System Requirements

OpenDLib is an HTTP (web) server-based software system, requiring only open source technologies and running on several major operating systems.

OpenDLib is distributed as a set of self-installing packages with instructions to use them in a simple way.

The system is developed as a Service Oriented Architecture application and it adopts XML as standard for internal (between services) and external messages exchange.

Linux and other Unix-like systems

The standard supported configuration of a distribution of OpenDLib for a typical Linux system is as follows:

- A 32-bit Intel i386-compatible microprocessor (e.g., Pentium IV or higher)
- Red Hat Linux, ver. 9
- Apache HTTP server for Linux / Unix, ver. 1.3.x (available from *apache.org*)
- Perl interpreter pre-built for Linux on Intel, ver. 5.8.0 or higher (a standard component of Red Hat Linux 9)
- Mod_Perl ver. 1.2.x (available from *perl.apache.org)*
- EmbPerl framework ver. 2.0 for the User Interface visualization (available from *perl.apache.org*)

Nevertheless, OpenDLib can be executed on any Unix-like system equipped with the software above.

Other operating systems

OpenDLib can be configured to run on other operating systems including, for example, Macintosh OS X and Windows platforms. However, at the present time, these additional configurations are not fully supported by the DLib Team.

User Side Requirements

Any of the following web browsers on Windows, Mac OS, or Unix/Linux with Javascript enabled:
- Internet Explorer ver. 6.0 or higher
- Netscape ver. 7.0 or higher
- any other Mozilla-based browser

www.opendlib.com

Links to OpenDLib digital libraries are available at the project home page.

OpenDLib

Digital Library Management System

Digital Libraries are no longer only the digital counterpart of physical libraries (or physical museums, video archives, etc.), rather they are complex networked systems able to support communication and collaboration among different, worldwide distributed user communities.

DIGITAL LIBRARY MANAGEMENT SYSTEM

Nowadays digital libraries are instruments for supporting communication and collaboration among worldwide distributed user communities. OpenDLib is a digital library management system that makes it possible to satisfy this demand by supporting a cost-effective digital library creation and operational model.

OpenDLib consists of a federation of services that can be customised to satisfy the requirements of a target user community. This federation can be expanded at any time by adding other community specific services. The entire set of services can be managed and hosted either by a single or by a multitude of organisations that collaborate on the maintenance of the shared digital library, each according to their own computational and human resources.

An orthogonal system facility enables different user groups to define their own virtual view of the shared digital library, tailored to the specific needs and policies of the group.

A federated digital library creation and operational model for enhancing cost-effectivenes
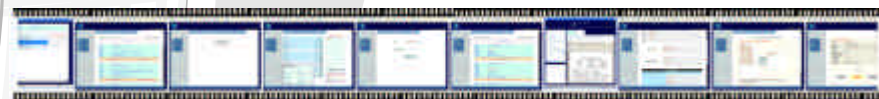
USER SERVICES

The basic release of OpenDLib provides services to support the submission, description, indexing, search, browsing, retrieval, access, preservation and visualization of documents.

Documents can be submitted as files in a chosen format or as URLs to documents stored elsewhere. They can be described using one or more metadata formats. The search service offers different search options: text free or fielded (with fields selected from a variety of known metadata formats); single or cross-language; with or without relevance feedback. Documents retrieved can be navigated over all their editions, versions, structures, metadata and formats. All the above services can be customised according to several dimensions such as, for example, metadata formats, controlled vocabularies, and browsable fields.

OpenDLib also provides other digital library specific services, such as the control of access policies on documents, and the management of "user-shelves" able to maintain document versions, result-sets, session results, and other information, etc. In addition, a number of administration functions are also given to support preservation of documents, document reviewing process, introduction of new collections, and handling of users and user group profiles.

Customisable services and expandability for meeting specific application requirements

DOCUMENTS

OpenDLib can handle a wide variety of document types with different formats, media, languages and structures. The same OpenDLib library can maintain, for example, a collection of journals and conference proceedings consisting of articles; a collection of theses in different languages, organised in chapters and sections; a collection of videos, structured into sequences and shots, and a collection of other documents represented only by the set of their bibliographic records.

OpenDLib can also manage new types of documents that have no physical counterpart, such composite documents consisting of the slides, video and audio recordings of a lecture, a seminar or courses. It can also maintain multiple editions, versions, and manifestations of the same document, each described by one or more metadata records in different formats.

The documents of an OpenDLib library are organised in a set of virtual collections, each characterised by its own access policy. Authorised people can define new collections dynamically by specifying definition criteria. In the same digital library, for example, it is possible to maintain a collection of grey literature accessible to all users and a collection of historical images accessible only to a specific group of researchers. Each collection is automatically updated whenever a new document matching the definition criteria is published in the library.

A rich document and collection model for handling a large variety of document aspects and access policies.

ARCHITECTURE

From the architectural point of view, OpenDLib consists of an open federation of services that can be distributed and replicated. This architecture provides a great flexibility in the management of the digital library. For example, an institution can decide to maintain an instance of the repository service, in order to have local control over its own documents, but to share all the other services with other institutions.

The architectural configuration is chosen when the digital library is set up, but can also be changed later to satisfy new emerging needs. For example, a replication of an Index instance can be created to reduce workloads when the number of search requests exceeds an established threshold, whereas an Index instance, able to serve queries in a language not previously supported, can be added to satisfy the needs of a new community of users. All these expansions can be done on the fly, i.e. without switching off the digital library. The OpenDLib architecture has been designed to be highly interoperable with other libraries. In particular, an OpenDLib library can act as both an OAI-PMH data and service provider. This implies that the metadata maintained by an OpenDLib digital library can be open to other libraries and, vice-versa, the OpenDLib services can access the metadata published by any other OAI-PMH compliant library.

A distributed service oriented architecture for achieving expandability and interoperability with other digital libraries

System Architecture

OpenDLib is a **networked system** hosted by servers belonging to supporting institutions. Services can be distributed or replicated on more than one server. The topology of the communication among the service instances allocated on different servers is dynamic since it takes load balancing and bandwidth monitoring techniques into account. The **open architecture** makes it possible to expand system functionality by adding new service components instead of re-building the whole system.
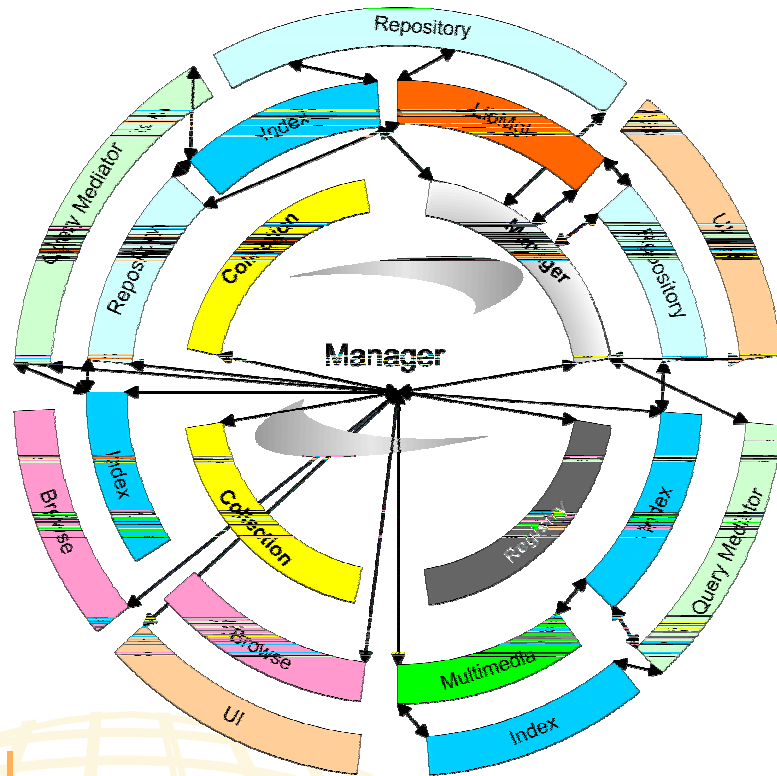
An OpenDLib networked federation of service instances. The arrows show the flow of service requests among the instances.

OpenDLib is growing over time along several dimensions, e.g. services, metadata formats supported, host servers, user communities, searchable metadata, manifestations handled, etc.

In particular, OpenDLib supports three kinds of dynamic service expansions:

✓ new services can be added;

✓ new instances of a replicated or distributed service can be mounted on either an existing or a new hosting server;

✓ service configurations can be modified so that they can handle new document types, new metadata formats and support new usages.

The overall functionality of OpenDLib is partitioned into a set of **well-defined interacting services** that provide:

✓ co-ordination functions (e.g. mutual re-configuration, distribution and replication handling, work-load distribution);

✓ basic utility functions, (e.g. user handling, rights management, information space mediators) ;

✓ application functions (e.g. acquisition, storage and preservation of documents, information search, browse and retrieval, dissemination).

Service interactions are more complex than in the client-server model since a service can act both as an information provider and as an information consumer.

OpenDLib is explicitly designed to support plug-and-play expansions.

OpenDLib

*D i g i t a l   L i b r a r y
M a n a g e m e n t
S y s t e m*

System Architecture

Services provided by the basic release of OpenDLib

Repository. Stores and disseminates documents that conform to the DoMDL document model which can represent structured, multilingual and multimedia documents.

Library Manager. Supports the submission, withdrawal, and replacement of documents through a complete review workflow.

OAI Harvester. Harvests content published by OAI-PMH (Protocol for Metadata Harvesting) compliant archives.

OAI Publisher. Publishes the content of the OpenDLib DL through the OAI-PMH protocol.

User Interface. Mediates between human interactions and the application services.

Manager. Maintains a continually updated picture of the status of the DL service federation and disseminates it on request to all the other services.

Query Mediator. Dispatches queries to Index service instances, according to availability.

Browse. Supports the construction of indexes to browse the library content. The Browse function is parametric with respect to the metadata formats, to the set of browsable fields, and to the set of formats for result sets.

Index. Accepts queries and returns documents matching those queries. The Index function is parametric with respect to the metadata formats, to the set of indexed fields, to the set of result sets formats and the language of the terms.

Collection. Mediates between the virtual dynamic organization of the content space, built according to the requirements of the DL community of users, and the concrete organization into basic collections of documents hold by publishing institutions.

Registry. Maintains information about the users, groups, and communities.

All the OpenDLib services are **configurable**, **customizable** and **extendible**. This means that the system can be used in a large variety of different DL application frameworks.

www.opendlib.com

OpenDLib

An OpenDLib **archive** is the basic unit in which documents are stored. A Digital Library (DL) empowered by OpenDLib can handle a great number of archives at the same time; these archives can be geographically distributed over a network.The physical organization of the archives is static and unique and reflects the specific needs of their creator, rather than the needs of the users that access them.

The OpenDLib developers have introduced an abstraction layer between the physical structure of the archives and the users. This layer is represented by the **information space,** the environment where search operations are executed. This environment is composed by a set of **collections** which are **virtual restricted views** of the content of a DL. For instance, one collection can contain all the documents produced by a specific institution and another can contain documents relevant to a given topic or search. A single collection can include documents that are managed by different archives and stored in different machines.

Users submit queries in the current information space, not in the whole set of archives. If users do not select a particular collection, a predefined information space is associated with their working sessions when they access the DL; this default environment contains all public collections. However, users can restrict their environment by selecting one or more collections from the list of those to which they have access rights.

www.opendlib.com

Though the construction of personalized collections, each user can perceive a different organization of the OpenDLib archives. Each user has his/her own **access rights** and at any time, he/she can access a different subset of the entire collection.

Another great advantage is that the physical location of the archives is completely hidden to users. Once they select a collection, the system knows where the queries should be executed, runs parallel jobs to remote archives, and merges the results appropriately.

In this sense, OpenDLib **virtualizes** physical archives.

An important result of the OpenDLib design is that collections are **dynamic views** of the information space. This means that, when a new digital object is submitted, if it satisfies the definition criteria of a collection, it is automatically included in the collection. Moreover, **data are not duplicated** in the library and if a digital object belongs to more than one collection, only one copy is stored.

D i g i t a l   L i b r a r y
M a n a g e m e n t
S y s t e m

OpenDLib

In the User Interface, the information space can be accessed by selecting "information space" on the left menu in the main page.

A list of collections is displayed. These collections are the active collections that compose the current information space. Users can customize their own space by adding/removing collections to/from the list of those available.

How can the information space be personalized?

Authorized users can create their own collections and decide to share them with other users. These operations are simple and immediate via the user interface, as shown in the following figures.

Document visualization

As result of their search or browse operations, users obtain a set of results pages with the list of documents that satisfy their request. OpenDLib manages digital objects that are compliant with its DoMDL model and visualizes them as a graphical rendering of this model. The visualization of these objects is more complex than the visualization of simple file.

Two visualization modes are available to the users: **tab page** and **top level**. With the tab page visualization, all object manifestations are displayed in the same window; with the top level mode, each manifestation is displayed in a new window.

Users can choose either mode, according to their preferences and to the type of digital objects to be visualized.
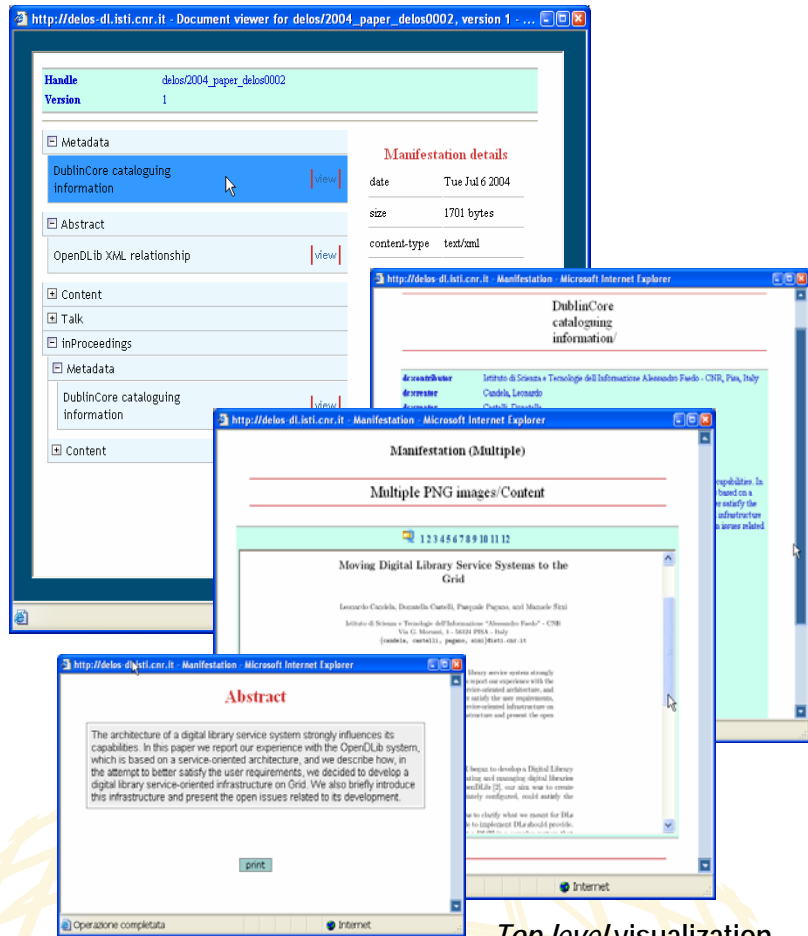
Top level visualization
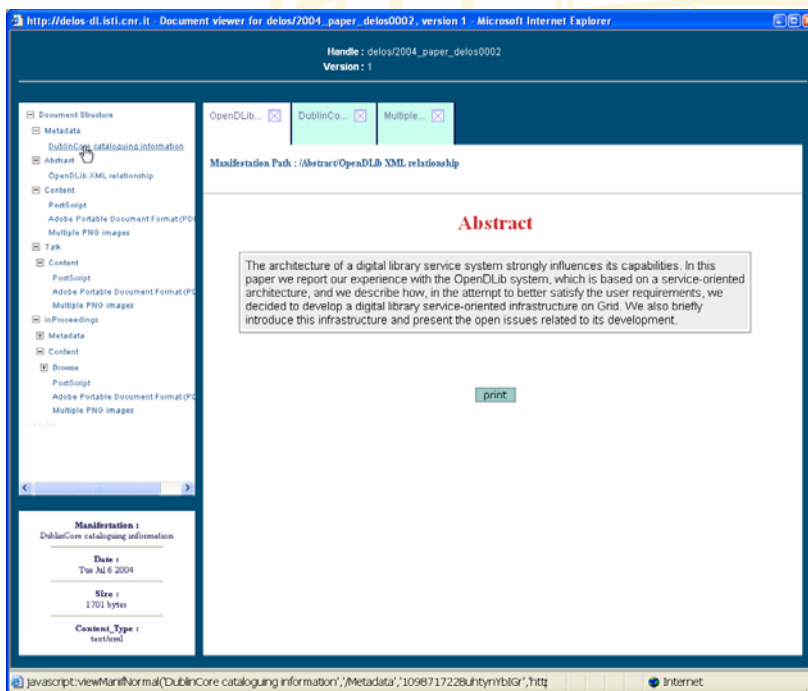
In both modes, a tree representing the structure of the retrieved digital object appears on the left of the first visualization page.

Although object data can be maintained in very different ways (stored locally or remotely, or as a reference to other data), this is transparent to the users who access them in the same manner.

Tab page visualization

According to the DoMDL specifications, a digital object is composed by a number of views and manifestations. By navigating the tree and clicking on the object manifestations, users can access the object data. OpenDLib is able to manage **every type of manifestation**, **regardless of the nature or dimension**. Multiple physical manifestations of the same view make it possible to select the one that is compatible with the software installed on the local machine. This is useful when no specific viewer for a given format (e.g, Microsoft Word) is available, but a suitable one (e.g., Acrobat Reader) is installed.

The images below show some examples of data visualizations on the OpenDLib user interface.

Document visualization
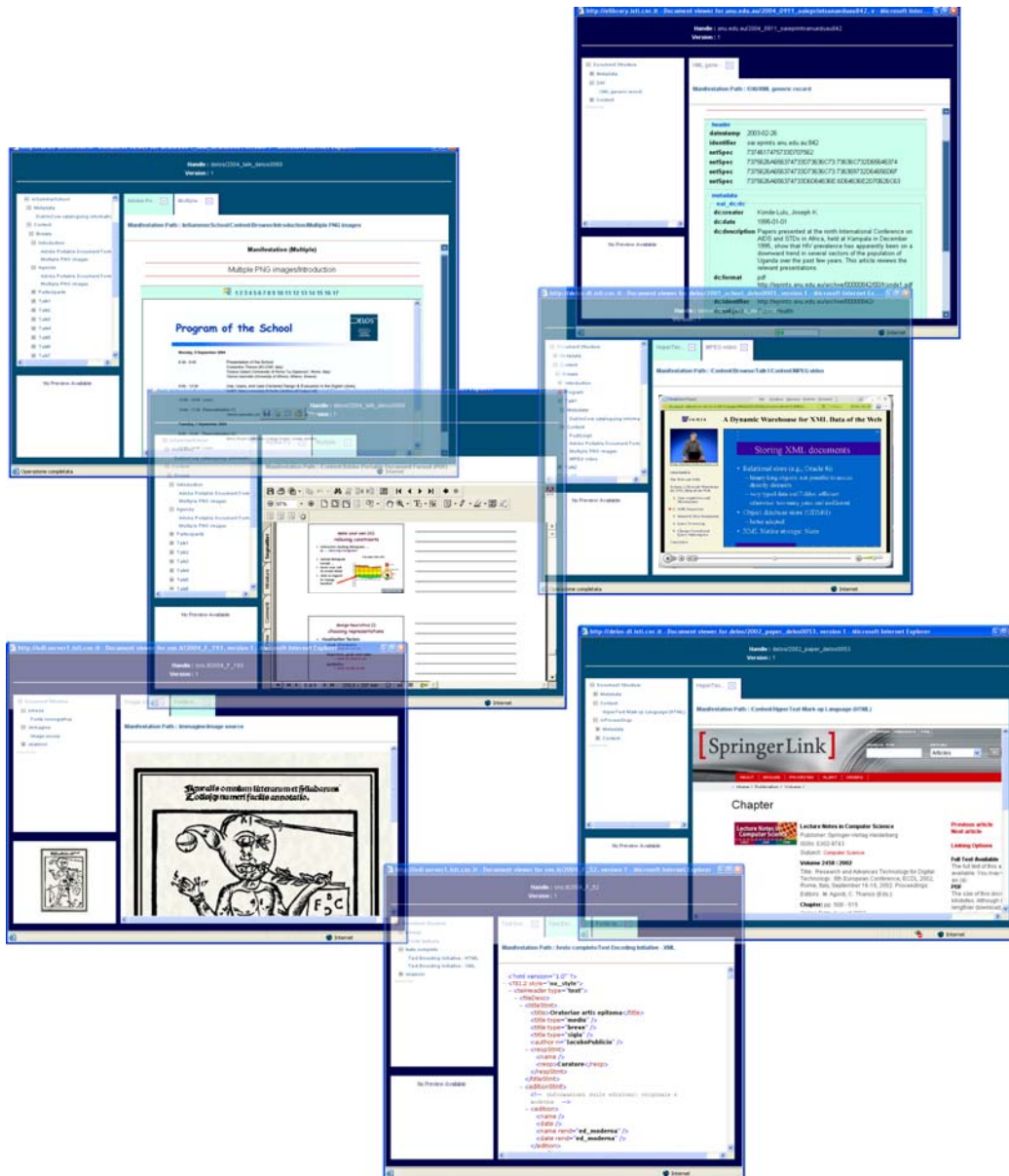
A major advantage when adopting the OpenDLib Digital Library Management System is the possibility to manage a great number of different document structures and types, and metadata formats in the same Digital Library (DL) using the same tool. This goal is achieved with the introduction of the OpenDLib **Document Model**.

OpenDLib does not manage document files directly; it manages "entities" (named **digital objects**) that conform to the Document Model for Digital Library (DoMDL). This model can represent a large variety of documents structures.

What is a Document Model?
A Document Model is the abstract representation of the physical or semantic structure of a given document.

According to this model, a digital object is identified by one or more **editions**. Each edition is split into a set of **views** that are logical partitions of the object. As the name suggests , a view is a "point of view" of the digital object; for instance, an object that represents a presentation of a paper in a conference could be partitioned in the following views:

- the PowerPoint presentation
- the video of the talk
- the audio of the talk
- a reference to the paper (already submitted in the library or stored elsewhere)

Descriptive metadata (expressed in a configurable and customizable set of formats) can be associated with both editions and views. These metadata are used by the Search Services to perform discover operations on the DL content. The metadata contain information about the edition or view which enables them to be identified and retrieved.

Each view contains its own **manifestations**. A manifestation is a physical representation of the object or one of its parts. The manifestations contained in a view are semantically equivalent (i.e. a PDF or a PostScript version of the same text), and the system can be configured to migrate automatically from one format to another.

DoMDL – Document Model for DL

The DoMDL model is shown in the figure below.
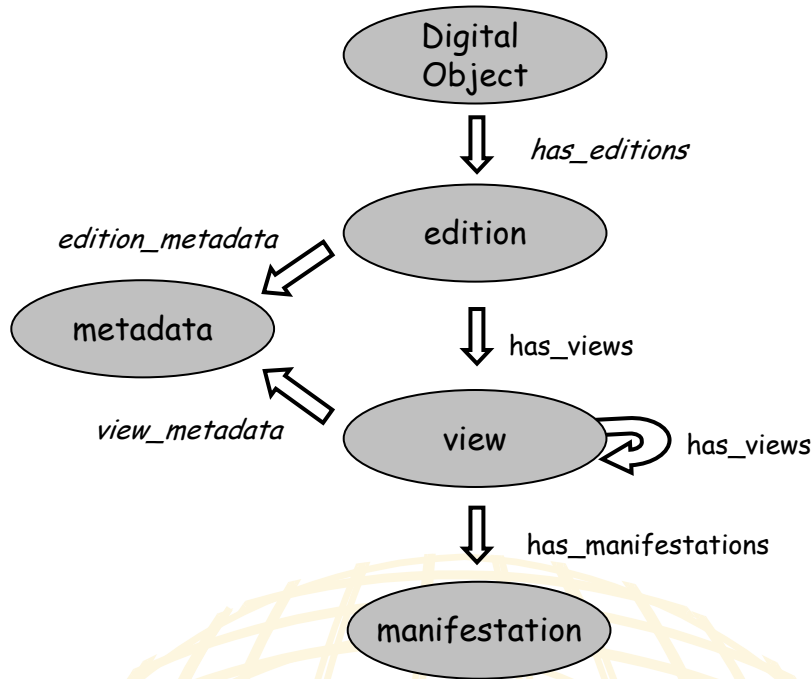
OpenDLib manages object manifestations with great flexibility. Every physical manifestation of the digital object can be:

- ✓ **locally stored**
- ✓ **retrieved** from a **remote server** and displayed at run time
- ✓ displayed in its **remote location**

This means that OpenDLib can be configured according to the **storage capabilities** of the machines on which it is installed. If a manifestation is too big to be stored in the local machine, a link to its remote location can be activated and the system will transparently redirect users to this location when they try to access the resource.
A manifestation can also be a reference to another object manifestation; through this mechanism, **data duplication can be avoided.**

A user-friendly interface leads authorized authors to submit their documents in an easy way. Submitters can chose among a number of predefined structures and just upload their data or create their own documents structures in order to fulfil their needs.

OpenDLib provides a set of dedicated advanced services (referred to as "Search Services") for discover operations on Digital Library content. These services include:

- Query Mediator
- Index
- Browse
- and, of course, User Interface

Thanks to these services, a user can **access the digital objects** maintained in the Digital Library (DL) that they are using.

The Search Services are independent modules with both networked and API interfaces. In an OpenDLib Digital Library, the number of instances of the Search Services and their physical distribution can be optimized by the administrator in order to improve performance and provide a high level of system reliability via backup instances. The Manager Service can reconfigure the map of the DL services when a service down occurs.

www.opendlib.com

While DL systems usually offer a well-defined (and limited) set of searchable fields and operators and only one metadata format managed, OpenDLib is a fully configurable system. It can handle **multiple metadata formats** at the same time and users can choose which one to

Result set	Search options	Search type
	Configure search ->	Simple
	Configure results page	Advanced
	View details	

search terms: ....

[search] [clear]

use in their queries. OpenDLib is also able to offer full text searches. When this option is selected terms are searched in the entire text of a document, not only in the metadata.
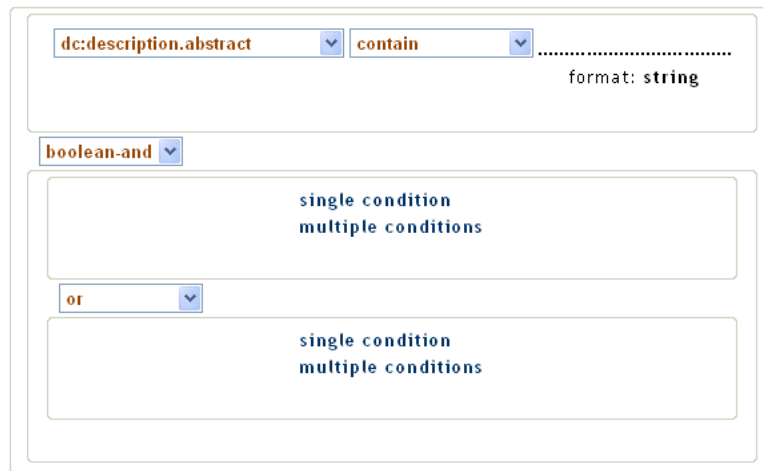
Other configurable options in Search Services include: selection of the language to be used in the query, selection of the metadata fields visualized in the results page, visualization mode (tabs page or top level), number of documents per page, and others. Each configuration parameter has its own default value, the user simply selects the **search working area** of the User Interface and then performs the query. The values of the parameters can be changed to meet the user needs.

Each value of these options is configurable and its availability depends on the global configuration of the OpenDLib instance. For example, if an Index service that indexes the Dublin Core format is being employed, the user can select Dublin Core as the query format.

D i g i t a l   L i b r a r y
M a n a g e m e n t
S y s t e m

OpenDLib

Search Services

In addition to the **free text search**, OpenDLib offers a powerful **advanced fielded query** mechanism which can be used to build complex queries **with no limitation** on the metadata used, the number and type of fields and operators, and the level of complexity.

When a result set is returned, a user can perform a **relevance feedback** query. By marking one or more document as relevant and submitting them, a new modified query is executed based upon the addition of terms extracted from the relevant documents. A new result set is returned.

Browse operations on a single collection of data are also supported. This functionality is especially useful when users do not have a clear idea of the content of the Digital Library they are working on.

www.opendlib.com