the types of electronic resource that users may want to retrieve. The record must contain electronic location and access information. For the CATRIONA model to function at a basic level, only minimal details are needed (e.g. title, author, URL).

(d) Records must be of a type that can be searched and retrieved using Z39.50.

This is a basic requirement for the CATRIONA model. Several record formats can currently be transported using Z39.50 including USMARC, the record format central to the CATRIONA proposals.

(f) Ability to store several alternative URLs in catalogue record.

A record should be able to supply more than one URL for a resource if more than one location exists.

ADDITIONAL REQUIREMENTS FOR FURTHER MODEL DEVELOPMENT

(b) Ability to record username and password of commercial services in suppressible (MARC) field.

This is required in order to enable and control access to commercial services.

(c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

This is required in order to enable local librarians to manage users' access to EIOs.

(e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.

Records should be able to describe Z39.50 OPACs, including subject description and access information such as the address of the host, number of the port, name of the database, preferred record syntax and type of attributes supported.

(g) Ability to record Z39.50 information needed to set up dialogues with non-MARC Z39.50 OPACs.

MARC records should be able to describe non-MARC Z39.50 OPACs and provide access information

(h) Records must be in a form that is compatible with developing Internet standards

Standards adopted in the CATRIONA model should be compatible with the emerging standards of Internet resource description, such as those being developed by the IETF (Internet Engineering Task Force) such as URCs.

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.

(Specific example of requirement (a))

(j) Ability to catalogue CD-ROM titles and record access information for them.

(Specific example of requirement (a))

F.3 Current Implementations of Cataloguing Requirements

In this section, the way in which different cataloguing schemes can accommodate the requirements of the CATRIONA model is examined.

In each case the information on the requirements (a) to (j) was sought in the key documents for each resource description scheme. In many cases it was not possible within the timescale of the project to determine whether or not the requirement was met. This does not mean that the requirement cannot be met, but only that it has not as yet been possible to obtain clear information on the question.

F.3.1 MARC Records

In 1992 OCLC began their first project in cataloguing Internet resources. (Dillon et al, 1993). This involved 30 cataloguers cataloguing 300 data resources, mainly documents. The results of this project fed into Proposal 93-4 (MARBI, 1993a) for cataloguing Internet resources. The most significant change resulting from Proposal 93-4 was the addition of a new field (856) "Electronic Location and Access". The field was accepted in January 1993 and is now formally part of the USMARC format. (See Note 1)

Together with other additions, the above enhancement to the USMARC record format means that it meets the CATRIONA requirements sufficiently for the model to function at a basic level. There are several advantages to using the MARC format for describing EIOs at present and for the near future: it is already widely used and so cataloguers are not required to learn a new system; it provides a good standard descriptive format, provides subject indexing, and it also works well with Z39.50.

There is the problem of differing national MARC formats to be considered as these enhancements were developed initially for USMARC and implementation in other MARC formats is very limited as yet.

BASIC REQUIREMENTS

(Note: the following description applies to the USMARC format up to Proposal 95-1)

(a) Ability to catalogue any and every possible type of Internet resource and service.

Yes - requirement met at a basic level. A wide range of Internet resources can potentially be described using MARC and this is to be tested in the current OCLC Project - Building a Catalog of Internet Resources. (OCLC, 1994).

Resources identified by MARBI as requiring to be catalogued:

OPACs, Bulletin Boards, Mailing List Servers, Data Archives, Computational Resources, White Pages, Network Information Centres, Full-text Databases, Numeric Databases (MARBI, 1991a), other types of citation database (MARBI, 1991b), Computer Discussion Lists and Forums, FTP Sites (since these are rapidly changing and numerous, it was questioned whether they should be controlled bibliographically), General Online Services (e.g. Prodigy, Compuserve), Campus Wide Information Systems, Distributed File Servers (Gopher, WAIS, World Wide Web) (MARBI, 1993b)

(d) Records must be in a form that can be searched and retrieved using Z39.50.

Yes. The USMARC record syntax is a basic Z39.50 record syntax.

(f) Ability to store several alternative URLs in catalogue record.

Yes. This is possible as 856\$u (USMARC) is a repeatable field. "The field is repeated if more than one URL needs to be recorded" (MARBI, 1993d).

ADDITIONAL REQUIREMENTS FOR FURTHER MODEL DEVELOPMENT

(b) Ability to record username and password of commercial services in suppressible MARC field.

It seems unlikely that this is met as it is probably specific to CATRIONA.

(c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

It seems unlikely that this is met as it is probably specific to CATRIONA.

(e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.

Probably specific to CATRIONA and unlikely to have been considered.

(g) Ability to record Z39.50 information needed to set up client-server dialogues with non-MARC Z39.50 OPACs.

Unlikely that this has been considered

(h) Records must be in a form that is compatible with developing Internet standards

Compatibility is envisaged between MARC and Internet resource descriptions or characteristics (URCs). "URCs will be mapped algorithmically into and out of MARC records; to the extent they are designed with this in mind, the Net and existing libraries will work better together. MARC need not be the syntactical wrapper for URCs, but the rules for encoding MARC records (AACR-2) [Anglo-American Cataloguing Rules, second edition) should inform those elements of the URC that are similar" (Weibel, 1994)

Various efforts are also underway to map between MARC and TEI headers (See below) which were partly modelled on MARC (Gaynor, 1994).

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.

Yes, at a basic level of detail (e.g. title and URL in 856\$u).

(j) Ability to catalogue CD-ROM titles of networked CD-ROMs and record access information for them.

Extremely unlikely that this is met.

Note 1: USMARC Format

Data elements for Electronic Resources

< > indicates change proposed in the most recent MARBI Proposal 95-1 "Changes to Field 856 (Electronic Location and Access) in the USMARC Bibliographic Format", December 2, 1994.

FIELD 856 (Additional explanation in italics)

Indicators

First Access method

0 Email

1 FTP

2 Remote login

< 3 Dial-up>

7 Method specified in subfield \$2

Second Undefined

Undefined

Subfield codes

\$a Host name (R) (R = Repeatable subfield)

\$b (NR) (formerly IP Address)

(NR = Non-Repeatable subfield)

\$c Compression information (R)
\$d Path (R)
\$f Electronic name (R)
\$g Electronic name_End of range (R)
\$h Processor of request (NR)
\$i Instruction (R)
<\$j BPS (NR)> (Bits Per Second)
\$k Password (NR)
\$l Logon/login (NR)
\$m Contact for access assistance (R)
\$n Name of location of host in subfield \$a
\$o Operating system
\$p Port (NR)
\$q File transfer mode (NR)
<\$r Settings (NR)> (Settings used for transfer of
data e.g. parity, number of bits to signal
end of a byte, number of bits per character)
\$s File size (R)
\$t Terminal emulation (R)
\$u Uniform Resource Locator (R)
<\$v Hours access method available>
\$w Record control number (R)
\$x Nonpublic note (R)
\$z Public note (R)

\$2 Access method (NR) (http, gopher, news, nntp, wais, file,
prospero)

\$3 Material specified (NR) (contains information that
indicates the part of the item to which the 856 field
applies)

(Most of the above subfields were proposed in 93-4, but some
were later additions (MARBI, 1993c). Subfield \$u was added in
December 1993: Proposal 94-3 "Addition of Subfield \$u
(Uniform Resource Locator) to Field 856 Electronic Location
and Access in the USMARC Holdings/Bibliographic Formats").

Note 2: USMARC Data Elements for Online Systems and
Services

The following data elements were identified by MARBI as
required in USMARC records for online systems and services
in Proposal 94-9 (MARBI, 1994a). Some of the elements have
been accommodated in field 856. Others are accommodated in
existing Bibliographic or Community Information Format fields
and changes have been made to these fields in some cases
(ibid). Discussion continues as to the most appropriate
locations for some of the elements. (MARBI, 1994b)

Data Element	USMARC Field
Name of the Resource	Bib./CIF 245 (Title)
Acronym/Initialism	Bib. 246 (Varying Form of Title: Format Integration)
Producer	Bib. 260 or 245 \$c (Statement of responsibility)
Distributor of the Resource	Bib./Hold. 856 (Electronic Location and Access)
Location	Bib./Hold. 856\$n (Name of location of host)
Contact Name and Address	856 \$m
Network Address(es)	Bib./Hold. 856 \$a (Host name)
Hours of Service	CIF 301 \$a (Hours, etc.)
Telephone	CIF 270 \$k (Telephone number) or \$j (Specialized telephone number)
Fax	CIF 270 \$l (Fax)
Network Access Instructions	Bib./Hold. 856
Terminal Emulation Supported	Bib./Hold. 856 \$t
Logon/Subscription Instructions	Bib./Hold. 856 \$l, \$z (Logon/login or Public note)

Logoff/Unsubscribe Instructions	Bib./Hold. 856 \$z (Public note)
Type of the Resource	Code "i" in 008/26 for computer file (proposed here) Bib. 516 for specific type (e.g. Online Public Access Catalog, Computer Forum, Bulletin Board, etc.)
Size of Resource	Bib. 300 (for number of records, etc.); Bib./Hold. 856\$s (File size if appropriate)
Frequency of Update	Bib. 310 (Current Frequency)
Language of Resource	Bib./CIF 546 (Language Note)
Profile of Resource	Bib./CIF 520 (Summary, Abstract)
Audience	Bib./CIF 521 (Target Audience)
Restrictions on Access	Bib. 506 (Restrictions on Access)
Authorization	Bib. 506 \$e (Authorization)
Source Machine	Bib. 538 (Technical Details)
Cost for Use	CIF 531 \$b (Fees)
Coverage	Bib. 513 (Type of Report and Period Covered)
Indexing Terms	Bib./CIF 6XX (Subject added entries)
Databases Available	Bib./CIF 505 (Contents)
Other Providers of Database	Bib. 775 (Other Edition Entry)
Documentation Available	Bib. 556 (Information about Documentation Note) 040 (Cataloging Source)
Responsibility for Record Maintenance	
Date/Time of Last Update of Directory Information	005 (Date and Time of of Last Transaction)
Local Access Information and Guidelines	Bib./Hold. 856 \$z

F.3.2 IAFA Templates

IAFA (Internet Anonymous FTP Archives) templates were designed to assist in searching through the vast amounts of data stored at FTP archive sites. Although Archie was available to do this, the metadata used by Archie is very sparse. IAFA templates were developed in order to provide more information to those searching archives, and to put the information into a standard format, so that it could be machine-searchable. (Weider, 1994) IAFA templates are defined for a range of resources. These include users, organisations, mailing lists, FTP archives, text files, sound files, software packages and so on. (Deutsch et al, 1994).

Although IAFA templates are a recent development, they are already being used, particularly in the ALIWEB (Archie Like Indexing in the Web) retrieval service (See Appendix E), and as the basis for PICA resource descriptions (See Appendix E and Dempsey, 1994). However, IAFA templates are not at present a Z39.50 registered syntax and cannot be used with existing clients and servers.

IAFA templates are based on attribute pairs, with the number of attribute pairs and the content of some of the values is left up to the encoder, which may lead to retrieval problems as the encoder may well be inconsistent e.g. in choice of keywords, although the intention is to restrict the syntax and semantics used to that the templates can be automatically collected and indexed (Weider & Deutsch, 1994).

BASIC REQUIREMENTS

(a) Ability to catalogue any and every possible type of Internet resource and service.

Yes - requirement met at a basic level. A wide range of resources and services can potentially be described in IAFA templates:

IAFA Template types

USER, for individuals (also a cluster which appears in other templates)

ORGANISATION, for organisations (also a cluster which appears in other templates)

SITEINFO, for information about an FTP archive
 LARCHIVE, for information about a logical archive within a physical archive mirror
 'MIRROR', for mirror unit
 'SERVICE', for mailing lists, video and sound feeds, and other forms of interactive information sources
 (Includes "online library catalogues ... interactive online information services such as WAIS, gopher, prospero,
 WWW orarchie"
 'DOCUMENT', for text files
 'IMAGE', for image files
 'SOUND', for sound files
 'SOFTWARE', for software packages
 'MAILARCHIVE', for mail archives
 'USENET', for USENET archives
 'FAQ', for Frequently Asked Questions (Deutsch et al, 1994)

(d) Records must be in a form that can be searched and retrieved using Z39.50.

No. IAFA templates are not yet registered as a record syntax for transportation using Z39.50.

(f) Ability to store several alternative URLs in catalogue record.

Seems possible. Example of document record gives two URLs for different formats of the same text (ibid p34).

ADDITIONAL REQUIREMENTS FOR FURTHER MODEL DEVELOPMENT

(b) Ability to record username and password of commercial services in suppressible (MARC) field.

It seems unlikely that this is met as it is probably specific to CATRIONA, although in the SERVICE template there is an entry for authentication information. (ibid p9). See Note 3

(c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

It seems unlikely that this is met as it is probably specific to CATRIONA.

(e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.

Probably specific to CATRIONA and unlikely to have been considered

(g) Ability to record Z39.50 information needed to set up client- server dialogues with non-MARC Z39.50 OPACs.

Unlikely that this has been considered.

(h) Records must be in a form that is compatible with developing Internet standards

Possible. IAFA templates and URCs (the current central developing Internet standard of description) have similar data elements and parseable formats and therefore interworking may be possible. (Koster, 1994).

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.

Likely, at a basic level. The 'SERVICE' template covers "online library catalogues, interactive online information services such as WAIS, gopher, prospero, WWW orarchie" (Deutsch et al, 1994 p9)

(j) Ability to catalogue CD-ROM titles and record access information for them.

Extremely unlikely

Note 3: Fields for the IAFA SERVICE template.

Template-Type:	SERVICE
Title:	Title of service.
URI:	URI of service.
Admin- (USER*) :	Contact information of person or group responsible for service administration (administrative contact).
Owner- (ORGANIZATION*) :	Information on organization responsible for this service.
Sponsoring- (ORGANIZATION*) :	Contact information for the organization sponsoring this site.
Description:	Free text description of service.
Authentication:	Authentication information. Free text field supplying login and password information (if necessary) or other method for authentication.
Registration:	How to register for this service if general access is not available.
Charging-Policy:	Free text field describing any

	charging mechanism in place. Additionally, fee structure may be included in this field.
Access-Policy:	Policies and restrictions for using this service.
Access-Times:	Time ranges for mandatory or preferred access of service.
Keywords:	Keywords appropriate for describing this service.
(Deutsch et al, 1994 p29-30)	

F.3.3 URC

The URI (Uniform Resource Identifiers) working group of the IETF (Internet Engineering Task Force) has been developing means of discovering, identifying and retrieving resources. Resources are identified by a URN and retrieved by a URL. "Describing the resource for purposes of discovery as well as making the binding between a resource's name and its location(s) is the role of the URC (Uniform Resource Characteristic)". (Daniel, 1994b, p3)

In this section the URC is considered as a means of encoding meta-information about a given network resource (ibid). There is also a proposed URC service, in which URNs are resolved into URLs (Daniel, 1994b).

The URC as it is at the moment is designed to be simple to understand and encode, extendible, compatible with a wide range of systems and with existing technology. (Mealling, 1994). The disadvantage of this approach is that its simplicity and flexibility - records can be encoded by anyone, with the content of the title and abstract values left up to the encoder - may present retrieval problems.

Like IAFA templates the URC is based on attribute-value pairs. See Note 3 for the current minimal set, "although no attribute- value pairs are required to be part of a URC" (ibid).

BASIC REQUIREMENTS

a) Ability to catalogue any and every possible type of Internet resource and service.

Requirement met at a much lower level than in MARC or IAFA templates.

The emphasis of URCs at present seems to be on describing retrievable resources. The examples given (ibid) are for documents, i.e. text files. However, one of the requirements is that "it must be possible to add new types of elements to the URC without breaking previous applications" (Daniel, 1994b p16)

(d) Records must be in a form that can be searched and retrieved using Z39.50.

No. URCs are not yet registered as a record syntax for transportation using Z39.50.

(f) Ability to store several alternative URLs in catalogue record.

Yes. An example is given with two URLs. (Mealling, 1994).

ADDITIONAL REQUIREMENTS FOR FURTHER MODEL DEVELOPMENT

(b) Ability to record username and password of commercial services in suppressible (MARC) field.

No.

(c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

No

(e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.

No

(g) Ability to record Z39.50 information needed to set up client- server dialogues with non-MARC Z39.50 OPACs.

No

(h) Records must be in a form that is compatible with developing Internet standards

Yes. The URC is the current central developing Internet standard of description.

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.

Likely at a basic level.

(j) Ability to catalogue CD-ROM titles and record access information for them.

No

Note 4: URC Structure

Attributes:	Values:
URN	Uniform Resource Name
URL	Uniform Resource Locator
LIFN	Location Independent File Name (to be developed)
Author	Author of the Item (there are no requirements as to how a name should be written)
TTL	Time To Live (measured in seconds, this sets a time limit on the attribute/value pair preceding it and is meant as a caching aid)
Abstract	A short abstract about the item. Any characters are admissible (Mealling, 1994).

Note 5: Examples of URC attribute-value pairs

Attribute:	Value
URN:	IANA:626:oit.5647
URL:	http://www.gatech.edu/ietf/urc.encoding.html
LIFN:	md5:2039874029834059283709475029387405 928374095827394875
Author:	Michael Mealling
TTL:	86400
Abstract:	This document explores the various flight patterns and speeds of unladen African and European swallows. A companion document concerning the relative velocities of swallows laden with coconuts is available. (Mealling, 1994).

F.3.4 TEI Header

The TEI (Text Encoding Initiative) has developed guidelines for the preparation and exchange of electronic texts mainly for research in the humanities. Meta-information relating to the encoded text is contained in the TEI header which can be in a different file from the main text (as an Independent header). The TEI header consists of four elements: File Description; Encoding Description; Profile Description and Revision Description. The only mandatory element is File Description, which provides a bibliographic description very similar to standard descriptions: ACCR-2, ISBD (International Standard Bibliographic Description) (G) and USMARC. (Giordano, 1994a) (See Note 6 for key elements in FileDesc).

TEI headers have advantages in that they are flexible - there are very few mandatory elements, and they make use of the standard language SGML. There are potential retrieval problems, however, as the values are mostly open (ibid). Other limitations for use in the CATRIONA model is that access via URL is not yet accommodated, and the resource descriptions required in the CATRIONA model must cover a wide range of systems and services as well as electronic texts.

BASIC REQUIREMENTS

(a) Ability to catalogue any and every possible type of Internet resource and service.

No. TEI has been developed for encoding and describing electronic texts.

(d) Records must be in a form that can be searched and retrieved using Z39.50.

No. TEI headers are not yet registered as a record syntax for transportation using Z39.50.

(f) Ability to store several alternative URLs in catalogue record.

No. There is place at present allocated for a URL in a TEI header. (Giordano, 1994b)

ADDITIONAL REQUIREMENTS FOR FURTHER MODEL DEVELOPMENT

(b) Ability to record username and password of commercial services in suppressible (MARC) field.

No.

(c) Ability to record that this is a special catalogue record only to be retrieved/displayed if the search is a local one.

No.

(e) Ability to catalogue and subject index other Z39.50 OPACs and record Z39.50 information needed to set up client-server dialogues.

No.

(g) Ability to record Z39.50 information needed to set up client- server dialogues with non-MARC Z39.50 OPACs.

No.

(h) Records must be in a form that is compatible with developing Internet standards.

Mapping with MARC is possible with some of the elements, although the TEI header contains information which does not have equivalent MARC fields. There has also been recent discussion of a TEI-URC relationship (Herr-Hoyman, 1994).

(i) Ability to catalogue Internet indexing services such as Archie, Veronica etc.

No.

(j) Ability to catalogue CD-ROM titles and record access information for them.

No.

Note 6: TEI Header Elements

The following is a list of a few of the key components of the TEI header selected from an extensive list of possible elements.

- * required. Some subelements are required, others optional or recommended:
 - required; subelements are required or optional:
 - recommended
 - optional
 - required
 - optional
 - recommended
 - * required. As much information as possible should be provided to identify the source.
 - * very highly recommended, especially for projects, collections, or corpora.
 - * recommended
 - * required in the independent header when available.
- (Sperberg-McQueen & Burnard, 1994, Chapter 24)

F.3.5 Other Developments: GILS

GILS (Government Information Locator Service) is a US-based project promoted by the US Information Infrastructure Task Force and the Office of Management and Budget. The purpose of GILS is to identify and describe government information resources and provide help for users to obtain the information. The public will be able to use GILS either directly through the Internet or indirectly through various intermediaries including public libraries. GILS will use the Z39.50 standard in order to make information as widely available as possible. Because GILS will be a decentralised service, there will be a wide range of implementations and the need to agree on a reference base of stable standards including Z39.50 to achieve interoperability was recognised. In addition, a GILS Profile has been developed which gives a complete specification of GILS as it makes use of ANSI Z39.50, but also specifies the features of GILS that are outwith the scope of ANSI Z39.50. In the GILS draft profile it is planned that GILS records will be available in three Z39.50 record syntaxes: USMARC, GRS and SUTRS.

A GILS Core will be set up including records for all information locators that catalogue publicly accessible government-funded information resources. It is expected that the Core will be no more than 100,000 locator records.

The locator records will conform to the GILS Core Element standards. The definitions of the Core Elements identify the corresponding USMARC field for each element. Listed below are the mandatory elements as well as optional elements of relevance to the CATRIONA requirements.

Note 7: GILS Core Elements

Title:	(USMARC Tag 245\$a)
Control Identifier:	(USMARC Tag 001)
Abstract:	(USMARC Tag 520)
Purpose:	(USMARC Tag 500)
Originator:	(USMARC Tag 710\$a)
Access Constraints:	(USMARC Tag 506)
Use Constraints:	(USMARC Tag 540)
Availability:	(USMARC Tag 037\$b)
Resource Description:	Optional sub-element (USMARC Tag 037\$f)
Order Process:	Mandatory sub-element (how to obtain the information resource from this distributor) (USMARC Tag 037\$c)
Technical Prerequisites:	Optional sub-element (USMARC Tag 538)
Available Time Period:	Optional sub-element
Available Linkage:	This optional sub-element occurs no more than once per Availability element. It provides the information needed to contact an automated system made available by this distributor, expressed in a form that can be interpreted by a computer (i.e., URI). Available linkages are appropriate to reference other locators, facilitate electronic delivery of off-the-shelf information products, or guide the user to data systems that support analysis and synthesis of information. (USMARC Tag 856\$u)
Available Linkage Type:	This optional sub-element occurs if there is an Available Linkage described. It provides the data center type (i.e., MIME) for the referenced URI. (USMARC Tag 856 first indicator/856\$2)
Point of Contact for further information:	This mandatory element occurs once per locator record. (USMARC Tag 856\$m for electronic resources)

(Christian, 1994)

[Return to Contents Page](#)

Appendix G. CATRIONA - A Case Study: Napier University Library

G.1. Background

Napier University is a new University, formerly Napier Polytechnic of Edinburgh, spread across several campuses. Napier University Library (NUL) has full-service library facilities at three major campuses, with a fourth and largest campus currently under development. The student population is just over 9000, and there are 1000 staff. The University has an Ethernet LAN which is scheduled for upgrading and connection to the Edinburgh Metropolitan Area Network (MAN). Access to the Library automated catalogue is via the LAN; there are no public terminals or PCs connected directly to the Library system.

NUL uses the Dynix library automation system. During the selection process for this system, the Depute Librarian wrote to the Technical Director of Dynix Library Systems (UK) Ltd at the beginning of 1990 regarding development of the system:

The Library wishes to develop the existing environment ... to attain the following:

1. The Dynix integrated catalogue should act as a shell to a file of machine-readable documents such as Polytechnic papers, AV material, contents descriptions and evaluations, etc. System users should be prompted that a particular catalogue record has such a document attached to it, and should be given the opportunity to display the document contents on screen and/or a printer. ... Certain users should be able to download the document file onto a personal workstation, amend it using Polytechnic standard software, and upload back into the file.
2. ... A graphical user interface should allow users to display Dynix data on screen ... and to cut and paste data from Dynix ...
3. ... users should be able to use Dynix, DOS application and other machine-readable data in an integrated information environment, allowing the transfer of data in a simple manner between the various system files. Dynix functions and data should be fully integrated with other applications such as word- processing, spreadsheet analysis, image storage and retrieval, and multimedia information analysis and processing.

To some extent, the first development has now been attained, and the latest Dynix upgrade (currently planned for installation at Napier towards the end of 1995) allows graphical images to be attached to standard MARC records. Editing and uploading facilities are not currently well-developed. The second objective was reached by using a shareware Windows terminal emulation package, which allows multiple Dynix sessions to be displayed on the same Windows desktop, and for data to be moved between them using the Windows Clipboard. The third objective is under development by Ameritech Library Services, the current owners of Dynix, using a Windows environment.

The Strategic Plan 1994-98 of NUL contains proposals which include:

Library emphasising service to clients, but trying to increase their self-help and self-service capabilities

Considered mixes of printed, audiovisual and electronic information provision and of policies for holdings and access

Electronic access ... to research information from JANET and the Internet, University-generated teaching packages and University administrative, advisory and regulatory information

G.2 Current Situation

NUL believes an integrated, one-stop, public-access catalogue to be the best tool to meet Strategic Plan objectives on self- service and a mixture of information formats. The catalogue comprises UKMARC standard records, using Anglo-American Cataloguing Rules, 2nd edition, 1988 revision, 2nd level of description, for most of NULs stock of physical formats, including monographs, serials, sheet music, music scores, video recordings, audio Compact Discs, and stand-alone CD- ROMs. In addition, attempts have been made to catalogue CD-ROMs and Windows/DOS packages which are available only via the University LAN, and therefore have no physical manifestation as such. The resulting integration of access to these resources via the catalogue is illustrated by Screen 1.

Screen 1: Integrating resources concerning the Guardian newspaper: title search

Alignment of title numbers indicates whether the material is available at the local campus library.

	Your search: The Guardian	
	TITLE/AUTHOR	EDITION
DATE		
	1. The bedside Guardian Webb, W. L.	27
1978	2. The Guardian	
	3. Guardian: biography of a newspaper Ayerst, David	
1971		

4. The Guardian on CD-ROM
Guardian Newspapers Ltd
5. The Guardian on CD-ROM including The Observer
Guardian Newspapers Ltd
6. Raising finance: the Guardian guide for the sma
Woodcock, Clive 2nd ed

1985
- - - - 6 titles, End of List - -
- -

This screen integrates access to several resources:

- Title 1 is a book containing extracts from the newspaper.
- Title 2 is the printed newspaper.
- Title 3 is a book about the newspaper.
- Title 4 is the previous title of the newspaper archive on CD- ROM
- Title 5 is the current title of the newspaper archive on CD- ROM.
- Title 6 is a book written by staff of the newspaper.

Selecting any particular title results in the display of more detailed descriptive information in card format, as in Screen 2.

Screen 2: Card format display of selected title

DYNIX BIB # 104199

TITLE	The Guardian on CD-ROM including The Observer
	The Guardian The Observer The Guardian on CD-ROM
NAME(S)	1) Guardian Newspapers Ltd
PUBLISHER	[London] ; Chadwyck-Healey ;
SIZE, ETC.	computer laser optical disk
NOTES	Issues cumulate within the year Continues: The Guardian on CD-ROM, 1994 'Guardian Newspapers Limited'
SUBJECT	1) 072.1
FORMERLY	The Guardian on CD-ROM
FREQUENCY	Quarterly

- - - - End of Title Info - - -

Location and availability status is displayed on a subsequent holdings screen. The CD-ROM versions have no physical format because they are mounted on the LAN, and therefore have no particular campus location and are always available. There are no data links between the MARC records in the catalogue and the electronic files because Dynix cannot support them; instead, the link is made explicit to the user in a dummy holdings record, as in Screen 3. This information could also be incorporated as a local note, MARC tag 509, in the bibliographic record; it has to be attached to the holdings record in this instance because it does not apply to the other holdings which refer to the manuals.

Screen 3: Dummy holdings record instructing user how to access the resource

Hyphens are used to fill the mandatory Shelfmark field in each holdings record

Author Guardian Newspapers Ltd
Title The Guardian on CD-ROM including The

Observer Holds: 0

	#	CALL #		STATUS	BARC
ODE		LIB			
	1.	Circulation desk		On shelf	002185245
S		- Manual			
	2.	Circulation desk	On shelf	002071171	
M		- Manual			
	3.	Circulation desk	On shelf	002071510	
C		- Manual			
	4.	See Network menu	CD-ROM Network	-144891	
		CN			
		-			

Holding 4 refers to the CD-ROM mounted on the LAN. It would be misleading for users, and cause operational problems, to locate the CD-ROM at any of the campus libraries; NUL has therefore set up an extra library, coded CN, to indicate where the material is held. A barcode is mandatory for each holding; in this case a dummy or negative barcode is used because the resource has no physical manifestation.

Users who wish to access the CD-ROM must log-off from the catalogue, and log-on to the University Network CD-ROM Menu, from where the title can be selected. When finished with the CD-ROM, users must log-off from the CD-ROM Menu, and log-on to the catalogue if they wish to continue searching for Library materials.

In addition to networked CD-ROMs, NUL is having to cope with an increasing number of other electronic resources, held locally on the LAN or available via JANET. These include DOS abstract and indexing services, multi-media packages, BIDS, and the OPACs of other libraries. In all instances, dummy records have to be used in the catalogue, and users have to carry out a cumbersome switch between different systems to access the resources they have identified.

G.3 Developing the Current Approach

NUL anticipates a rapid increase in the number of electronic information resources it provides. In addition to the non-print resources already mentioned, NUL expects to have to stock or provide access to Windows/Multi-media teaching and reference packages, CD-i discs, and CD-Video discs, all of which could be mounted on the LAN and connected directly to an appropriate PC workstation. Many print-based reference services are being replaced by electronic versions distributed by both local carriers such as CD-ROM or DOS file and remote dial-up systems similar to BIDS. NUL also wishes to offer seamless access to Internet resources and services, for researchers and students, which have been deemed of interest and relevance. NUL does not believe that menu-driven, hierarchical access tools will be able to cope with more than a few hundred electronic resources, whether on the LAN, WAN or MAN, and in any case requires that access and navigation tools for these resources to be integrated with those for traditional resources held as physical items in stock so that self-service is encouraged, and Library staff are not required to mediate between identification of a resource and access to it.

For example, it is expected that the Guardian will introduce an online version of the newspaper to be accessed over the Internet/JANET, in the very near future, as several other newspapers have done. It may also be the case that another organisation offers a service collating literary and arts reviews from several UK national newspapers, including the Guardian, and mounts the results on the Internet. Other hypothetical products could include a CD-i on editing a newspaper, using the Guardian as the example, and an online file of images of press photographs, etc. NUL would attempt to integrate identification of these extra resources by creating catalogue records similar to the CD-ROM resource, with associated dummy holdings and explicit connection details, as in Screens 4 and 5.

Screen 4: Potential development of the existing approach: title search

```
Your search: The Guardian
              TITLE/AUTHOR                                EDITION
DATE
1. The bedside Guardian                                27
   Webb, W. L.
1978
```

2. Editing the Guardian: an interactive experienc

Compact Eyes Productions

3. The Guardian

4. Guardian: biography of a newspaper

Ayerst, David

1971

5. The Guardian on CD-ROM

Guardian Newspapers Ltd

6. The Guardian on CD-ROM including The Observer

Guardian Newspapers Ltd

7. The Guardian online

Guardian Newspapers Ltd

8. The photo-journalist: an archive of famous pho

Press Photo Archives plc

9. Press reviews: an Internet service

Reviews in the News, Inc.

10. Raising finance: the Guardian guide for the sma

Woodcock, Clive

2nd ed

1985

- - - -10 titles, End of List - -

- -

Screen 5: Potential development of the existing approach: title
detail screen

DYNIX BIB # 999999

TITLE The Guardian online
The Guardian
The Observer

NAME(S) 1) Guardian Newspapers Ltd

PUBLISHER [London] ; Guardian Online Ltd ;

NOTES Online access via JANET
Access available to registered users of the
University Network
Mosaic:

<http://gol.news.co.uk:6560/www/gol.html>

SUBJECT 1) 072.1

- - - - End of Title Info - - -

-

In this case, location and availability information is incorporated into the MARC record; the dummy holdings record would be similar to the existing example. The user is still required to log- off from the catalogue, having noted the http address, and log- on to a University JANET/Internet browser service. If the user subsequently needs to check the catalogue, say for an original printed copy of an issue of the newspaper to see its typography and page layout, then the whole process has to be explicitly reversed.

Another anticipated problem with this method is the proliferation of standard instructions and location codes needed to given explicit access information. Access may be via one of several different direct, dial-up facilities or pre-set menu systems supporting World-Wide Web, Gopher, Telnet, etc.

A further problem with providing explicit connection information is that it may change. The catalogue record will require maintenance to ensure the correct URL or other network address, whatever model evolves. The user may, however, try to use an old address, noted during a previous catalogue search, without checking the latest information. This is analogous to a user going direct to the shelf for a book previously consulted, without checking the catalogue, only to find the item has been reclassified, respined, and transferred to another branch library.

G.4 The Preferred Approach

NUL anticipates a major potential problem in the sheer quantity of networked resources that may require cataloguing or some other access and navigation system. NUL would seek to identify and provided integrated access to resources likely to be of a high interest and relevance to its users, but would also wish for extended access to other similar resources via other libraries. This idea already applies to physical resources: if a user cannot find what they want in NULs catalogue, they are encouraged to search the catalogues of other libraries in the vicinity, for which there are reciprocal user agreements. NUL sees the use of Z39.50 protocols as an essential development to allow extended searching beyond NUL itself. NUL is not concerned with the source of the catalogue records: they may be held in the local database, or the catalogue of another local library, or by the resource provider. The resource has no physical manifestation and its location is networked information space; concepts of shelfmark, library, holdings, circulation status, and other storage parameters, are replaced by filename, URL, usage license, browser, viewer, and other access parameters.

The preferred approach for NUL is thus a system which allows users to identify all types of resources held locally, using a common search interface, and to extend the search to resources held by other libraries without reformulation. If the identified resource is electronic, the user should be able to access it within whatever license arrangements have been made, by proceeding directly from the catalogue record without having to reverse and take another route. Once use of the resource is complete, the user should find themselves back at the catalogue record, ready to proceed to check other retrieved resources, or to carry out another search.

In other words, disregarding any other aspects of a graphical user interface, NUL would like to offer its users Screen 6 rather than Screen 5.

Screen 6: The active catalogue record

DYNIX BIB # 999999

TITLE	The Guardian online The Guardian The Observer
NAME(S)	1) Guardian Newspapers Ltd
PUBLISHER	[London] ; Guardian Online Ltd ;
SUBJECT	1) 072.1
ACCESS	Click on button to use this service

- - - - End of Title Info - -

--

[Return to Contents Page](#)

Appendix H. Reports of Meetings on Z39.50 Developments

H.1 Report of a Joint Meeting on Z39.50

The Library Association Information Technology Group (Central Scotland) and the Cataloguing & Indexing Group in Scotland held at joint meeting on Z39.50, with a demonstration of the SALSER database, at Glasgow University Library during the afternoon of 18 January 1995.

The principal speaker, on Z39.50, was Lorcan Dempsey, Director of the UK Office for Library Networking.

Lorcan began his talk by noting the growth of the network environment, the existence of location devices such as URL and URN, and the importance of meta-data such as finding lists and library catalogues. Other features included

client/server system architectures, and a number of different resource spaces such as Telnet, WWW, Gopher, and FTP. Although these spaces are becoming linked, most network resources of interest to libraries used Telnet, which has many shortcomings. Increasing connectedness between systems or resources was resulting in a lot more data flowing around networks.

The library resource space is currently characterised by monolithic applications using terminals as the main peripheral device. It is fragmented, and navigation tools are small-scale and not as elegant as they might be (clunky).

The establishment of the Z39.50 standard is an attempt to resolve some of these problems. It is a protocol for translating a local database search request into a standard form which can be interpreted by a foreign database into its own searching procedures, so that the original request does not require rekeying or reformulation. Although it currently focuses on searching and retrieving bibliographic information, it is intended to be generalisable to non-bibliographic datasets such as statistics, chemical formulae, etc. The results of a Z39.50 search are always in the form of a set of records which may be in the format, for example, of a standard OPAC card image.

A number of standard Z39.50 attribute sets are evolving, to cover different datasets; these include bib_1 for bibliographic data, GILS for US Government information, and STAS for chemical data. These attribute sets define search types, such as Title, and operators, such as =, which can act on them. Several different record syntaxes are currently catered for, including MARC, SUTRS (for simple text), and the aforementioned OPAC (simple bibliographic data with some holdings and circulation status data).

Projects proposed, underway, or completed, using Z39.50 include DBV-OSI2 in Germany, IRIS in Ireland, LONDON LINK, and ALSA. The last had been proposed for FIGIT funding by UKOLN, but had been turned-down. The British Library is working on a project for a general OPAC to European national libraries, and a lot of work is being carried out in the Netherlands by PICA.

The following points were made in summing up the current situation:

Z39.50 itself is only a part of the network picture.

Distributed library systems are in an immature state.

There is no driving focus in the UK for a distributed, virtual library.

There is no culture of resource-sharing in the UK, an essential component of the network approach.

There are no significant UK resources available to improve this situation.

In response to questions regarding the enhancement of Z39.50 searching to provide search control structures such as menus for databases to be searched, time-out facilities, and automatic search halting if a hit was made on a unique key such as ISBN, it was said that little progress was being made, although implementors were aware of such issues.

After a coffee-break, the meeting was given a demonstration of the SALSER system by Bruce Royan, Director of Information Services at the University of Stirling. SALSER is a finding tool for serials holdings in Scottish libraries. It essentially consists of a Union list of serial titles, with brief location and holdings information. It has been designed so that the SALSER serial title can be linked to the full catalogue of the holding libraries using Z39.50, so that issue detail and availability can be displayed to the user. Eventually, the whole system may be supplanted by a general Z39.50 interface, but this will not happen until every member library has an automated serials catalogue, with an appropriate Z39.50 client.

H.2 Z39.50 Pre-Implementors Group Meeting

Aston University Library, 18 October 1994

The meeting began with a demonstration of the SLS Libertas 6.3 Z39.50 client. The facility to select groups of servers for distributed searching was effectively demonstrated. (See Appendix D for a list of client features).

Status reports of Z39.50 implementation were given from the British Library, Manchester Computer Centre,

Specialist Computer Group, Bangor University Library, Fretwell Downing and Strathclyde University Library. The current status of the CATRIONA project was reported on and there was an opportunity to present the aims of the project in general terms.

Fretwell Downing gave a description of the related IRIS and DALI Projects (See Related Projects and Developments). The work of EWOS EGLIB was then presented by Erik Lorenz Petersen, and was followed by a discussion in which concern was expressed at the need for progress of European information retrieval standards development in relation to what is happening in the US.

It was decided to host a Z39.50 seminar in January 1995 (Z39.50 UK Wide: What's Going On?) in order to inform the LIS community of what is happening in Z39.50 implementation in the UK.

H.3 Aston University Seminar: Z39.50 UK Wide: What's Going On

?

A seminar at Aston University organised on behalf of the Z39.50 Pre-Implementor's Group Aston University Library, 24 January 1995

The seminar, which was attended by systems suppliers, software vendors and LIS professionals, was designed to introduce the latest developments in Z39.50 implementation to the LIS community. A keynote address on Z39.50 strategy in the UK was given by Lorcan Dempsey of UKOLN. After outlining the Z39.50 protocol, he emphasised the current obstacles to progress in the UK: the immaturity of current distributed library systems; the lack of significant UK resources accessible via Z39.50 and lack of resource-sharing; and the need for experience of how the servers will work. This was followed by descriptions of a wide range of UK and European Z39.50/SR projects. The CATRIONA project was mentioned and was enquired about in during the question time which followed.

The major part of the meeting was allocated to systems suppliers who demonstrated various clients or outlined development plans. SLS gave a demonstration of Libertas 6.3, GEAC gave a demonstration of GeoPac, and Bryan Alton of IRIS Document Delivery Services Limited gave a demonstration of the IRIS client. This was followed by several updates on development plans. The MIRO System Architecture for searching online hosts and which uses Z39.50 was outlined by Satellites International. Neil Smith gave a presentation of the British Library plans to develop a Z39.50 server and discussed the ION project. The DALI project and the future trend towards providing services to the student desktop was outlined by Fretwell-Downing, as well as the need for LIS professionals to become involved as gatekeepers. Peter Smith of LASER discussed various aspects of their projects, including London Link. The position of BLCMP was described by Robert Watson who said that they may make their BLCMP database accessible via Z39.50 if there is sufficient demand.

The meeting closed with an observation from Lorcan Dempsey that the proposals for the Libraries Programme in the EC provide evidence that there is a need to bridge the awareness gap between those who know what the technology can do and those who are responsible for providing funding. He indicated that UKOLN would be interested in supporting such measures, which might also involve the Z39.50 PIG.

H.4 Meeting of the SCOLAR Technical Sub-Committee on Z39.50

The meeting of the SCOLAR Technical Sub-Committee which was held on 5 December, 1994, was addressed by Ameritech representatives Tom MacDonald, Maribeth Ward and John Kolman. After an outline of Z39.50, the Ameritech client WinPAC which was originally developed by NOTIS, was introduced. (WinPAC is one of the clients that was examined during the CATRIONA project. See Appendix D). Future plans for WinPAC which include expanded access were discussed, as well as the need for developments on the Z39.50 server front such as different indexing techniques and more sophisticated software and hardware.

John Kolman then discussed Z39.50 in more depth, detailing the various requests and responses involved in a Z39.50 session. The potential interoperability problems between clients and servers which can arise due to differing combinations of search types, attribute sets and record syntaxes selected were described. Finally the developments proposed for Z39.50 Version 3 were mentioned including the record syntaxes SUTRS and GRS, as well as the Explain facility which allows an origin to obtain details of the implementation of a target, including databases available for searching, attribute sets and record syntax.

Appendix I.

BUBL Subject Tree Project and Relationship to CATRIONA

I.1 BUBL Subject Tree Initiative

The BUBL Subject Tree Initiative was begun in the final quarter of 1993 and is now a significant part of the BUBL Service with over 5000 links by June 1994. BUBL was the first national UK service to offer subject-based access to Internet services and resources and is still the only UK national service whose subject tree covers all main subjects. It is also unusual, perhaps unique, in that it provides a composite tree covering both Gopher-based resources and World-Wide Web resources. The tree may be accessed in a number of ways:

- JANET (Joint Academic Network) X.29 call
- TCP/IP Telnet call
- via Gopher
- via World-Wide Web

However, it is seen at its best if accessed by a World-Wide Web client. The URLs for the service are:

<http://www.bubl.bath.ac.uk/BUBL/Tree.html>

and

<http://www.bubl.bath.ac.uk/BUBL/Treealphabet.html>

One aim of the initiative was to provide the academic community in the UK with improved access to Internet services and resources. It demonstrably improves the user's interface with these resources and so improves the quality and efficiency of the access. The project has other aims which are equally innovative:

- to encourage the UK LIS community to exploit academic networks for co-operative and work-sharing purposes
- to train and educate LIS professionals by involving them in a real network project
- to establish a role for LIS professionals in the provision of network-based information services.

The project seeks to meet these aims by involving the LIS community in the creation and maintenance of the Subject Tree. Although the main tree was largely created by the BUBL team, they were also assisted by a large group of subject specialists distributed across the UK (and beyond). This group also take responsibility for the maintenance and further development of some parts of the tree and other specialists are being sought to cover those subject areas not already covered. They communicate with each other and with BUBL (and NISS - see below) via the Mailbase electronic mail discussion list LIS- Subjects. The Subject Tree Initiative is, in fact, only one of the many ways in which BUBL seeks to meet its aims. The primary aim of the service is to encourage, develop, co-ordinate and support the emerging LIS networking community in the UK and to promote its interests.

It is a fundamental assumption of the BUBL enterprise that its various sub-projects are examples of good practice. The Subject Tree Initiative is a good illustration of this, the following characteristics of it being regarded as examples of good practice, embodying principles which are capable of being applied in other contexts:

- the exploitation of networks for co-operation and work sharing purposes
- the application of LIS skills to the networked environment
- the training of LIS professionals through active involvement in real projects

The following description of the Subject Tree Initiative should give a clearer picture of the initiative and how it works:

The Subject Tree is based on the Gopher and World-Wide Web software technologies. These enable BUBL to provide access both to files held on the BUBL machine itself and, more importantly, to other remote services and resources. Link information like the following:

GOPHER

Name=BUBL Information Service

Type=1

Port=7070

Path=

Host=ukoln.bath.ac.uk

WORLD-WIDE WEB

<http://www.bubl.bath.ac.uk>

allows the user to jump from an item found under a particular subject heading to either a file held on BUBL or a service or resource anywhere on the Internet. The task of the subject specialist is to discover valuable resources and transmit either the resource itself or the link information to BUBL (and also NISS) with an indication of which part of the Subject Tree it should go under. The use of gopher and WWW links allows items to be 'pointed at' from a number of subject areas thus solving the difficulties of resources of interest in a number of subject disciplines.

Resource discovery takes place via a number of mechanisms:

- searches of the Internet using Veronica and other tools
- searches of subject guides on BUBL
- browsing of other subject trees and services in the world at large
- joining electronic mail discussion lists with a particular subject orientation

Once a resource is discovered, either the resource itself or the link information is sent to the Mailbase list LIS-Subjects from which it is picked up by BUBL and also by NISS.

I.2 The BUBL Subject Tree and CATRIONA

BUBL regards the Subject Tree as a useful interim measure. In the long term the only solution resource discovery on the Internet is likely to be based on local cataloguing and national and international co-operative cataloguing. It is also likely that the voluntary effort on which the BUBL Subject Tree is based will have problems in scaling up to cope with the expansion of electronic information. These are some of the reasons behind the CATRIONA project which arose from the BUBL Subject Tree Initiative. If CATRIONA Phase II goes ahead it may ultimately replace subject trees by replacing them with an improved service, but this is not likely to happen for some time to come.

I.3 Further Information

Further information on the CATRIONA feasibility study, on other related information and projects, and on the BUBL subject tree will be found on the BUBL Information Service at URLs:

<http://www.bubl.bath.ac.uk/BUBL/maincatriona.html>

<gopher://bubl.bath.ac.uk:7070/11/Link/Catriona>

<http://www.bubl.bath.ac.uk/BUBL/Tree.html>

<http://www.bubl.bath.ac.uk/BUBL/Treealphabet.html>

<gopher://www.bubl.bath.ac.uk:7070/11/Link/Tree>

This information will be added to and enhanced by the BUBL Information Service as information vital to the library community and its users regardless of whether or not the CATRIONA project continues into Phase II and beyond.

[Return to Contents Page](#)

Appendix J. List of Project Participants, Advisors and Correspondents

The organisations and individuals who participated in the CATRIONA Project are listed below. Participation was at

a wide range of levels.

GEAC Computers Limited
DYNIX Library Systems (UK) Ltd.
Fretwell-Downing Informatics Ltd
SLS Information Systems
SIRSI Limited
MDIS (McDonnell Information Systems)
SCG (Specialist Computer Group)
SilverPlatter Information, Ltd
OCLC (Online Computer Library Center)
RLG (Research Libraries Group)
CURL (Consortium of University Research Libraries)
SCURL (Scottish Confederation of University and Research Libraries)
CIGS (Cataloguing and Indexing Group Scotland)
BLCMP Library Services Ltd
Lorcan Dempsey, UKOLN: the UK Office for Library and Information Networking
Hunter Monroe, Eric Lease Morgan (Alex Project - See Appendix E)
The British Library
Glasgow University Library
Stirling University Library
Abertay University Library
Paisley University Library
Edinburgh University Library
Glasgow Caledonian University Library
Strathclyde University Library
National Library of Scotland
Napier University Library
Heriot Watt University Library
Aberdeen University Library

[Return to Contents Page](#)

Appendix K.

Workshop on Cataloguing Electronic Resources

This Workshop was one of two presented during the 7th annual OPAC Forum, organised by Cataloguing & Indexing Group in Scotland (CIGS) and held in the Library of the University of Abertay Dundee on 8 November 1994. The Workshop leader was Gordon Dunsire, Information Systems Librarian at Napier University, Edinburgh; the notes on which this report is based were taken by Lynn Corrigan, Assistant Librarian (Information Systems), also of Napier University.

The Workshop was structured in two parts: a short presentation by the Workshop leader; followed by an open discussion on the issues raised.

Presentation

The presentation gave an overview of some of the issues concerning resources, records, and the catalogue.

Many libraries purchase electronic resources such as CD- ROMs for local use; these may substitute for printed material. Such resources can be catalogued in an integrated fashion with other resources; the information contained is the primary aspect, not the physical format. Nowadays, however, we also have to contend with networked information - virtual resources which have no physical manifestation. Existing cataloguing rules blur the distinction between intellectual works and their physical format because the mode of use or access may very well affect the intellectual structure of the work. Current rules are weak when physical manifestation is nebulous. A second problem is that Internet and other networked resources have no fixed location. They can exist in more than one

location, or the location can change over time. Maintenance of the catalogue record becomes complicated. The principles of ownership, copyright and access are also more difficult to apply under these circumstances.

Electronic resources can be direct equivalents of their analogue forms, including text, graphics, and sound. To access machine-readable text, hypertext and multimedia software, users require programs such as viewers and browsers, often specific to the resource. Such programs may give users the means to alter a document - electronic resources are particularly amenable to change.

Documents need no longer be flat; as well as being non-linear or multidimensional, they may no longer represent a single intellectual concept.

Electronic resources can be primary, secondary or tertiary, in the traditional sense. An electronic resource may be a work, an abstract of the work, or a catalogue of records of many works. Electronic catalogues of catalogues are urgently required on wide area networks; the electronic manifestation of Russells catalogue of catalogues which reference themselves is just around the corner.

The existing international cataloguing standards of AACR, MARC, and ISBD have evolved in a print-based culture. While they are adequate for describing intellectual works, they have shortcomings in an electronic environment. ISBD has little to say on displaying bibliographic information in graphical user interfaces, AACR lumps all machine-readable files together, and MARC requires modification to store data associated with electronic information objects (EIOs). Additional MARC tags and subfields are needed to store the electronic address of an EIO, analogous to the shelfmark of a physical resource. Information about access, including passwords, log-ons, and editing and copying rights is much more resource-specific than for traditional materials, and should also be stored as part of the bibliographic record. The implicit access route for a physical library may be to join the library and obtain a readers card; the electronic equivalent is likely to be different depending on the particular EIO, and must be made explicit to the user to avoid the need for professional mediation.

A trend in automated catalogues is increased linkages between records. Implicit links are a primary feature of bibliographic catalogues, in the shape of authors, titles, and subjects. Explicit links are now seen as desirable for improving access to and integration of information; these links can be between analytic entries and the principal bibliographic record, or between electronic abstracting and indexing products and bibliographic and holdings records. Analytics share the lack of a separate physical presence with EIOs; they are embedded in other works. Hyperdocuments may consist of intellectually separate works linked together electronically; again, concepts of intellectual ownership have to be revised. Such links may be dynamic, and may be amended automatically by intelligent software; even the concept of a work needs reappraisal. The electronic catalogue of electronic information resources thus faces problems concerning description, location, access (to the resource), content, ownership, version (or edition), and integration, all of which are subject to a much greater degree of change than for traditional resources.

The main advantage of such a catalogue is its own manifestation: it does not need to be physically integrated to be electronically integrated. The records do not have to be stored in a single data file; they can be brought together from many different files to create a virtual catalogue. From the users point of view, it does not matter whether the resource is accessed in China or Canada, and it is of no interest that the catalogue record being used for access is itself stored in Dundee or Dresden.

Discussion

The fundamental assumption, that this meta-catalogue is a good thing, was queried. Many users are only interested in local resources, and the old-fashioned OPAC might be sufficient for their needs. It was suggested that local really meant available, and in that sense all network EIOs could be considered local if it was possible to deliver them to the OPAC or workstation screen.

The CATRIONA Project was mentioned; this aims to produce a catalogue that integrates resources of all types and provides a navigation tool to identify EIOs. The catalogue approach reduces netsurfing and prevents time-wasting; on the one hand there is no need for random browsing (surfing), and on the other access is controlled by the catalogue record itself.

It was pointed out that there is confusion between the catalogue and deskpace. The same PC or other

microcomputer could be used to access the OPAC and to download and manipulate the actual EIO. In some circumstances, users would not be able to carry out a quick search on the OPAC for local, physical materials because all machines were being used to prepare bibliographies. There might therefore be a need to consider disabling the retrieve facility on some OPACs so that they were dedicated to standard catalogue searches only.

An example of the desire for a fully integrated catalogue was The Guardian newspaper: many libraries would already have records for the newspaper itself, collections of essays from the paper (The bedside Guardian), books about the newspaper, illustrations of typography and page layout taken from the newspaper, and the publishers annual report and accounts. To this could soon be added records for the newspaper on-line, an abstracting/indexing service for the paper, a directory of journalists and other staff employed on the paper, etc. Any library user seeking information about the Guardian would expect to retrieve catalogue records for all of these resources in a single search.

There was discussion on what changes might be required to the MARC standard; the USMARC initiative on the 856 tag was mentioned, together with the OCLC research report on the use of MARC for cataloguing networked resources. There was a general discussion on the role of the librarian in the electronic information environment. Librarians could provide information by storing and providing access at a local level (catalogue only what you own), by providing specific, detailed access only at a remote level (catalogue what you think the user will be interested in), by providing general, non-specific access (just the access channel, with little navigation), or a mixture of all modes.

There was potential confusion between communication and information; opening a communication channel was not the same thing as using it to transmit information. The workshop agreed that the librarians skills encompassed both; information requires expert organisation, but communication factors had to be taken into account in designing the user interface.

Alternatives to CATRIONA, or any system using traditional cataloguing methods, were discussed. It was pointed out that traditional computer solutions, such as menus and icon groups, worked well when only a few tens of alternatives were to be organised, but these methods failed when thousands or millions of separate items had to be categorised. Brute force methods of keyword indexing would not be sufficient for a user to identify a particular resource because of the vagaries of language and the perversity of information creators. Controlled language approaches, such as authority files and subject indexing, were necessary to improve precision.

It was agreed that a major problem in cataloguing remote EIOs would be a change of home location; catalogue records would then point to the wrong address and the resource would not be retrievable. Cataloguers would not have the time to constantly check the currency of Universal Resource Locators (URLs) or other addresses. An automated process could identify invalid URLs, but would not be able to determine the correct URL. It was suggested that one solution might be for EIOs to be catalogued in the same locality as they were created; if the creator changed the URL, they would notify a colleague cataloguer to update to record. Other libraries would see this catalogue record in their own OPACs, if they chose to do so, although it would actually be stored remotely, and brought to the local OPAC by a Z39.50 or similar search. The alternative approach of copying the record to the local catalogue would not resolve the original problem of a URL changed elsewhere. Such methods would require a far greater level of organisation of the Internet, including rules for network publishing, resource cataloguing, and location co-ordination than was currently the case.

Librarians would not be able to achieve this on their own; practical co-operation between libraries is much talked-about and little acted-upon. The workshop thought that the best hope was for one or more central agencies to organise such work: advising creators on CIP for EIOs, identifying and cataloguing important resources of general interest, and making such catalogue records available to libraries. OCLC was already moving in this direction. Librarians would still have to do it for themselves, a la CATRIONA, for specific resources of narrow interest.

The workshop also identified a need for better training in information exploitation by providers, mediators and users. And a final observation: only one library represented at the workshop had actually attempted to catalogue networked resources, and that was confined to resources such as CD-ROMs mounted on the local network. While the proportion of such resources to total stock remained low, reasonable navigation and access could be achieved using non-integrated menus or printed documentation, but as their numbers increased, libraries would have to tackle these problems head-on.

[Return to Contents Page](#)