NetLibra: Toward Networking Library Systems

Tomohiro Kokogawa, Mitsuhiro Yuitou, and Masakazu Nishino
NTT-West Research and Development Center
t.kokogawa@rdc.west.ntt.co.jp

Yoshiaki Seki
NTT-East Research and Development Center

Abstract

The combination of computers and Internet technologies has accelerated research on an environment for searching, sharing, and providing multimedia contents, such as the digital library system. The scalability of the system and its co-operation with other digital libraries will be key issues in the near future. This paper propose an advanced digital library architecture called Networking Library, and describes a prototype system "NetLibra", that can integrate several digital libraries using distributed object technologies and networks. NetLibra provides a common distributed platform based on networking and contents distribution middleware. The NetLibra prototype system includes copyright management services.

Keywords

Digital libraries, CORBA, TINA, Copyright management, SHD

1. Introduction

The growth of Internet technologies and the rapid spread of network infrastructures are stimulating interest in information distribution environments, such as the World Wide Web (WWW). Now millions of people all over the world search, share, and provide multimedia contents via public networks every day. There is a huge amount of multimedia contents available. More effective contents distribution is one of the most important issues in Internet and Intranet services. The Digital Library (DL) is an effective technology for the distribution of multimedia contents.

There have been many DL research projects in the last few years[1]. In the US, the DL is treated as one of the most important parts of the Global Information Infrastructure (GII). Digital Library Initiatives (DLI) projects funded by the National Science Foundation (NSF), DARPA and NASA have been conducted by six universities starting in 1996 [2]. As a DLI project, Stanford University discussed and built the prototype DL testbed Infobus [3]. It was constructed on distributed object-oriented platform based on CORBA (Common Object Request Broker Architecture) [4] to study the interoperability of DL systems. The Japan Information Processing Development Center (JIPDEC) is also conducting next generation DL projects and has developed a new DL architecture [5]. They also discussed the CORBA-based DL architecture in a working group. Such architectures are being widely discussed to build a DL system that has sufficient scalability and interoperability.

This paper presents the concept of a next-generation DL based on networking technology, the Networking Library (NL) [6]. We discuss the DL and NL requirements, and propose a prototype of NL called "NetLibra" [7]. A NetLibra prototype system has been built and its performance evaluated.

2. Requirements and approach

2.1 Requirements for DL

System requirements for DL have been widely discussed by librarians, DL researchers, contents providers, and users. There are four main requirements.

(1) Scalability

Larger real libraries such as university libraries and public libraries in big cities may contain million of books. The DL system has to manage such contents and bibliographies.

In addition, the DL system should support all digital contents in the following formats as examples;

Documents: plain text, PDF, SGML, HTML,
and particular formats of word processors

(such as Microsoft Office)
Images: GIF, JPEG, and TIFF
Sounds: SMF, AIFF, and MP3 (MPEG layer 3)
Movies: AVI, MPEG, and QuickTime

(2) Interoperability and co-operation

The DL system must offer not only a contents retrieval service, but also a variety of other services to support librarians or DL administrators. These services include information sharing, Q&A, user management, and so on. These services should co-operate with each other.

There are some existing library services for networks such as the Online Public Access Catalog (OPAC) services based on the WWW or Z39.50 [8]. The new DL service should fully co-operate with such existing services, or should offer an effective conversion function.

(3) Copyright protection

Digital contents can be copied and be distributed much more easily than books. Copyright protection is one of the most important problems in contents distribution services, especially in DL services.

There are two viewpoints to copyright protection. One viewpoint is the prevention and detection of illegal use. The DL system should offer an adequate level of security using tools such as user authentication, access control, encryption of contents, and digital watermarking of contents.

Another viewpoint is a copyright management. To encourage the reasonable use of contents, functions to purchase a license for access, and to pay the fees collected to the originator are needed.

(4) Usability

As a public service, The DL system should offer easy of use to many users and should give them reliable results. Easy to operate services using common client applications such as WWW browsers, quick retrieval of contents, and appropriate navigation to their favorite contents are needed.

For librarians and library administrators, DL management services should reduce their operation costs.

2.2 Approach to Networking Library

A lot of DLs exist or are being constructed. These DLs have been designed as individual elements, so they have different interfaces, different GUIs, and different services, and limited contents.

Users must access them individually in order to assure themselves of adequate coverage. Constructing a new large-scale DL is difficult because of the problems development cost, the cost and time of gathering contents, and the cost of management. Ensuring a sufficient level of copyright protection and security among DLs are also important issues for contents distribution. Some existing DLs have to limit access to hide the lack of copyright protection.

One effective approach to solve these problem is the construction of a common platform that allows the interconnection and co-operation of DLs on the public network. That is, set of DLs constructed on the platform behaves just as one large-scale DL. We call this proposed concept of DL the Networking Library (NL). NL means not only a DL system based on distributed databases, but also a set of DL systems with different backgrounds (venders, services offered, locations, purposes, and so on).

The common platform for NL needs two typical functions.

(1) Networking functions

To get acceptable NL performance, the networking functions used to connect the DLs are very important. Both existing and new DLs will be connected to the NL, so these functions should be used commonly to unify the interface and to hide the networking complexities, such as network protocols, from the service applications.

(2) Contents distribution functions

Several service functions such as fulltext retrieval are essential in many DL services. In addition, copyright management is also important. Such kind of functions should be used in common to reduce the cost of constructing new services and to ensure the usability and security of the whole NL system.

3. NetLibra architecture

3.1 Software structure

Based on our concept of NL, we proposed the prototype system of NL: NetLibra. NetLibra is a common platform for multiple DLs, and has two middleware layers: networking layer and contents distribution layer. Figure 1 shows the architecture of NetLibra.
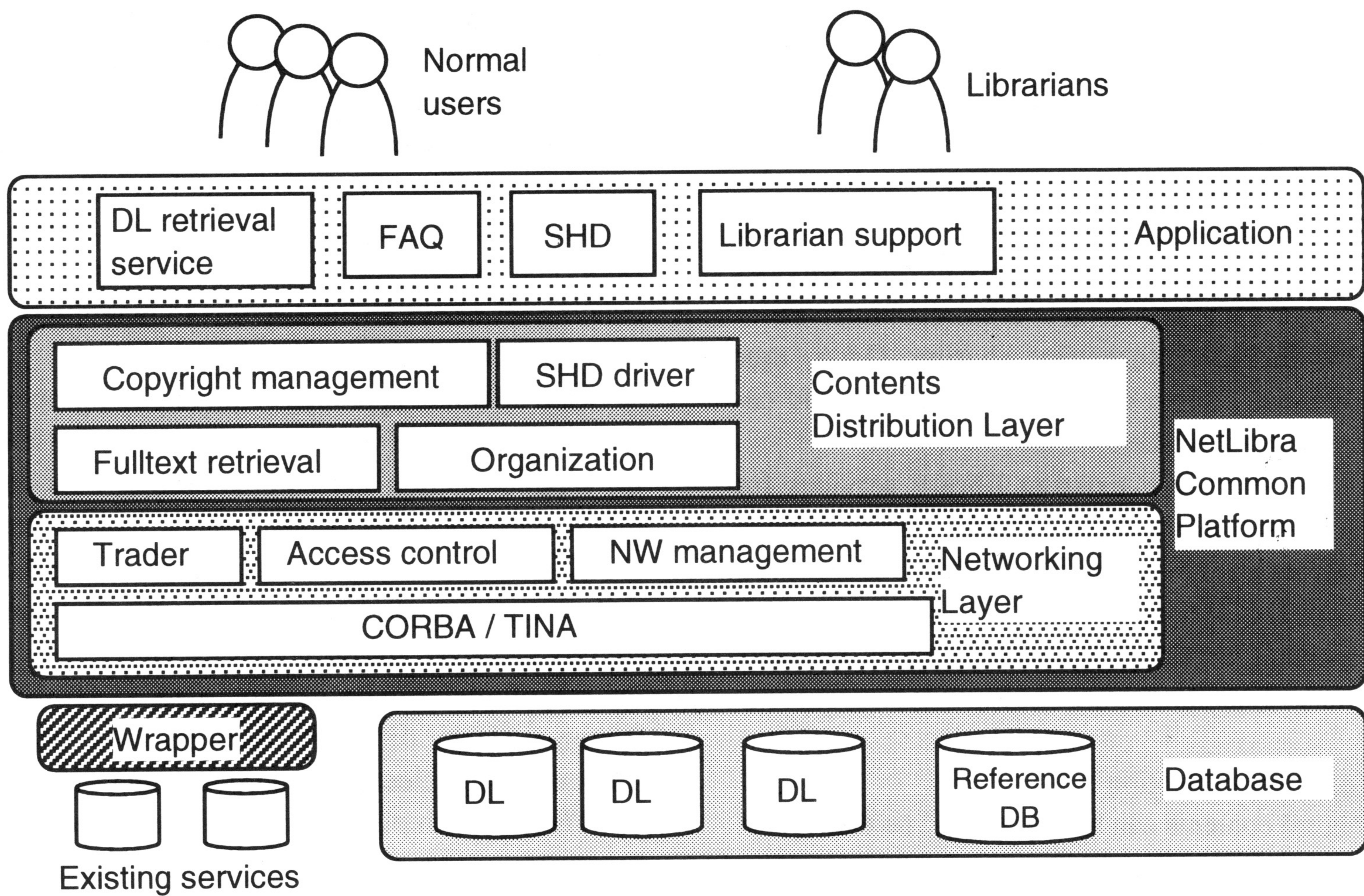
3.2 Networking layer

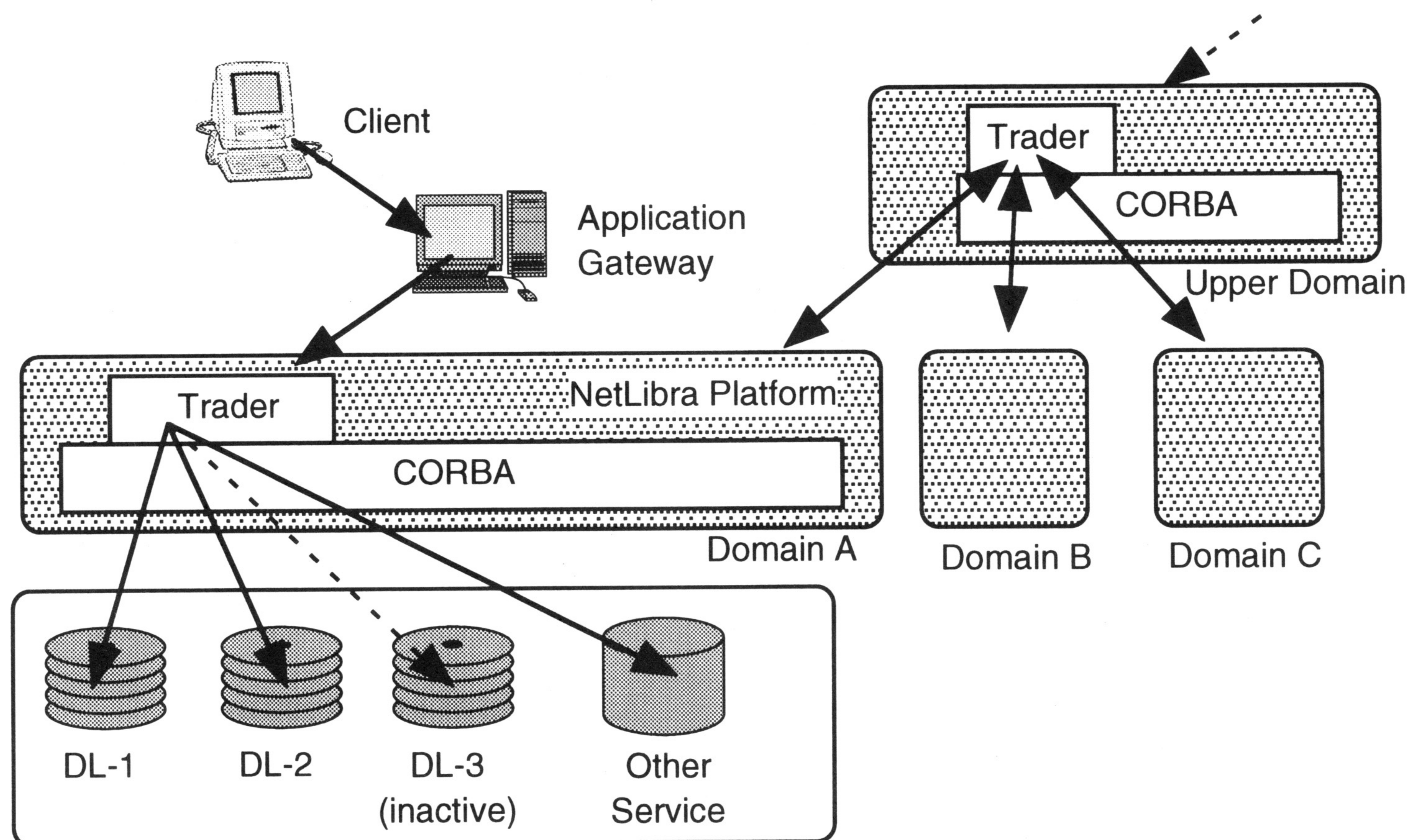Figure 1. NetLibra architecture.

Figure 2. Multiple DL access using trader.

This layer offers a function to control networks and distributed systems, such as network management and access control. It hides the networking complexities from the upper layer. This layer is based on a distributed object-oriented architecture such as CORBA or TINA (Telecommunication Information Networking Architecture) [9]. CORBA is becoming a popular object-oriented architecture for distributed systems. TINA is similar to CORBA and was mainly developed for network carriers. TINA is a more useful framework with regard to networking

functions such as connection management and session management, and so is useful for distributing large volumes of multimedia contents across networks effectively.

NetLibra offers multiple DL access services that are based on the CORBA trading service. When a DL server (database) is connected to the NetLibra platform, the service information of the DL server is registered as an object reference to the the trader, a CORBA function to offer trading service is created. When a user requests a service, the NetLibra application gateway receives the request and accesses to the trader to get the list of registered active services. The user can access any active service, without considering which service is registered, where is the service on the network, and whether the service is active or inactive. The trader can also manage the service methods, service names, several service parameters, and result data types. These parameters are useful for the detailed classification of services.

Figure 2 shows an example of the dynamic multiple DL retrieval service based on the use of the trader. In this case, the user can access all active DL services on NetLibra at the same time, and can treat them just as one DL service by merging their retrieval results.

This feature also benefits library administrators. They may easily add, manage, and remove any DL service individually, without considering the effect on other services available on NetLibra.

In the NetLibra system, one trader may be used for each domain. To connect NetLibra services on other domains, traders on each domain may be linked via a trader on an upper domain (federation trader).

3.3 Contents distribution layer

This layer includes contents distribution services that are used by many applications, such as full text search, contents linking, and copyright management.

For NetLibra, we developed a copyright management service. As typical copyright management functions, we integrated the InfoProtect [10] technology into NetLibra. InfoProtect has two basic functions. One is image distribution using partial-disclosure encoding. After the user registers an image to get its license and to pays the copyright fee via a purchase application, InfoProtect sends the key needed to recover the full image. Another function is digital watermarking to prevent illegal use of the image. When the key is used to decode the image, InfoProtect also embeds the buyer's ID as a digital watermark into the image.

A DL must handle a variety of multimedia contents. SHD (Super High Definition) [11] is one of the most useful image distribution services to
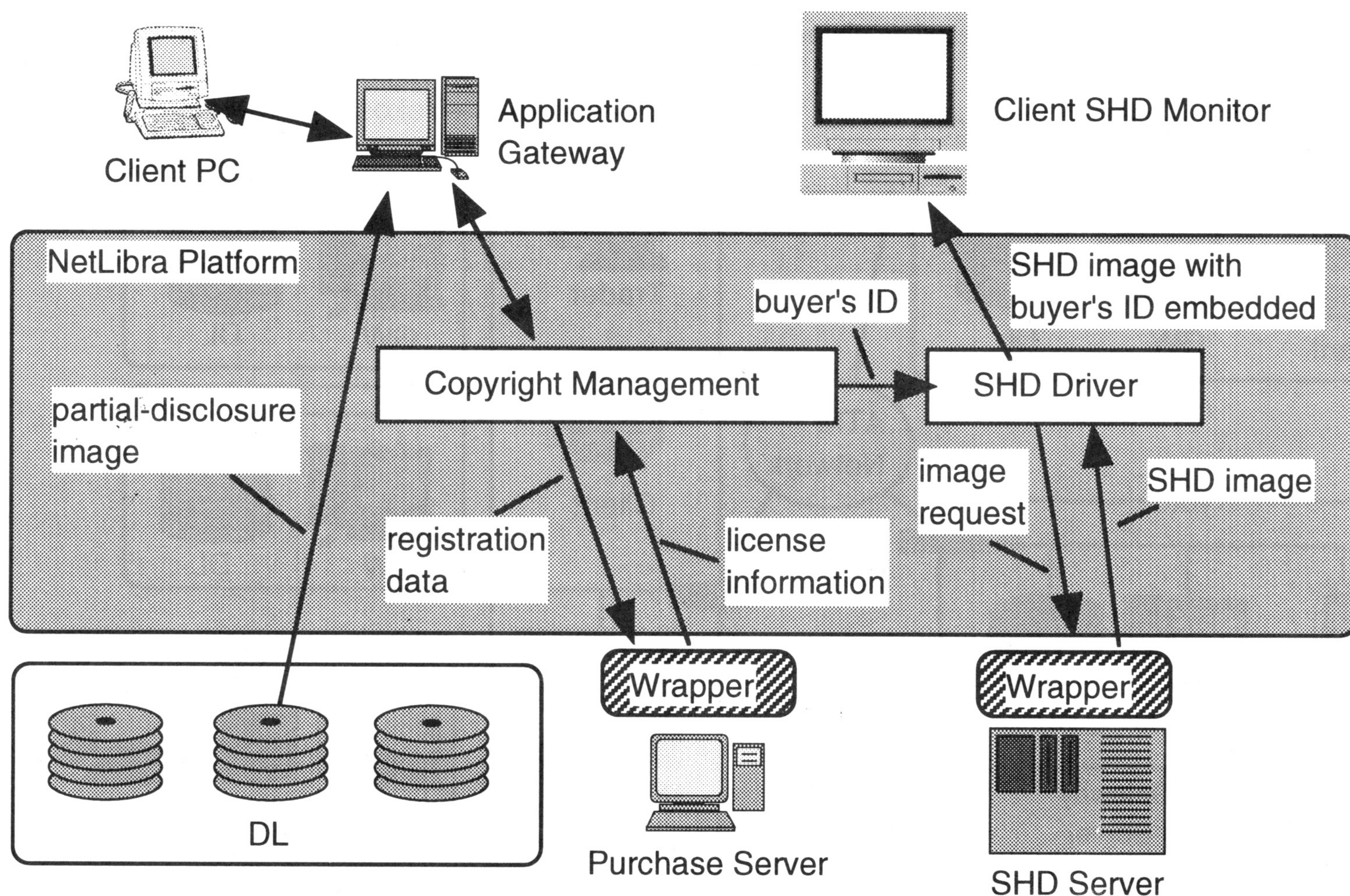

Figure 3. SHD service sequence on NetLibra including copyright management.

provide high-quality images or movies via wide-band networks. In this case, copyright protection may be extremely important because of the quality and value of SHD images. Figure 3 shows an example how copyright management on NetLibra would handle SHD images. At first the user previews low-quality images in his client PC (partial-disclosure encoding). After registering the payment of copyright fee, he gets the real high-quality image on his SHD monitor into which is embedded a digital watermarks. Both middleware of copyright management and SHD function are contained in the layer, so the co-operation service is effectively integrated. This is an example of the value-added service model that co-operate with copyright management.

3.4 Other features

For NetLibra we also prepared other techniques for making real NL services.

(1) Metadata design

There are many schemes for cataloging DL contents such as USMARC. The Dublin Core facility, which points a metadata architecture with fifteen basic elements [12], is becoming well accepted in DL. The metadata of NetLibra are based on Dublin Core metadata elements to manage DL contents efficiently and to reduce the cost of converting other DL services' metadata or cataloging schemes. NetLibra has over forty metadata elements

including copyright management.

(2) Librarian support service

The networking architecture of DL offers benefits not only to users but also librarians and library administrators. Librarians have their own know how about reference services. In many case, their know how should be shared with other librarians as FAQ (Frequency Asked Questions). NetLibra offers these services as librarian support in addition to library management services such as user management, contents input and so on.

(3) Wrapper for existing services

Some services such as old style DL services will continue to exist. To connect these services via NetLibra, small wrapper middleware modules that offer the NetLibra style interface have been prepared.

4. Prototype evaluation

4.1 Prototype system

Figure 4 shows the NetLibra prototype system configuration. This system connects three locations (Yokosuka, Musashino, and Kyoto) via an ATM network. There are five DL servers, with three of them in the Yokosuka location. Each DL server offers a multimedia database that is built on NetLibra platform based on CORBA. Each has their

GW: Application Gateway
DL: Digital Library Server
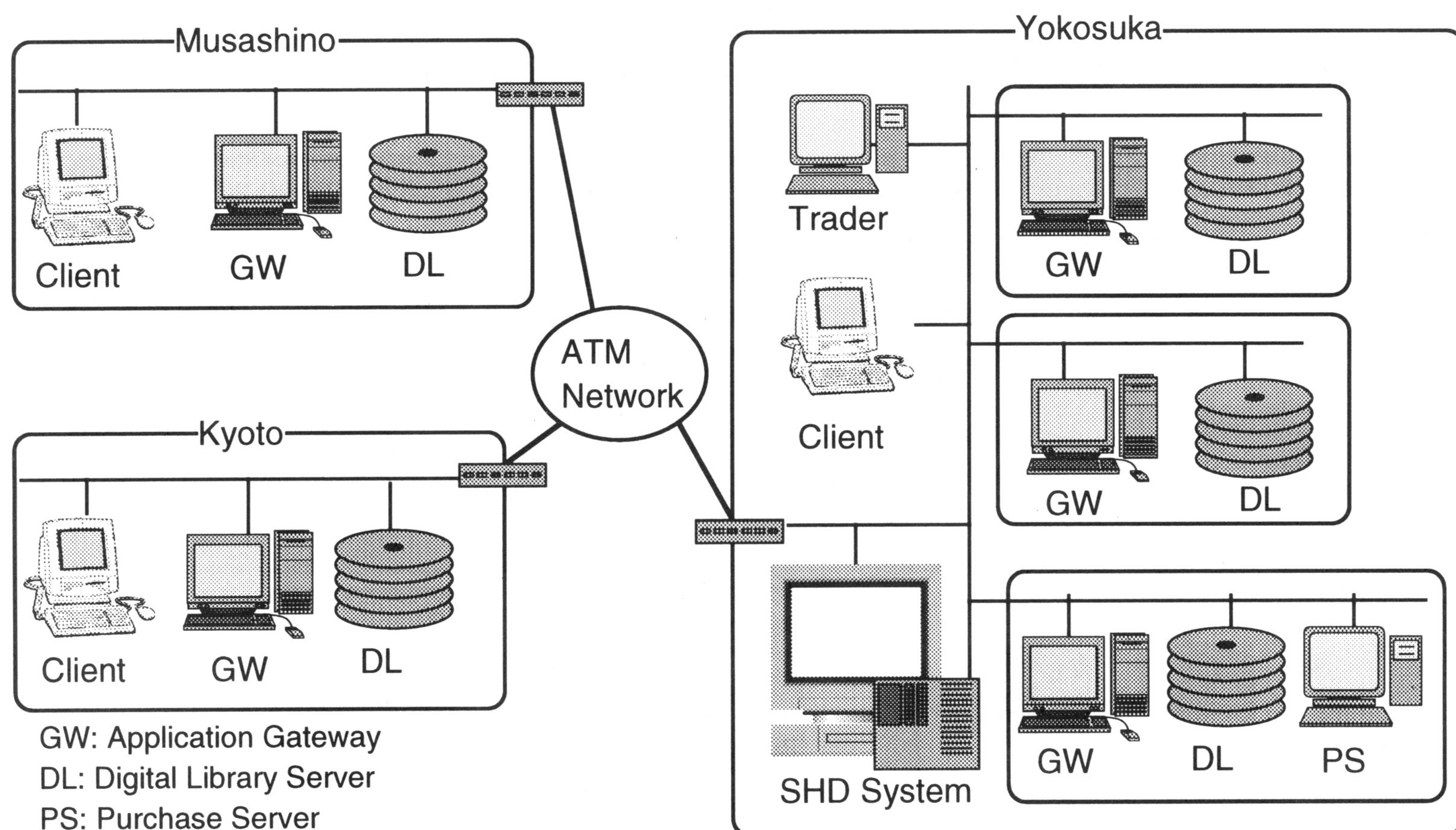PS: Purchase Server

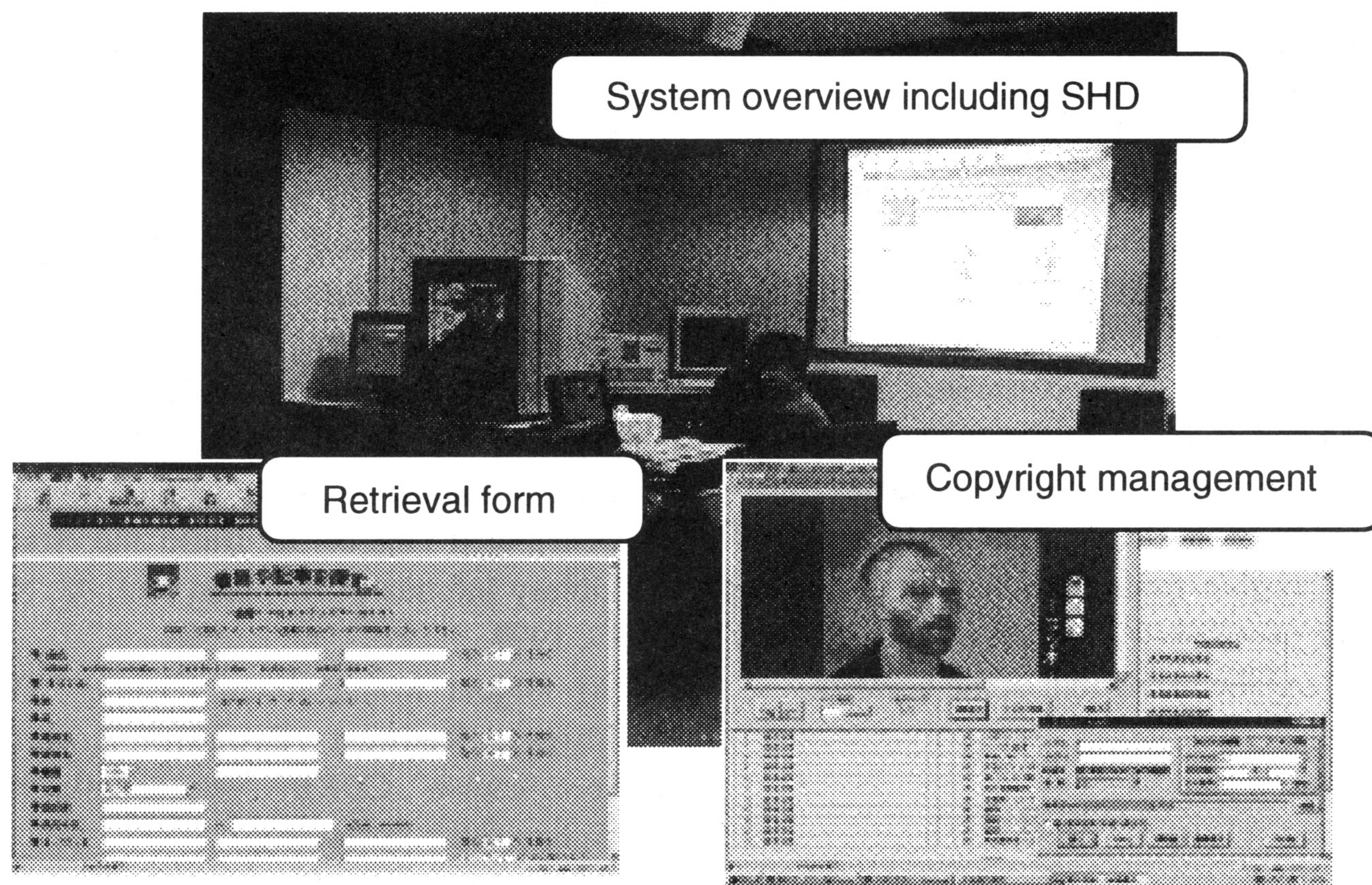Figure 4. NetLibra experimental system configuration.
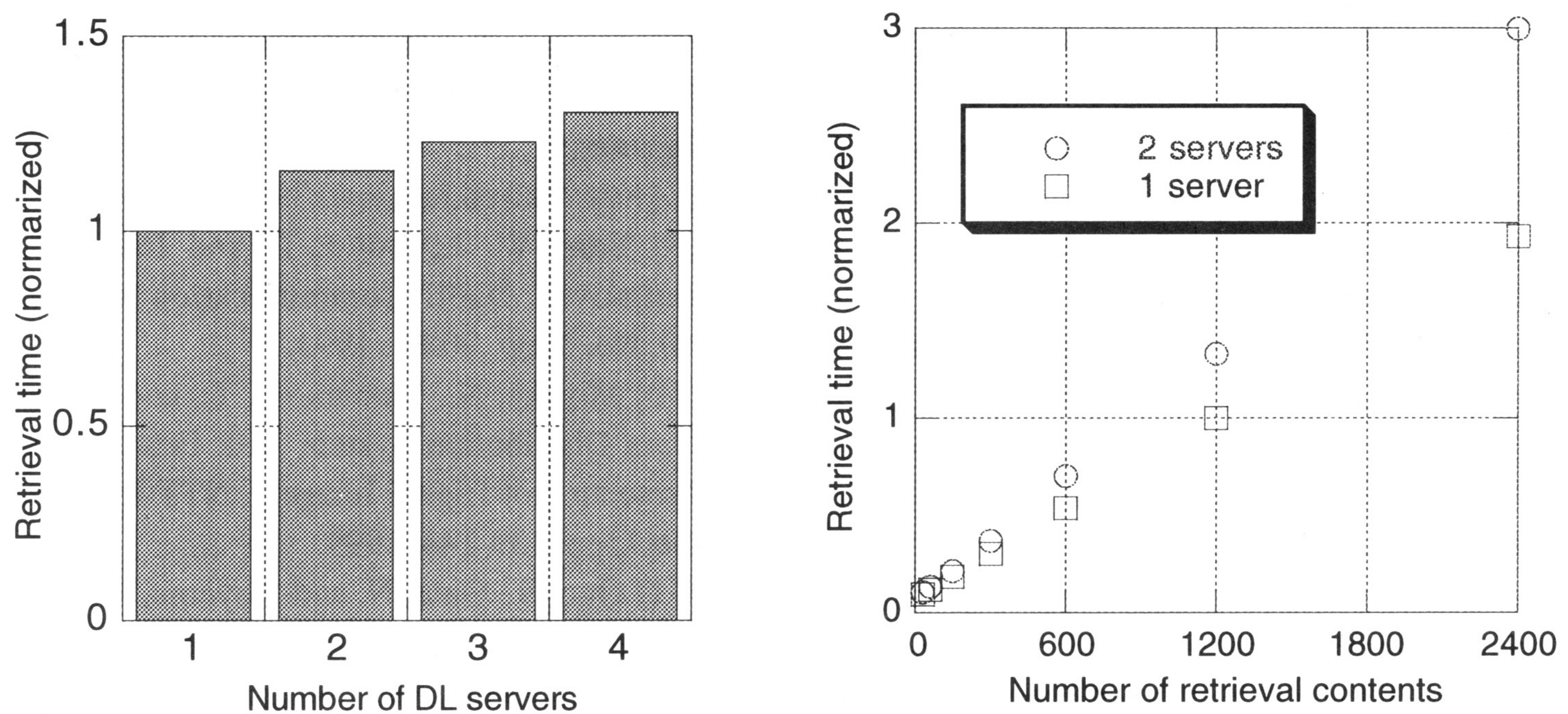
Figure 5. NetLibra overview.

Figure 6. Retrieval Results.

own application gateway, but any DL server in the system can be accessed via any gateway by using NetLibra. The trader is located on an independent workstation in Yokosuka. Other servers, such as librarian support, FAQ service, and user management service were also built on the platform. Several functions such as a partial implementation of the copyright management server and an SHD system are connected to the platform via wrappers as to offer the NetLibra style interface. Applications such as DL retrieval are built as CGI (Common Gateway Interface) or Java Applet components to allow operation via typical WWW browsers. Figure 5 shows an example of service.

4.2 Performance evaluation

To evaluate the effectiveness of the NetLibra architecture, we examined the performance of typical DL services. Figure 6 shows the performance of multiple DL retrieval from one of the gateways in Yokosuka. The left side of Figure 6 shows the change in retrieval time versus the number of DL services. The deterioration of the performance as the

number of DLs is increased is not so serious. The main problem is the performance of the merging process in the gateway, which merges the retrieval results of each DLs into one form.

The right side of Figure 6 also shows the performance of DL retrieval. The horizontal axis shows the number of retrieval contents selected by one retrieval operation. The retrieval time increased almost linearly with the increase in retrieval contents. In this case, the performance of the DL databases and that of the CGI process in the gateway are the main factors determining performance.

These results show the practical performance of NetLibra, but further tuning is needed considering the overhead of CORBA, database and gateway performance, and the bandwidth of the network.

5. Summary

We proposed the concept of an advanced digital library called Networking Library, and developed the prototype system NetLibra. NetLibra architecture is based on two middleware layers: networking layer and contents distribution layer. Networking layer hides the networking complexities from upper layer and unificates each DL as one large-scale DL. Contents distribution layer prepares informative function such as copyright management to make services effectively. This architecture provides a common platform not only for digital library services but also any kind of multimedia contents distribution service such as digital museums. Such services provide users informative and reliable ways to access large kind of contents on networks.

We made a NetLibra prototype system that connected three locations and evaluated its performance. Although it needs more tuning before commercial use, evaluation results show its potential in a multiple DL environment. The prototype system also implemented copyright management and SHD services as middleware, important function in the next-generation digital library systems.

As future work, we will develop and tune up the system design for practical use. Additional networking functions to optimize the use of networks in NetLibra, such as QoS (Quality of Service) control, connection management, and session management, will be integrated using TINA as a distributed object-oriented platform. These functions will be useful to distribute a variety of contents such as large-scale image (such as SHD image) and streaming video on networks effectively.

We will also evaluate the usefulness of other middleware functions such as copyright management, and integrate them as needed.

References

[1] Schatz, B. and Chen, H.: "Digital Libraries: Technological Advances and Social Impacts", Computer, Vol. 32, No. 2, pp. 45-50, 1999.
[2] http://dli.grainger.uiuc.edu/national.htm
[3] Paepcke, A., Baldonado, M.Q.W., Chang, C-C.K., Cousins, S., and Garcia-Molina, H.: "Using Distributed Objects to Build the Stanford Digital Library Infobus", Computer, Vol. 32, No. 2, pp. 80-87, 1999.
[4] Object Management Group: "The Common Object Request Broker; Architecture and Specification", CORBA Version 2.0, 1995.
[5] http://dlib.jipdec.or.jp/
[6] Yoshikai, N., Kayano, T., Wakahara, T. and Nishino, M.: "Networking Library over Distributed Object-oriented Environments", IPSJ Tech. Report, DPS-85-38, pp. 219-224, 1997 (in Japanese).
[7] Kayano, T., Nishino, M., Seki, Y., and Kokogawa, T.: "Distributed Networking Library: NetLibra", 57th IPSJ Conference, 3-1, 1998 (in Japanese).
[8] http://lcweb.loc.gov/z3950/agency/
[9] Barr, W. J., Boyd, T. and Inoue Y.: "TINA Initiative", IEEE Communication Magazine, Vol. 40, No. 3, pp. 70-81, 1993.
[10] Miyake, N., Yamamoto, T., and Tamai, M.: "A Data Distribution System with Featuring Copyright Protection of Digital Images", NTT Review, Vol. 11, No.1, pp. 139-144, 1999.
[11] Ono, S., Ohta, N. and Aoyama, T.: "Super High Definition Images Beyond HDTV", Artech House Publisher, 1995.
[12] http://purl.org/DC/index.htm