Chapter: 8
IIM Unified Intranet Portal System (IIM-UIPS): A Design

8.1 Introduction
8.2 Scope of the system
8.3 Functional Requirement Analysis
8.4 Architecture, System Analysis and Design
8.5 Discussions
8.6 Conclusion
8.7 References
8.1 INTRODUCTION

IIM Unified Intranet Portal System (IIM-UIPS) exploits both the technologies of intranet and portal to provide a unified interface to heterogeneous resources available both remote and local in a distributed environment among IIMs. The design of IIM-UIPS is based on the information collected through the questionnaire that includes:

- Information Requirements of users (end users as well as staff) of IIMs both local and remote
- Information Resources available
- Tools and techniques to make these resources available and accessible
- The information gaps between the users and resources

IIM-UIPS model provides a framework that would facilitate acquiring, organizing, sharing electronic resources among IIMs. The primary purpose of IIM-UIPS is to establish a portal framework to allow users to have seamless access to range of print as well as digital information resources which are “pushed” to them according to their teaching, learning and research needs.

8.1.1 CURRENT EFFORTS BY IIMS IN ELECTRONIC RESOURCES MANAGEMENT

All the IIMs use same library management software called ‘Libsys’. Libsys is integrated library management software to support automation of main functions of a library including acquisition, cataloging, circulation, and serial control. It provides a web based online catalog, union catalog and ILL facilities. Using Libsys, all IIMs have their OPACs web enabled. Each IIM has its own library web portal through which the electronic resources are made available for access and use. But, these portals are basically providing browsing of resources (hierarchy) and search a particular resource through its
native search interface. IIM-K has developed its library portal based on the Greenstone software to provide access to some of its digital resource collections.

8.1.2 CONSORTIUM APPROACH

The consortium approach adopted by IIM-UIPS is based both on centralized and decentralized models. In addition to collective acquisition of resources at the consortium level, the individual IIMs can identify and evaluate new resources based on their local needs. The IIM-UIPS will facilitate broadcasting these resources to make other participating IIMs aware and provide mechanisms to share across are license to cross share is obtained. In otherwords, acquisition and organization of resources are based on both centralized and decentralized models while utilization is based on the decentralized model. This means that the autonomy of an individual IIM in terms of information acquisition and organization is not diluted while providing a mechanism to share and access. However, the model needs a consensus on adopting a standard based approach among the participating IIMs. The figure (Fig. 8.1) below explains the consortium environment for resource sharing among IIMs (Pandian, etal, 2002).
8.2 SCOPE OF THE IIM-UIPS

The IIM-UIPS is defined as a single, easy to use, integrated, and coherent web enabled system, which provides

- A single point of access to all the resources across IIMs
- Unified login (including one user ID)
- One common user interface, i.e. one presentation structure
- One uniform user-friendly retrieval system
- Direct access to electronic media and unified request service
- Patron-initiated online requests of resources and interlibrary loan facilities
It will support various user types:

- Academic (Faculty, students, research staff, visitors)
- Non academic
- Library and System Staff

It will support various type of Resources:

- OPAC (IIMs +)
- A & I databases (Econlit, Socio file, ...)
- Fulltext databases (Business Source Premier, ABI/INFORM, ...)
- Numerical and statistical databases (CMIE, ERS, Indiastat, ISI emerging markets, ...)
- E-journals (Science Direct, Blackwell, Kluwer, Wiley, ...)
- E-books (netlibrary, ebrary, ...)
- Case studies
- Working papers
- Other in-house publications

Type of protocols/standards that the system will support:

- Z39.50
- XML
- HTTP
- OAI-PMH
- OpenURL
- LDAP
- Shibboleth
- NCIP/SIP2

It will support various locations of material:
Integrated access to all forms of resources available has become a necessity. Such access maximizes faculty, staff, and student productivity while assisting them in becoming knowledgeable, and critical users of the wide array of digital information.

The IIM-UIPS on each campus will present the users with a series of screens designed to guide in framing a search and in selecting appropriate information resources. It will allow searching of multiple resources with a single command, and will offer links to additional interactive help when needed. Search results will be integrated and displayed in a useful and coherent manner. The system will assign each user a profile to permit librarians to customize the IIM interface and information sources most appropriate for the individual user. The system will also authenticate the user statuses as legitimate borrower, and will permit direct requesting of books and articles from another IIM library without going through the interlibrary loan department. IIM-UIPS system broadly consists of the following components (Figure 8.2):

![Figure 8.2: IIM-UIPS System components](image-url)
IIM-UIPS is based on a combination of both centralized and distributed service models in order to meet the diverse needs of users and heterogeneous nature of resources.

8.3 FUNCTIONAL REQUIREMENT ANALYSIS OF IIM-UIPS

In order to design a framework, it is necessary to understand the functional requirements of the proposed IIM-UIPS. Following are the requirements of end users, library staff and the systems.

8.3.1 END USERS REQUIREMENTS

An end user from any of the participating libraries of IIMs enters the system (mainly through web browser) and identifies himself and after successful login he starts searching by selecting one or more resources. He may submit a query to multiple resources of which some are local, some are from other IIMs and some are from third parties. These resources include OPACs, ebooks, ejournals, A&I databases, full text databases, statistical and other numerical databases, case studies, etc. The results are returned to the user as a list of bibliographic records (descriptive metadata).

He browses the records to which full text documents may be attached (either copyright material or non copyright material, either locally mounted or remotely, either own holdings or holdings from third parties, including internet resources that were described in the databases).

From the bibliographic record that contains such a link the end user simply clicks the show document button and the document is fetched from its location. The latter is transparent to the end user as the system takes care of the tracing of the exact location of the document.
He is able to print the desired document. He can request for an item that is not locally available but with other IIM libraries. He is able to save his search history for the session. He may also be able to create his own profile that can be matched with profiles of resources available from time to time and deliver the results in his mailbox if desired so.

He is also provided with suggest a new item to be acquired for the library. He can view his transactions with regard to acquisition requests, check in and check out of items, ILL requests, search history and saved sessions.

End users may virtually be at any place (library, office, at home) without risking exclusion from the IIM Intranet Unified system as long as they can (1) gain access to the system in one way or the other and (2) present a valid username and password combination at times of login. This makes it a truly universal service.

8.3.2 LIBRARY STAFF REQUIREMENTS

Following are the library staff requirements in managing the IIM-UIPS system:

When new content is acquired the library system management staff needs to add the material to the system. This material may have been obtained from publishers, from subscription agents (copyright material) as well as from faculties of the own institute (noncopyright material). Sometimes librarians even add descriptions of Internet resources to the local databases including the URL where the document can be obtained.

Each time that new content is acquired and which differs in structure from a previous one a new content loader must be built. The system management staff's main task is to maintain the link between the bibliographic description and the full text document and to store various types of material in the appropriate databases/systems. Another task of library system management staff is to control the access of (groups of) users. These staff shall have a facility with which they can maintain access regimes for the local end users
that are affiliated to the local library/institute. But they even can maintain access regimes necessary to control the access by third parties with which agreements were made.

Furthermore library management would like to receive periodically an overview of the use of the system in quantitative terms. The data that is logged in the system is regularly transferred to the logdata loading system. This subsystem processes all the logdata that was produced and stores the desired data in a database. This database is the source for the library manager to produce the overviews he wishes to make and uses the software he is familiar with for doing that.

8.3.3 SYSTEM REQUIREMENTS

In order to meet requirements of both the end users and library staff, the system will need to be designed as follows:

General:

- The system should be web based
- The system should allow a single sing-on
- The system should present cohesive view of all the collections across IIMs
- The system should provide the users both browsing and searching
- The system should allow searching both on native interfaces of the resources and through a common search option
- The system should support Current Awareness services, mylibrary – user profile and ILL
- The system should support generating various statistical report

Querying

- a facility to display searchable fields (e.g. pull down menu)
a facility to query a metadatabase to find information about available resources
a facility to select one, more or the complete list of resources at once
a facility to search multiple resources at once
a facility to use controlled vocabulary during search

Query Results

a facility to skip multiple pages at once in a large result set
a facility to highlight the search terms in the record data (given query and search fields)
a facility to use parts of the record data as a search term in a new query (keywords, author fields, journal titles/codes)
a facility to reuse/refine (augment) record result sets (throughout session)
a facility to (un)mark records (throughout session)
a facility to save, print or email a marked set of records collected throughout session

Broadening and narrowing down search terms

a facility that suggests the end user what databases/sources to use best
a facility that discourages the selection of particular sources for particular queries
a facility that refines or does suggestions to refine queries
a facility that warns the user of a possible information overload before a query is submitted (possibly by learning from previous submitted queries)

Subject Match

End users may need to query various databases. These databases may not contain the same terms with the same semantic meaning. As a demonstrator the system shall contain a facility that aims at subject matching of search terms across the resources.
Result Optimization

- a facility to normalize the search results into a common display through the browser
- a facility to identify duplicate records found
- a facility that does rank the more appropriate records higher in the hit list. Ranking is pursued by the system rather than by the database where the records come from.
- a facility to sort the result sets obtained from multiple resources.

Access control

- There should be only *one* login (or user identification / authentication) process per session (per user) - even when accessing an external source/holding of a deliverable document therefore the IIM-UIPS system must supply or negotiate access (dependent of course on the rights of the current user) to such external deliverable document sources - transparently to the user.

8.4 ARCHITECTURE, SYSTEM ANALYSIS AND DESIGN

8.4.1 IIM-UIPS ARCHITECTURE

There are two main characters to determine the architecture of an information retrieval system: the size and diversity of data source. To support the above-mentioned requirements, a system for accessing heterogeneous information sources needs mechanisms for hiding the differences between sources, for identifying sources likely to contain relevant information, and for combining results. System scalability, extensibility and customizability are also important because the system needs to adapt as the environment or the user's needs change. Scalability is essential for handling the exponentially growing number of available information sources, as is the case on the WWW. No single WWW crawler indexes the entire WWW and no single corporate
database or search engine provides one-stop access to global information. Extensibility is critical because new information sources and new interfaces, protocols, and formats emerge constantly.

Users and system administrators should be able to extend the system with minimum manual effort. Ideally, the system should be able to discover new information sources automatically and know how to talk to them. A system should also be able to adapt to a changing environment because information sources, especially external ones, may change unexpectedly. Internally, an unified access system therefore needs to employ a data model and query language that are rich enough to subsume the data and query representation capabilities of other systems, and it should provide a flexible abstraction mechanism which hides the interface details of an information source. Customizability is important because it is unrealistic to expect a single generic system to be able to handle all application domains and information seeking tasks as competently as a tailored system. Therefore, an architecture in which individual components can be customized to a particular application domain is desirable. A customized results integration component, for instance, can embed knowledge about domain-specific vocabularies, transformation rules, or similarity measures to improve retrieval effectiveness.

Because of the very nature of the diverse resources distributed across and diverse user needs in IIM-UIPS environment, an intelligent system to these resources is needed to provide the services and interfaces required by readers. In general terms, there are three basic models, which could support accessing these resources and associated scholarly resources (Swan, et al., 2004):

- Centralized - both metadata and the resources themselves are kept on a centralized system. The advantages of this model are that the agency running the service would have overall administration of the whole process and be able to standardize the protocols used, select the software that provided the most appropriate set of storage and output capabilities and to develop facilities that maximized search capabilities (categorization of the data, subject
classification, etc.). However, with all administrative and maintenance functions centralized, it is an expensive option.

- Distributed—all metadata and resources remain in their source locations, and metadata are cross-searched on the fly. In this model proposed services would obtain metadata in real time, as the user asked for it, and point the user at the digital resource, which would be located in a distributed archive. The service would cross-search all available archives, using the Z39.50 protocol, or SRW/SRU, and present the results to the user.

- Harvesting—a hybrid model—metadata are harvested into a central searchable database but also remains distributed among the original data providers, while the resources themselves remain distributed.

In otherwords, a more realistic approach points to only two possible types of approaches for consideration:

Federation or Harvesting: The federated search, or metasearch, is a service that provides unified query interfaces to multiple search engines. It requires each search engine to implement a joint distributed search protocol; moreover, as it needs post-process search results in real time, it presents significant scalability problems. In this environment, the data is located on multiple resources and is searched by clients using standard protocols such as Z39.50, ZING/SRW/SRU, OpenURL or XML Query. There is no centralization of the data. A central search engine sends a user's search request to every participating online database, having first translated it so that it can be understood by each database. The remote databases carry out their searches locally and return the results to the central search engine. The central engine waits for all the results to come in then it checks the quality of the incoming responses, groups apparent duplicate hits and then sorts and displays the results.
A harvesting approach collects data from heterogeneous sources in advance; therefore, it is more realistic in dealing with large number of digital libraries. Harvesting approaches have the additional attractive property that they allow data enhancing procedures to be run on the collected data. Enhancements such as normalization, augmentation and restructuring are applied to data originating from different sources in order to create consistent end-user services. In a harvesting scenario, these activities can be dealt with in a batch. Protocol such as OAI-PMH supports this approach.

The difference between Z39.50 and OAI, according to Pederson, is that OAI's uses a metadata-harvesting approach, whereas the protocol Z39.50 is a search-and-retrieval protocol. Both Z39.50 and OAI provide "federated searching", which enables users "to gather information from multiple related resources through a single interface". That way one search query can present search results from many resources. This mechanism raises the completeness of the available information so that the user does not have to search several resources separately. Z39.50 uses an "online, real-time connection between the searcher's system and one or more targets using a thick and complex set of communications protocols". A query is transmitted to many Z39.50 targets within sessions, and the results from each query are collected, analyzed and presented to the user. "The advantages of Z39.50 are its ability to search remote resources through a common user interface and the immediacy with which it accesses current information in real time". The mechanism of establishing online connections with remote servers makes Z39.50 a subject to the instability. This is a disadvantage of Z39.50, whereas the OAI does not need to establish and maintain such interactive connections among all the original repositories. The OAI creates "pre-built metadata collections" and collects all the metadata it aggregates from the data providers in advance (Pederson, G. 2000).

Both the approaches have some advantages and disadvantages. On one hand, federated search method works best among a limited number of large, powerful databases in a real time environment, there are scaling problems in the management of searches that are run at large numbers of servers; one has to worry about servers that are unavailable (and with enough servers, at least one always will be unavailable), and performance tends to be
constrained by the performance of the slowest individual server participating in the federation of servers. In these circumstances the user has to wait for a lot of record transfer and post-processing before seeing a result. Federated search performance sensitive to participating server response time, result size, and network bandwidth. On the other hand, in the harvesting environment, a user is actually querying on a metadata repository, which has already been constructed from metadata harvested from the resources. Hence, there is a possibility of losing the currency of information.

However, in the IIM Consortia environment, possibilities of using both the approaches exist, in addition to custom scripting for some resources where they are no standards to support accessing these resources in order to provide a truly unified environment.

Comprehensive review (Chapter 7) of various unified models reveals that there is no single model which can be directly deployed in the IIM-UIPS environment with ease as every solution provided has been developed to meet the specific needs of the that particular organization(s) involved. Commercial solutions are customizable to suit the needs, but are prohibitively expensive that rules out an option for considering any commercial solutions. Academic and research and open source models are again specific to meet the individual organization's requirements. However, the proposed system architecture for the IIM-UIPS is influenced by combination of several such models in order to meet the requirements.

The architectures of the European Library and DecomateII are closer to meeting the requirements, which form base for the design of IIM-UIPS system architecture. The IIM-UIPS framework combines a distributed search model with the OAI model for federating access. The collections that can support search and retrieve via federating search tools (Z39.50, OpenURL, XML) can be directly integrated into our framework with real-time federated searching. It is therefore not necessary to federate access to them by harvesting their metadata into a shared index using OAI. For those collections that cannot be integrated into the framework in this way, we propose creating metadata repository of OAI-metadata. The most straightforward way to gather metadata for the OAI repository
is with OAI harvesting, provided what is being harvested is an OAI-compliant resource. However many quality resources within IIMs digital collections are not OAI-harvestable in their current state. For these collections, we propose custom scripting for capturing, extracting, mapping and importing non-OAI metadata into an OAI repository. But, the significant difference in approach is that IIM-UIPS uses an intelligent Interface system that will talk to the resource control unit to decide which resource can be searched for a given search query. The resource control unit also acts as a controlled vocabulary system where the search terms are mapped with controlled vocabulary that in turn links to the right resources. The vocabulary-controlled system essentially consists of the controlled terms and the resource identifier.

Therefore, the broad overview of the architecture for IIM-UIPS would be as given in the following figure (Fig. 8.3):
The above architecture design of IIM-UIPS was driven by the following functional requirements. 1) Identification, acquisition, organization of information resources (collection manager), 2) Providing access to these resources through the appropriate tools and techniques (federation manager and harvesting manager), 3) building up of resource control system for retrieval efficiency to enhance the precision and recall values (resource control manager), 4) single sign-on system for the users to access resources across the
to normalize, deduplicate and present the search results retrieved from various heterogeneous and distributed resources.

Figure 8.3 illustrates the core components of the IIM-UIPS architecture and the interactions among them. The architecture builds on a number of fundamental concepts developed over the past several years of digital library research. The components of IIM-UIPS are:

- User Manager
- Collection Manager
- Federation Manager
- Harvesting Manager
- Resource Control Manager
- Presentation Manager

8.4.1.1 USER MANAGER

User Manager is responsible for adding new users, modifying the attributes of the existing users, and removal of users.

8.4.1.2 COLLECTION MANAGER

Collection Manager is responsible for adding new information resources, modifying the existing resources attributes, removal of information resources. Collections are identified, evaluated and added through various processes in a consortia environment.

8.4.1.3 FEDERATION MANAGER

Federation Manager is responsible for providing federated search to the select resources based on the users search requirements facilitated by the intelligent unified interface.
system. Z39.50, OpenURL and XML protocols support these functionalities of federated manager.

8.4.1.4 HARVESTING MANAGER

Harvesting Manager is responsible for building the Metadata repository of select resources. These are based on the OAI-PMH protocol.

8.4.1.5 RESOURCE CONTROL MANAGER

This is an important system responsible for building a reference pointer to the resources both federated and harvested. This comprises of controlled vocabularies developed over the resources with source identifier. When a user submits query, the system facilitates semantic mapping with the controlled vocabularies and intelligently identifies the resources that satisfies the users search query.

8.4.1.6 PRESENTATION MANAGER

Presentation Manager is responsible for normalizing the search results from various resources and providing a common display of search results through a web browser.

8.4.2 SYSTEM ANALYSIS AND DESIGN

All the above system components are interconnected and interrelated. The functionalities of the above components are accomplished through three main interfaces as described below:

- IIM-UIPS User Interface
- IIM-UIPS Librarian Interface
- IIM-UIPS System Interface
8.4.2.1 IIM-UIPS USER INTERFACE

When a user accesses the IIM-UIPS portal:

- He will see a list of collections and catalogues and may select those he would like to search (collection level services)
- Then he enters the query (search service)
- He may get help in entering the correct terms from a name authority file (authority service)
- The result will be a list of short records that enables him to select the relevant ones (identification and description)
- Then he retrieves the full record (retrieval)
- and will further inspect the metadata to determine whether this is the object he wants to access (identification and description)
- Next he will move up or down in the object hierarchy or jump to related metadata records (navigation)
- or he may decide to retrieve the object (link service)
- Before that he wants to know whether he has permission or the rights to access the object or how he can obtain permission (authorization)
- and whether he has the right equipment (hard- and software)
- Finally metadata might be exchanged with other systems or downloaded

Of the three interfaces, user interface is an important interface, which provides users access to right resources at the least time possible through an intelligent resource controller. Following figure (Fig. 8.4) illustrates the user interface environment.
The Presentation layer in this interface is a web browser. The browser connects to the unified login system of IIM-UIPS for users authentication. Based on the attributes passed by the authentication system, the interfaces provides the list of resources which the users can browse, search across the resources, normalize the search results and finally get access to full text if available. These resources include the resources available in the originating IIM, the common OPAC, the commonly subscribed resources, and cross-shared resources of all IIMs.

When a user accesses the IIM-UIPS, the user credentials are authenticated through an authentication system. On successful login, the user gets access to various resources hosted on the IIM-UIPS. The user will be provided with a My-Library module wherein
the required information based on the already constructed user profile is pushed to the users. Optionally, the user can browse through the resources (arranged in classified order) or choose a common search interface to search across the resources. The workflow in the user environment is shown as flowchart diagram in the figure (Fig. 8.5):
Figure 8.5: IIM-UIPS User Interface Flow Chart

User

Authentication?

Y

IIM Unified Intranet Portal System

N

Public Domain resources (Internet, OPACs, etc.)

End

My Library (Profiles based)

Want More?

Y

IIM Intranet Resources

N

End

Want More?

Yes

Institution (IIM-A, B, C, I, K, L)

IIM Unified Search system

End

Other

Resource type (OPAC, e-journals, databases and other e-resources)

 Classified/Hierarchy

Yes

No

End
8.4.2.2 IIM-UIPS LIBRARIAN INTERFACE

IIM-UIPS Librarian Interface allows the library staff of IIMs to manage the following:

- **Users**
  - add users,
  - modify users attributes,
  - remove users,

- **Resources**
  - add resources,
  - modify the resource attributes, and
  - remove resources.

The figure (Fig. 8.6) illustrates the IIM-UIPS Librarian Interface environment.
A typical selection of electronic resources for consortia and evaluation process might involve the following steps (DLF, 2004):

- Build request record for trial;
- Flag resource as trial;
- Notify interested parties that trial is live and provide the access instructions and expiration date;
- Allow staff (and, optionally, users) to access the trial via stored URL;
- Send reminder to those notified of a trial that expiration date is near.
Record opinions of key players and final purchase decision, allowing for both central or local input of opinions and funding commitments into notes fields in request record;

Prompt licensing/acquisitions contact to report purchase decision. Most common options: approved, rejected, or on hold;

For resources that are approved for purchase, additional statuses and actions would be required, including:

- Notify licensing contact to obtain and negotiate license, and remind this contact if license is not completed by a given deadline
- Notify access contact if status has not changed to "live" by deadline
- Notify cataloging and selector/product sponsor as well as other interested parties when access is available

8.4.2.3 IIM-UIPS SYSTEM INTERFACE

The Intelligent Unified interface identifies based on the user's query the resources, issue appropriate search commands through appropriate protocols/standards, retrieve the records, deduplicate the records and normalize the records for uniform display. Once the relevant items selected, the user can link to the full text record if available.

Each IIM will act as both host and server. IIM-System interface will act as client to access Z39.50 targets (IIM OPAC servers), Metadata repository for locally developed resources and other oai-enabled subscribed resources and OpenURL resolver for OpenURL based resources.

IIM-UIPS System interface comprises of various modules:

- Authentication
- Search and results
- Resource Control
- Metadata Repository
Federated search system
• Interface to Library Management system
• Report Generator

For the various modules of the IIM-System interface to work, various standards and protocols have to be adopted. The primary ones are:

• For interfacing with local authentication or authorization systems - LDAP and Shibboleth
• For distributed searching and resource discovery - Z39.50, HTTP, and XML
• For document delivery services - ILL protocol
• For context sensitive Link services - OpenURL
• For harvesting metadata records – OAI-PMH
• For interfacing with library management software – SIP2, NCIP

Following figure (Fig.8.8) illustrates the various modules of the IIM-System Interface and the protocols/standards that support interaction of these modules:
Figure 8.8: IIM-ULPS System Interface Environment
8.4.2.3.1 AUTHENTICATION AND AUTHORIZATION

The Authentication and authorization system will be one single authentication gateway to all authorized resources and extended services. The authentication method should be easy to use and at the same time it must be secure: secure data transport between client and server. The core access management system relies on standard (e.g., LDAP) or emerging protocols (e.g., Shibboleth) to distribute identity verification (authentication) of distinct communities of users. In other words, the user’s “home” institution performs user identity and capability management. Supported identity services range from those requiring individual logon to group-based mechanisms, such as organizational proxies and network topology identity (IP address), to anonymous identity (Lagoze, et al. 2002).

Shibboleth, the next generation protocol is emerging as a promising protocol for distributed authentication management effectively. It is a standard that can be used for cross-institutional access no matter what authentication mechanism the Portal uses for local users, and no matter what mechanism the remote user's home institution uses (LDAP, CAS, etc...). In this system, there need not be a common user database for all the IIMs. Shibboleth takes care of passing the attributes of users from originating IIM to the target IIMs and provides access to the resources based on the attributes of users.

The Shibboleth model defines three roles that are involved in access management for electronic resources (Carmody, 2001):

- The browser user (for example, the library patron accessing electronic resources)
- The origin site (the institution to which the user is affiliated; responsible for authenticating the user and providing attributes)
- The target site (the resource provider; responsible for granting access based on the attribute information about the user provided by the origin)

Shibboleth works by retaining the authentication/authorization information locally while releasing the remote side only the attributes it needs to make its authorization decisions, thus preserving the end user’s privacy to the highest degree possible. The user’s home
institution will retain the administration of user identifies and attributes instead of instead of the remote systems needing to assign site-specific user IDs and passwords.

IIM Consortium scenarios are complicated by the fact that individual users may have multiple identities among the member institutions. For example, a staff member at one IIM may be a graduate student at another. Or, a faculty member of one IIM may visit another IIM library. In these cases it would be ideal if the entitlements from all the identities could be merged, so users could access resources to which any of their identities are eligible. Few, if any, library portal systems support this now; multiple identities are generally handled by providing a mechanism for patrons to specify a particular institutional affiliation before accessing resources through the portal. Such a mechanism should be easy to implement in Shibboleth environment. The following figure (Fig. 8.9) explain this better (Gourly, 2003):

Figure 8.9: IIM-UEPS Shibboleth enabled authentication interface (adapted from Gourly, 2003)
Within a consortium that shares library resources it is possible that some member institution campuses have Shibboleth origins sites set up and others don't. In order to provide uniform resource access to library resources through Shibboleth, the consortium would need to provide Shibboleth origin sites (based on the library patron directory) for those members that don't have their own.

### 8.4.3.2.2 IIM-UIPS SEARCH INTERFACE

IIM Search Interface provides user with a search window wherein the search keyword(s) are entered. The search system also provides the user the variants of search terms from which user can select the appropriate words. This is particularly useful when user types incorrect spelling of the search keywords (author name, etc.) The search keywords are mapped against the resource control database to identify the resources that meet the search criteria. It lists first the source names. User can select the sources (single or multiple) and system employs appropriate search mechanism to the targeted sources and retrieve the metadata records. The user then can access full text of the metadata records if available. The search system also provides the user option of searching the targeted sources with their native search systems. Following figure (Fig. 8.10) illustrates this search interface flowchart.
Figure 8.10: IIM-UIPS Search Interface Flowchart

8.4.3.2.3 IIM-UIPS METADATA REPOSITORY (IIM-MR)
The IIM-MR provides a facility for storing metadata entities and the relationships between those entities, creating a platform on which to build essential services. It provides the following basic functions:

- central storage of metadata harvested from resources
- output interfaces that provide metadata to services, such as search and browse services
- input interfaces that enable ingest services to provide new metadata and update existing metadata

There are three main components in the framework (Ding, 2003). They are Metadata Registry Repository (MRR), Schema Repository (SR), and Terminology Reference Repository (TRR). MRR is a system that registers and manages the metadata the specific digital library system will support. MRR will be used as a mediator for semantic information (data). It is responsible for the unique identification, registration and service of metadata. SR contains the schema information of local databases, such as table names, metadata element names in the database, value, and referential integrities. SR will support a mapping to the MRR according to its semantic. By doing so, not only will SR show high precision for end-user queries, but also it contributes to solve the semantic heterogeneity of data. TRR deals with the heterogeneous terms that may exist in the local databases and indexing databases. It manages the relationships among different terms, such as generalization, specialization and inheritance. TRR will also parse the user queries because it can integrate the term without requiring all resources to use the same terms.

The figures (Fig.8.11 and Fig. 8.12) show the process involved in developing the metadata repository.
It may be noted that both the above figures illustrate harvesting more of oai-enabled sources. Not all potential sources of metadata will necessarily be OAI-PMH data providers. It may be possible to make use of Z39.50 or SRW/SRU interfaces as well, in order to be able to harvest from any appropriate provider. The following diagram (Fig. 8.13) illustrates how metadata could be harvested and processed for entry into a metadata store using the OAI-PMH and other protocols (Swan, et al, 2004).
Swan, Alma et al. (Swan, et al., 2004) recommended that metadata is harvested 'as is' and saved locally in a temporary store prior to further processing. This has a number of benefits when compared to the alternative of processing on the fly:

- It substantially reduces the harvesting time
- The possibilities of encountering network errors are decreased
- It allows for more comprehensive and safer pre-processing of candidate metadata before it enters the database, e.g. setting or verifying metadata semantics (which may vary between data providers), or identifying duplicates or different versions of resources already catalogued in the database

### 8.4.3.2.4 IIM-UIPS FEDERATED SEARCH SYSTEM

As the resources available in IIMs have different protocols and standards support, IIM-UIPS federated search system will need to interface with all these protocols and standards. These resources are grouped essentially into the following modules:

- Z39.50 module
- OpenURL module
The federated search system of IIM-UIPS will therefore look like the following figure (Fig.8.14):

8.4.3.2.4.1 Z39.50 MODULE

IIM-UIPS system interface to z39.50-based resources comprises of

- IIM Z39.50 client
- IIM Z39.50 server

Resource registration and resource configuration are the two actions which enable access to z39.50 based resources.
The following figure (Fig. 8.15) shows how an IIM can act both as Z39.50 client as well as Z39.50 server. A user from one IIM submits through z39.50 client a query, which will be converted into a searchable query, and search is performed on the z39.50 server, results are retrieved and presented to users through z39.50 client (Ding, 2003).

Figure 8.15: IIM-UIPS Z39.50 Model of Information Retrieval (Adapted from Ding, 2003)

The figure (Fig. 8.16) below explains how a z39.50 search session takes place (Velegrakis, et al., 1999):

Figure 8.16: IIM-UIPS Z39.50 based search session (Adapted from Velegrakis, et al., 1999)
Module 1 is responsible for network communication with the client and is based on the Yaz toolkit. When it receives a search request it decodes it into appropriate query structures. More specifically, it produces the syntax tree of the query that is included in the search request and sends it to Module 2. When a response has to be sent back to the client, this module is responsible for the transformation of the answer to the appropriate network format.

Module 2 is used only during the search process. When it receives the syntax tree of a Z39.50 query, it translates it to a preliminary DL expression that is sent to Module 4. After the query execution, it receives the id and the cardinality of the result set (not the data themselves) and forwards this information to Module 1 to be sent back to the client.

Module 3 is used only during the retrieval process. After receiving a Z39.50 result set id it communicates with Module 5 to get the retrieved records in the form of structures. The task of Module 3 is then to encode the returned structures in one of the record formats defined in the Z39.50 profile (i.e., XML) in order to send the retrieved records back to Module 1. Modules 4 and 5 essentially form the DL-based wrapper for the underlying source (dotted line in Fig.8.16). Module 4 loads the source schema and the AP mappings (Tbox) from a configuration, i.e., while the data reside in the source (virtual Abox) and can only be cached in the DL system. When it receives a DL query from Module 2, it rewrites it according to the defined AP mappings and the paths to central concept of interest and forwards the resulting expression for evaluation to the underlying source. Finally, Module 5 converts the retrieved objects of the central concept by taking into account the mappings of the Z39.50 Record Elements to the source data.

8.4.3.2.4.2 OPENURL MODULE

For OpenURL to work, a resolution server is needed that knows what full-text journals are available and how to link to them as well as how to link to local print holdings and other local services being offered. The information provider, also referred to as the source, must be OpenURL-enabled to redirect the linking request to the local resolution server. The local resolution server resolves the link from the information provider to the
appropriate copy of the full-text and provides the data needed to locate the full-text. In addition to the link to the full text, the local resolution server offers the user other localization options such as the ability to link to local library holdings, link to an Interlibrary Loan request, etc. The localization of linking allows libraries to provide access to hybrid collections that include both electronic and print journal. The overview of OpenURL architecture is provided in Fig. 8.17.

In order to enable an OpenURL based search,

- Install an OpenURL resolver locally
  - SFX, LinkFinder Plus
- Load OpenURL resolver with local holdings information
  - Provider, journals, coverage
- Tell source vendors to enable OpenURL functionality
  - Identify your OpenURL resolver to them
- Once enabled, the “OpenURL-enabled” button will appear as an option

![Figure 8.17: The Schematic Diagram of OpenURL environment](image)

Following are the steps involved in using OpenURL enabled resources:

- A user or user agent (Requester) access an html page that contains an OpenURL
The OpenURL was inserted by a third party service providers that want to allow the requester to choose to receive service options related to the subject of the html page accessed.

- The requestor chooses to activate transportation of the encoded metadata in the OpenURL to the service component (Resolver) specified by the OpenURL. Clicking a link usually activates an OpenURL. The encoded metadata is passed to the resolver.

- The resolver interprets the encoded metadata based on business rules maintained in its local information service environment to identify one or more services to be returned to the requester. The business rules are not part of the OpenURL standard.

- The services are returned to the requester via an html page.

- The requester selects one or more services by clicking on the link(s) provided.

The figure (Fig.8.18) below illustrates the above steps better.

*Figure 8.18: IIM-UIPS OpenURL search session (Adapted from, Caltech)*
8.4.2.3.5 IIM-UIPS RESOURCE CONTROL SYSTEM

One of the current weak points of library portals, or of any cross-database searching service, is that even if a user's search is reformatted into the correct syntax and protocol of the target system, it may contain terms that are inappropriate. The user who has typed a particular term(s) in a search strategy may find that this term(s) gets good results in one database, but not so in another even though the database does have the relevant records. This is because the search terms are not common to the resources. Each resource uses their preferred term(s) for indexing. The solution is to make use of enhanced services based on controlled vocabularies, where terms in one index can be mapped into terms in another. There is much current interest and research in this area in respect to name and subject authority files, particularly thesauri (Davies, 2004).

One important issue for the IIM-UIPS is the ability to perform Metadata based searching across a range of different resources. This is particularly difficult when the data in the databases originates in different domains because usually different databases have their metadata coded using different standards. The number of resources that are available today continues to grow. Any attempts to search these resources run into the problem that the user's terminology does not match the database terminology. Use of controlled vocabularies of resources specific can bridge this gap. A search engine that is able to do this multi-server/cross-domain searching has to deal with two related problems (Bradley, 1998):

a. a technical one: there needs to be a way to merge metadata coming from different sources and possibly structured using different encoding conventions — MARC from one, DC from another, etc. — into a single coherent format which can be shown to a user; and

b. more seriously, a semantic one: there needs to be a way to interpret metadata element data from different sources in some consistent way so that, as far as
possible, any particular field when displayed to the user, contains data that can be interpreted by him/her in one way no matter what source the data came from.

IIM Resource Control System is an enhanced and intelligent vocabulary control system developed based on the search terms derived from metadata database and other sources. The figure (Fig. 8.19) below illustrates the process involved in building a resource control system based on vocabulary control.

![Diagram of IIM-UIPS Resource Control System]

**Figure 8.19: IIM-UIPS Resource Control System**

A basic controlled vocabulary consists of three elements: a list of the terms, and an equivalence relationship and resource identifier(s). Resource Identifier is a unique identifier that identifies the target resource. The equivalence relationship doesn't mean we can only use synonyms; it means that the terms are equivalent for our search purposes.

The user must be able to see a list of the top terms in the vocabulary. Users should be able to enter search terms directly, or to identify them by browsing
If an indexer enters a term directly to the Subject metadata, the software should check its validity against the controlled vocabulary applicable and complete any syntax requirements, such as identification of the source vocabulary. Any non-preferred term entered should be converted to the corresponding preferred term, and the indexer should be invited to confirm its appropriateness. If the term entered is not present in the source vocabulary, then it may be retained among the metadata, without any refinement, as an uncontrolled term.

When a user enters a search query, system first looks at the controlled vocabulary and select all of the related terms. Then it automatically replace the users' search query with a new one that groups all the related terms together, and send that to the search engine. This is called search query expansion; it's simple but efficient.

Used like this, a controlled vocabulary helps locating relevant resources and also expands the result set.

8.5 DISCUSSIONS

IIM-UIPS design framework demonstrates that a mix of both centralized and distributed models would serve the purpose of sharing heterogeneous and distributed resources among the IIMs. It also demonstrates that both harvesting (central and distributed) and federated system are required as they complement each other in order to achieve the desire user environment that makes the information sharable and accessible across IIMs through a single window. Since IIM-UIPS will need to interact with varied protocols and standards (harvest and federated), an intelligent resource control system is proposed to bring-in a common controlled index for both harvest and federated resources. The IIM-UIPS design framework exploits all the promising standards and protocols to make the system scalable and future enabled.

Although, SRU (search and retrieve URLs) based approach is a promising solution much simpler and effective than Z39.50 based approach, it may not be considered for IIM-UIPS environment due to the fact it is browser dependent
Chapter 8: IIIM Unified Intranet...

(Internet Explorer). The overall approach in the IIIM-IIIPS design framework is to try and use open sources (protocols and standards).

8.6 CONCLUSION

One solution is rarely right for all circumstances; one user interface or one search technique is unlikely to suit all circumstances. The tricks are to determine which solutions make most sense for which cases, come up with good ways to guide users to the solutions that make most sense for them. The IIIM-IIIPS is a mixed of both centralized and decentralized models. The design framework of IIIM-IIIPS is based on a number of investigations and enquiries into available standards and technologies, functional requirements of users, staff and systems. The technical starting point at the beginning of the design of IIIM-IIIPS was the heterogeneous nature of access to the data of the participating IIIM libraries. All the distributed resources that we are interested in may not be searchable in a standard way, nor can they all easily be made searchable. Therefore, in some environments, it makes more sense to harvest a subset of resources, to build a consolidated index, and then to search that consolidated index through the portal at the same time as the distributed resources. The IIIM-IIIPS, in fact, ends up relying on a mixed model of searching, with some resources distributed, and some harvested and centralized into a single searchable index.

The system proposed will act as a unified intranet portal to offer integrated access to the combined resources of the IIIM libraries. The system will have a central authentication mechanism to facilitate single sign on approach to access all the resources across. The system proposes an IIIM-IIIPS Metadata repository to pool the metadata from resources that are OAI compliant. The system also provides a custom scripted extraction of metadata from sources that do not support OAI. The IIIM-IIIPS federated system builds on two effective protocols viz., Z39.50 and OpenURL. The ILL request for an item from one library to another library is facilitated through the emerging protocols NCIP and SIP2. In order to make the search effective and to yield better search results by increasing the precision and recall value, the system proposes a unique IIIM-IIIPS
Resource Control System that builds on controlled vocabularies extracted and normalized from targeted sources. This intelligent resource control system acts as heuristic system to identify the target resources based on the nature of the query and structure and provides appropriate search mechanisms. In effect, The IIM-UIPS will facilitate bringing together on the user’s desktop disparate collections and will allow for cross-collection searching. It will present integrated results and will deliver digital objects.
8.7 REFERENCES


DOI:10:1108/02640470210432357

