Master Euro-Mediterraneo di Specializzazione sulle Tecnologie Multimediali
Napoli - 13 Novembre 2000

FROM TRADITIONAL TO DIGITAL LIBRARIES

Maria Bruna Baldacci

IEI-CNR

Pisa

Abstract.

A digital library can be regarded as a system that performs or supports at least the functions of a library in a context of distributed networked collections. This very intuitive definition that recalls concepts of traditional libraries will not hide the fact that substantial changes are implied when we introduce the idea of digital libraries, and that systems managing libraries of digital documents must be capable of handling many issues unknown to traditional library systems. The new physical state of documents not only affects all the procedures operating upon documents as physical objects - collections organization, collocation, loan, etc - but also the administrative procedures versus publishers and patrons, and versus the Library itself, when considered as an administrative institution.

However knowing concepts underlying traditional library functions/services does help understanding digital libraries, mainly because the "new" physical state has a lesser impact on procedures which treat the information contents of documents - cataloguing, indexing classifying - as well as on information retrieval behaviour of users.

The seminar will present the main concepts of information searching and retrieval and introduce to the digital environment, where documents may be of totally new types and new mechanisms are required for cataloguing as well as for discovering network information resources.

FROM TRADITONAL TO DIGITAL LIBRARTIES

Maria Bruna Baldacci

baldacci@iei.pi.cnr.it

Istituto di Elaborazione dell'Informazione

CNR - Pisa

www.iei.pi.cnr.it

- NAPOLI - Master Euro-Mediterraneo di Specializzazione sulle Tecnologie Multimediali
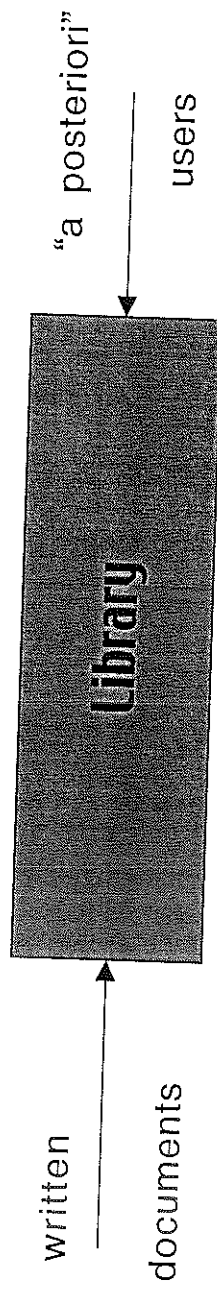 - 13 Novembre 2000

From traditional to digital libraries

■ *What is a digital library ?*

■ *"*A digital library has been defined as a system that performs or supports at least the functions of a library in a context of distributed networked collections of digital documents*"*

—But, what is a library ?

History in brief

Written documents make human communication independent from the speech

written documents → **Library**

"a posteriori" → users

Library as a "mediator"

History in brief

How libraries mediate between

authors communicating information

and

users searching for information?

The basic tool is

ORDERING

Physical ordering

- Documents are put in their places according to a given criterion:
  - Subject
  - Date
  - Type

Access is only possible according to that criterion

History in brief

■ **Making ordering separate from physical arrangement:** Ordering is not applied to physical documents; document representations are ordered instead.

- Each document is abstractly represented by describing its main bibliographic elements: author, title, subject, n.of pages......physical location

- Each representation is headed with those bibliographic elements which are judged to be useful to search the document they refer to. Such elements are called Access points

- Headed representations are ordered into a Catalog

■ *Catalogs allow access to each document through any of its access points*

History in brief

Cataloguing has been until now the main library service because it realizes the tool to search and retrieve information

- The development of library services has been strictly dependent on the development of information search and retrieval techniques

- The use of computers has empowered traditional information search tools [card catalogs] and given rise to innovative services
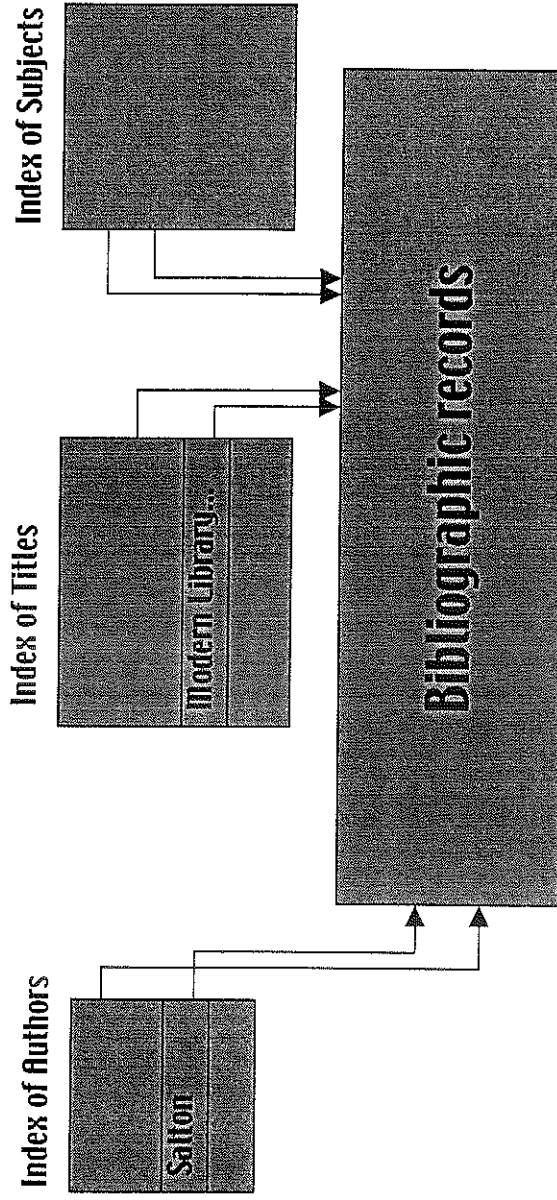
[1]

OPAC = On-line Public Access Catalog

i.e., catalog of electronic bibliographic records searchable by remote users

[2]

■ An electronic bibliographic record is a formal representation of cataloguing data

■ A bibliographic record is composed by a set of fields; a specific field for each bibliographic element.

| AU=Salton, G. | TIT=Modern Library Services: ...... | DATE=1979 |

Modern Library Services - OPAC

[3]

Computers are used to hold bibliographic records in an electronic archive and to create a set of indices to such an archive.

Modern Library Services - OPAC   *[4]*

■ While card catalogs allows searching by only one access point at a time, OPAC records are searchable by specifying more than one of indexed fields:

AUT = Salton, Gerald

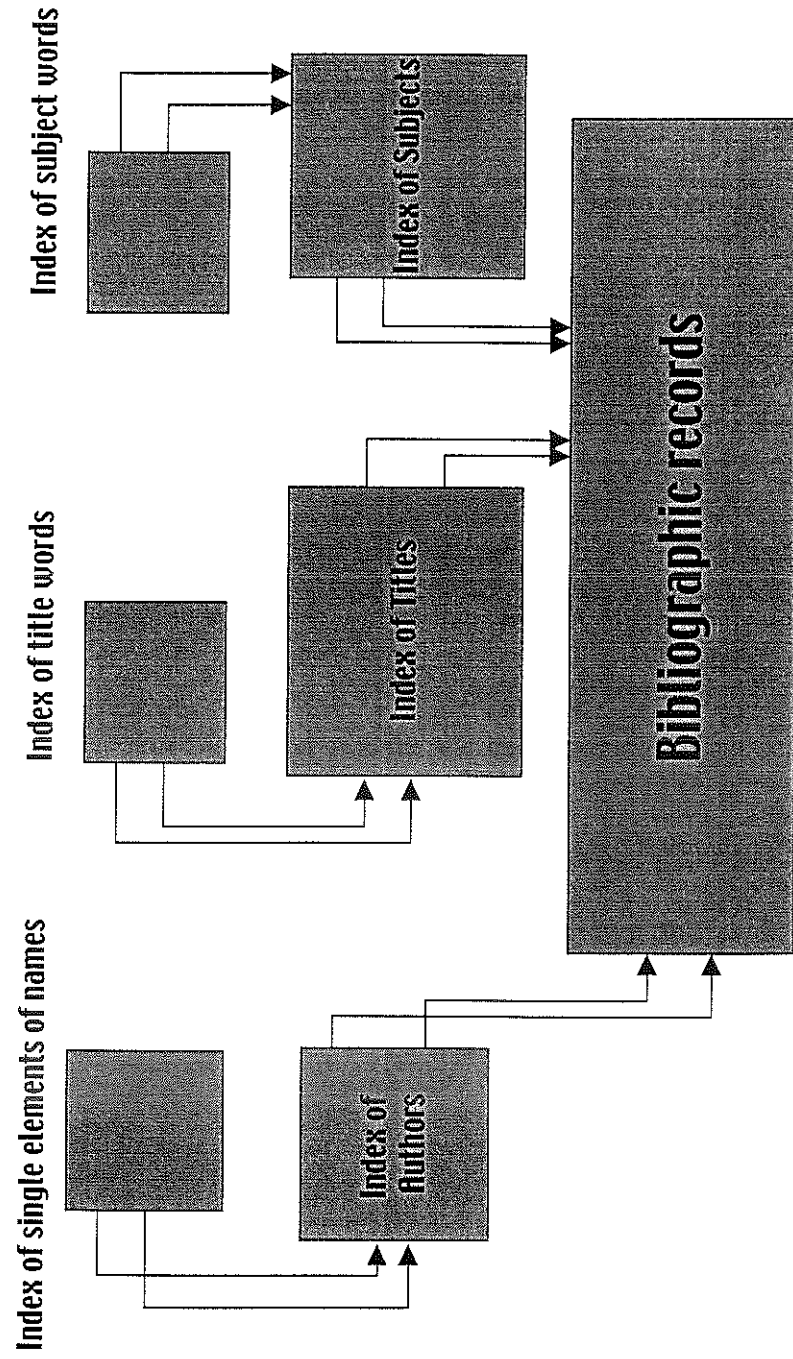   AND

TIT = Modern Library Services

     AND

    ....

Subject searching by a controlled vocabulary

- The subject of documents is described using terms of a controlled vocabulary

- Users express their queries using terms of the controlled vocabulary

If the same "language" is used, the successful matching between user queries and subject description is likely to happen!!!

Modern Library Services - OPAC   [6]

Index of subject words

Index of Subjects

Index of title words

Index of Titles

Index of single elements of names

Index of Authors

Bibliographic records

Word indexes allows searching with one or more words of indexed fields

AUT = Salton

AND

TIT = Library

■ **SYSTEM-USER INTERFACE: The way how a system presents itself externally to allow users to perform system operation**

■ *In modern OPACs search operations become more and more easy with the use of technologies for*

friendly user interfaces with
on-line helps

■ *From* TIT = Library Services

to Any = Library

thus retrieving all the document records where the word "library" appears in any of the bibliographic fields

From more PRECISION to more RECALL

ICCU Home Page

Back Forward Stop Refresh Home Favorites History Search AutoFill Larger Smaller Print Mail Preferences

Address: http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame Go

Live Home Page Apple Computer Apple Support Apple Store Microsoft MacTopia Office for Macintosh

ICCU

Ricerca Base

Istituto Centrale per il Catalogo Unico
Indice SBN

Modulo di Ricerca Base

(**Legenda**: ● *Lista delle frasi*; ● *Lista delle parole*)

Autore ● ●

Titolo ● ●

Soggetto ● ●

Classificazione:

Numero ●

Descrizione ● ●   rappresentazione e ricerca delle informazioni

Tutti i campi

Cerca

Cerca

Home
Indietro
E-Mail
Aiuto
Avanti
Accessi

Pagina Precedente

Favorites History Search Page Holder

ICCU Home Page

Address: http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame Go

Live Home Page Apple Computer Apple Support Apple Store Microsoft MacTopia Office for Macintosh

Istituto Centrale per il Catalogo Unico
Indice SBN

Risultati della Ricerca Base

(**Legenda**: *Testo completo del titolo nel caso di titoli molto lunghi*)

Criteri usati nella ricerca:
Titolo=rappresentazione e ricerca delle informazioni

Risultato della ricerca: 1 documenti

No.		Visualizza Tutto Visualizza Selezione Raffina la Ricerca ▲
(1/1)	☐	[Monografia] - **Baldacci, Maria Bruna** - *Rappresentazione e ricerca delle informazioni : come comunicare attraverso sistemi informativi automatizzati / Maria Bruna Baldacci* - Roma - 1988 (IT\ICCU\CFI\0155404)

No.		Visualizza Tutto Visualizza Selezione Raffina la Ricerca ▲

ICCU

▲ Ricerca Base

Favorites History Search Page Holder

Home Auto
Indietro Avanti
Chiudi Access

Internet zone

ICCU Home Page

ICCU

ICCU
ICCU

▲ Ricerca Base

Istituto Centrale per il Catalogo Unico
Indice SBN

Presentazione in Formato Sutrs

#1

Livello bibliografico:	Monografia
Tipo documento:	Testo a stampa
Autore:	Baldacci Maria Bruna
Titolo:	Rappresentazione e ricerca delle informazioni : come comunicare attraverso sistemi informativi automatizzati / Maria Bruna Baldacci
Pubblicazione:	Roma : NIS, 1988
Descrizione fisica:	127 p. ; 24 cm
Collezione:	Beni culturali
Numeri:	Bibliografia Nazionale - IT
Nomi:	Baldacci Maria Bruna
Soggetti:	● Archivi di dati
	● Documentazione - Automazione
Classificazione:	025.04 - ATTIVITÀ DELLE BIBLIOTECHE.ARCHIVIAZIONE DELL'INFORMAZIONE E SISTEMI DI RECUPERO
Paese di pubblicazione:	IT
Lingua di pubblicazione:	ITA
Localizzazioni:	● AN0092 - Biblioteca comunale Planettiana - Jesi - AN
	● AQ0047 - Biblioteca provinciale Salvatore Tommasi - L'Aquila - AQ
	● BA0018 - Biblioteca nazionale Sagarriga Visconti-Volpi - Bari - BA
	● BG0025 - Biblioteca comunale - Sezione Moderna e Scientifica Ciro Caversazza - Bergamo - BG
	● BO0199 - Biblioteca comunale - Imola - BO
	● BO0210 - Biblioteca comunale Giulio Cesare Croce - San Giovanni in Persiceto - BO
	● BO0287 - Biblioteca comunale centrale di Palazzo Montanari - Bologna - BO
	● BO0304 - Biblioteca comunale dell'Archiginnasio - Bologna - BO
	● BO0471 - Biblioteca del Dipartimento di Scienze Statistiche P.

Home
Indietro
E. Altri

Aiuto
Avanti
Accessi

Favorites History Search Page Holder

ICCU Home Page

Address: http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame

Back Forward Stop Refresh Home Favorites History Search AutoFill Larger Smaller Print Mail Preferences

Live Home Page · Apple Computer · Apple Support · Apple Store · Microsoft MacTopia · Office for Macintosh

ICCU

► Ricerca Base

Istituto Centrale per il Catalogo Unico
Indice SBN

Risultati della Ricerca Base

(*Legenda*: ✱ *testo completo del titolo nel caso di titoli molto lunghi*)

Criteri usati nella ricerca:
Soggetto—*Documentazione - Automazione*

Risultato della ricerca : 38 documenti

No.			Visualizza Tutto Visualizza Selezione Raffina la Ricerca ▲ ▲ ✱
☐ (1/38)			[Monografia] - Bettelli, Oscar - *Dati, relazioni & associazioni / Oscar Bettelli* - Milano - ☐WD 1-1991 (ITNCCUAQ N0001308)
☐ (2/38)			[Monografia] - *Advances in intelligent retrieval : Informatics 8 : proceedings of a conference jointly sponsored by Aslib, the Aslib Informatics Group, and the Information Retrieval* ✱ - London - 1985 (ITNCCUBVE000247 1)
☐ (3/38)			[Monografia] - Ashoford, John - *Text retrieval and document databases / John Ashford, Peter Willet* - Lund - ☐WD 1- 1988 (ITNCCUBVE0003679)
☐ (4/38)			[Monografia] - Convey, John - *Online information retrieval : an introductory manual to principles and practice / John Convey* - London - 1989 (ITNCCUBVE0002719)
☐ (5/38)			[Monografia] - Salton, Gerard - *Introduction to modern information retrieval / Gerard Salton, Michael J. McGill* - New York [etc.] - ☐WD 1-1983 (ITNCCUBVE0019968)
☐ (6/38)			[Monografia] - Cox, John - *Keyguide to information sources in online and CD-ROM database searching / John Cox* - London [etc.] - 1991 (ITNCCUBVE0021465)
☐			[Monografia] - *Manual of online search strategies / edited by C. J. Armstrong and J. A. Large* - Aldershot - ☐WD 1-1992 (ITNCCUBVE0022240)

Home

Indietro

E-Mail

Aiuto

Avanti

Accessi

Favorites · History · Search · Page Holder

Visualizza solo i riferimenti selection

ICCU Home Page

Back Forward Stop Refresh Home Favorites History Search AutoFill Larger Smaller Print Mail Preferences

Address: http://opac.sbn.it/cgi-bin/IccuForm.pl?form=WebFrame

Live Home Page | Apple Computer | Apple Support | Apple Store | Microsoft MacTopia | Office for Macintosh

Istituto Centrale per il Catalogo Unico
Indice SBN

Modulo di Ricerca Base

[Legenda ● *Lista delle frasi* ● *Lista delle parole*]

▸ Ricerca Base

Autore ● ●	baldacci	Cerca
Titolo ● ●	biblioteche	
Soggetto ● ●		
Classificazione:		
Numero ● ●		
Descrizione ● ●		Cerca
Tutti i campi		

Esegui la ricerca

ICCU Home Page

Back Forward Stop Refresh Home Favorites History Search AutoFill Larger Smaller Print Mail Preferences

Address: http://opac.sbn.it/cgi-bin/iccuform.pl?form=WebFrame Go

Live Home Page · Apple Computer · Apple Support · Apple Store · Microsoft MacTopia · Office for Macintosh

ICCU
ICCUiccu

▶ Ricerca Base

Favorites \ History \ Search \ Page Holder

Home
Aiuto
Indietro
Avanti
E Mail
Accedi

Istituto Centrale per il Catalogo Unico
Indice SBN

Risultati della Ricerca Base

(**Legenda**: ★ *Sono completi i dati in cui vier visualizzato in Josept*)

Criteri usati nella ricerca:
Autore=Baldacci Titolo=biblioteche

Risultato della ricerca: 26 documenti

Visualizza Tutto   Visualizza Selezione   Raffina la Ricerca ▶   ▲   ▼

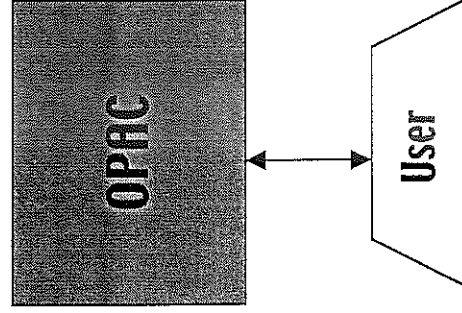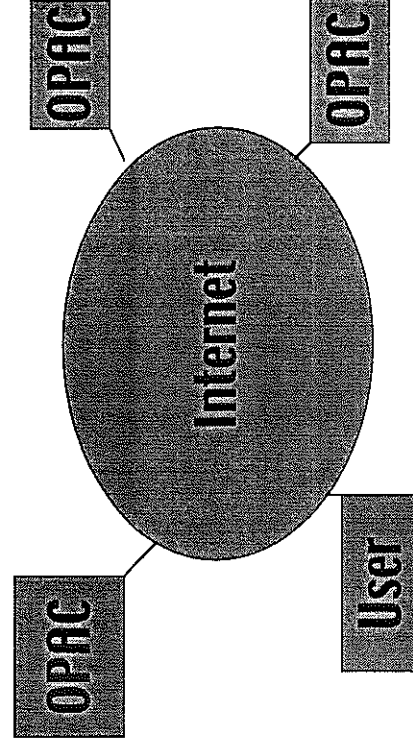| No.	
(1/26)	[Monografia] - **Baldacci, Maria Bruna** - *Informatica e biblioteche : automazione dei sistemi informativi bibliotecari* / Maria Bruna Baldacci, Renzo Sprugnoli - Roma - 1986 (ITNCCCUCFIV0013166)
(2/26)	[Monografia] - **Baldacci, Maria Bruna** - 2 : *La base dei dati dell'ATLAS* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli e N. Crevani ... [et al.] - [Pisa] - 1986 (ITNCCUCFIV0040944)
(3/26)	[Monografia] - **Baldacci, Maria Bruna** - *Automazione delle biblioteche con elaboratori mini e micro* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli - [Pisa] - [1986?] - (ITNCCUCFIV0040946)
(4/26)	[Monografia] - [3 : *Introduzione allo studio delle geocarte nautiche di tipo medievale e la raccolta della Biblioteca comunale di Siena* / a cura di Osvaldo Baldacci - Firenze - 1990 (ITNCCUCFIV0161335)
(5/26)	[Monografia] - **Baldacci, Maria Bruna** - *L'automazione delle biblioteche universitarie* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli - [Ospedaletto - Pisa] (ITNCCUCFIV0139736)
(6/26)	[Monografia] - **Baldacci, Maria Bruna** - *ATLAS : sistema per l'automazione delle biblioteche* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli, Università degli studi di Pisa, Istituto di elaborazione dell'... [Ospedaletto - Pisa] (ITNCCUCFIV0065821)
(7/26)	[Monografia] - **Baldacci, Maria Bruna** - *ATLAS : sistema per l'automazione delle biblioteche : Registrazione fuori linea dei dati bibliografici : manuale per l'utente* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli ★ Ospedaletto - Pisa (ITNCCUCFIV0065875)
(8/26)	[Monografia] - **Baldacci, Maria Bruna** - *ATLAS : sistema per l'automazione delle biblioteche* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli ... [al.] - Pisa (ITNCCUCFIV0065889)
(9/26)	[Monografia] - **Baldacci, Maria Bruna** - *Nuove accessioni : manuale per l'utente* / M. B. Baldacci, O. G. Mancino, R. Sprugnoli e N. Crevani ... [et al.] - Ospedaletto - Pisa (ITNCCUCFIV0065904)

Internet zone

From on-line catalogs to distributed internet resources [1]

- Original on-line technology

- Users operate on "their" OPAC through a "dedicated" connection line which presents the OPAC interface on his/her terminal

- To-day "on-line" technology is a very different one

- Users can operate on any of the OPACs connected to the communication network [typically the Internet]. Communication happens through a network protocol

What is a protocol ?

- In the network communication context, a **protocol** is an agreed-upon format for transmitting data between two different devices

- An **application protocol** is designed to be applied in a specific operative context

From on-line catalogs to distributed internet resources   [3]

■ The most widely used network protocol:

http – hyper-text transfer protocol
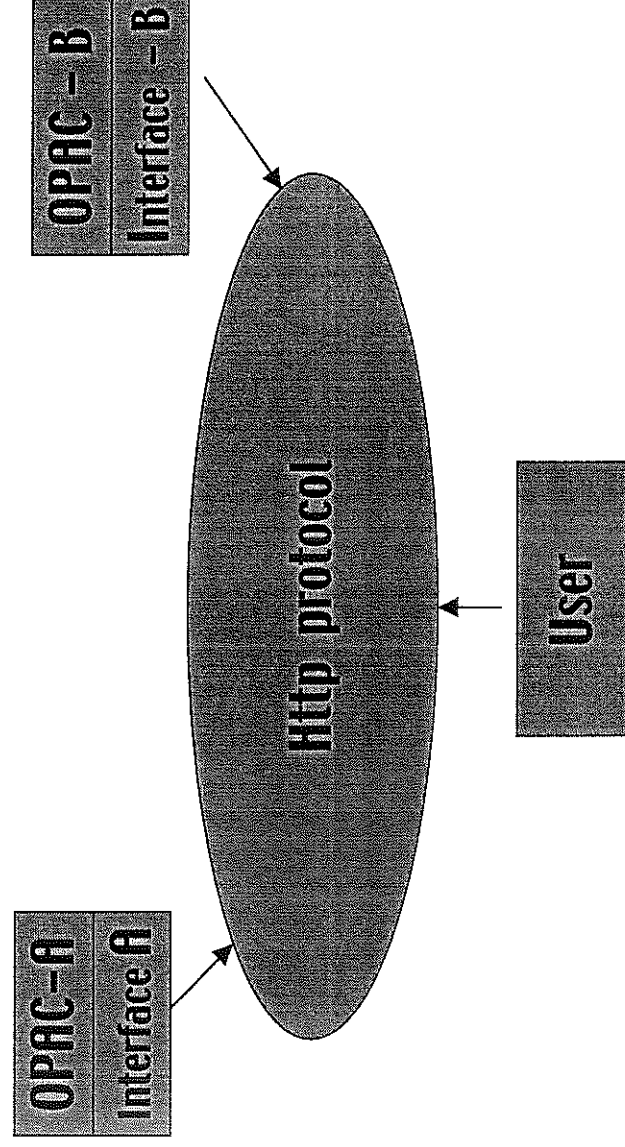
Http allows to access any information on the Web [World Wide Web]

The Web is a global information system deployed on the Internet

From on-line catalogs to distributed internet resources   [4]

Http:// *[OPAC address]*

■ It allow users

- to connect with any remote OPAC on the Internet

- to operate on that OPAC through its interface

Searching on distributed internet resources [1]

OPAC – B	
Interface – B	

OPAC–A	
Interface A	

Http protocol

User

Users have to operate with different interfaces when searching on different OPACs
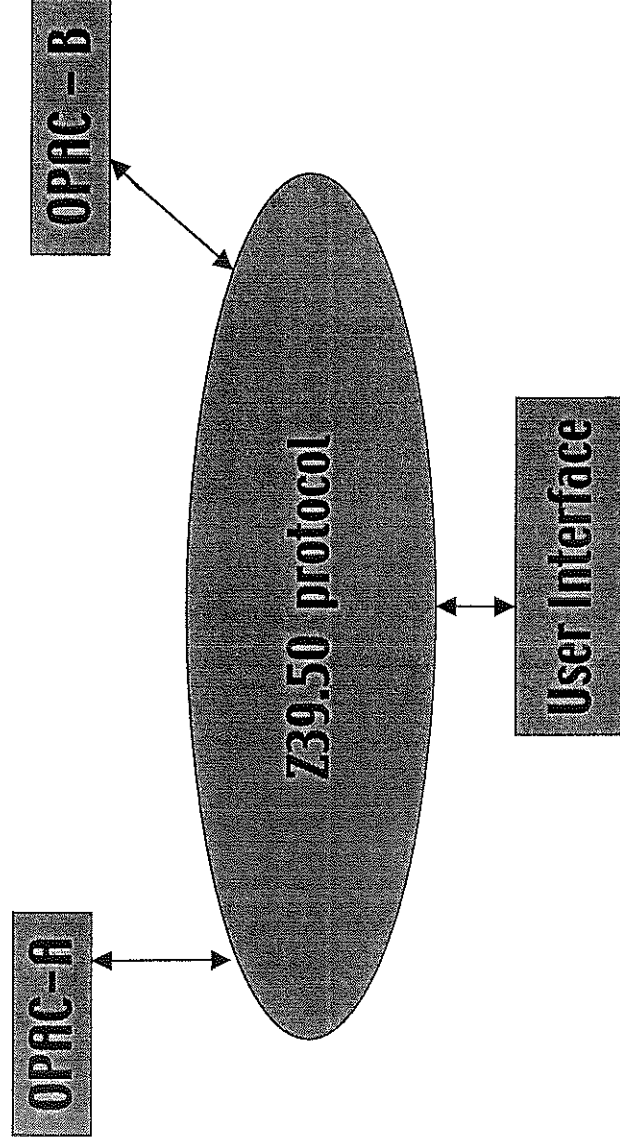
Each OPAC has its own user interface, possibly different from the interfaces of other OPACs because of:

   | Different style

   | Different language of instructions to users

   | Different names given to the same bibliographic fields

   | Different fields used as 'access points'

   | Different types of documents in the catalog

Searching on distributed internet resources   [3]

■ **Z39.50**

■ The standard application protocol
designed for information search and
retrieval over the Internet

Searching on distributed internet resources [4]

Any user can operate on any OPAC with the interface her/he is accustomed with

Interoperability [1]

■ How Z39.50 works:

The protocol states rules which allow two different systems, i.e.,

- a client [a user], which requests information, and
- a server [an OPAC], which holds searchable bibliographic records

to interchange understandable messages to perform information retrieval operations

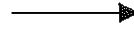| Client | → | Translation | → | Protocol Messages | → | Translation | → | Server |

Interoperability [2]

- Protocol messages standardize:

  - How the server is connected
  - How the user queries are represented
  - How returned bibliographic records are formatted [e.g.: Unimarc]

Interoperability [3]

- Z39.50 supports searching on a very large set of access points *[authors, title, subtitle, series...publisher, place of publication...]*, whereas OPACs generally support only a subset of them.

- Z39.50 supports searching on descriptive elements of different document types, having different access points *[modern books, manuscripts, museum objects...]*

- Z39.50 allows many options as to how access points are to be searched *[by a whole string, by one or more words, by truncated words, etc.]*

→

Searching PROFILES

Interoperability   [4]

- A **profile** defines a set of access points and associate search modes which complying systems agree to implement in order to assure **interoperability**

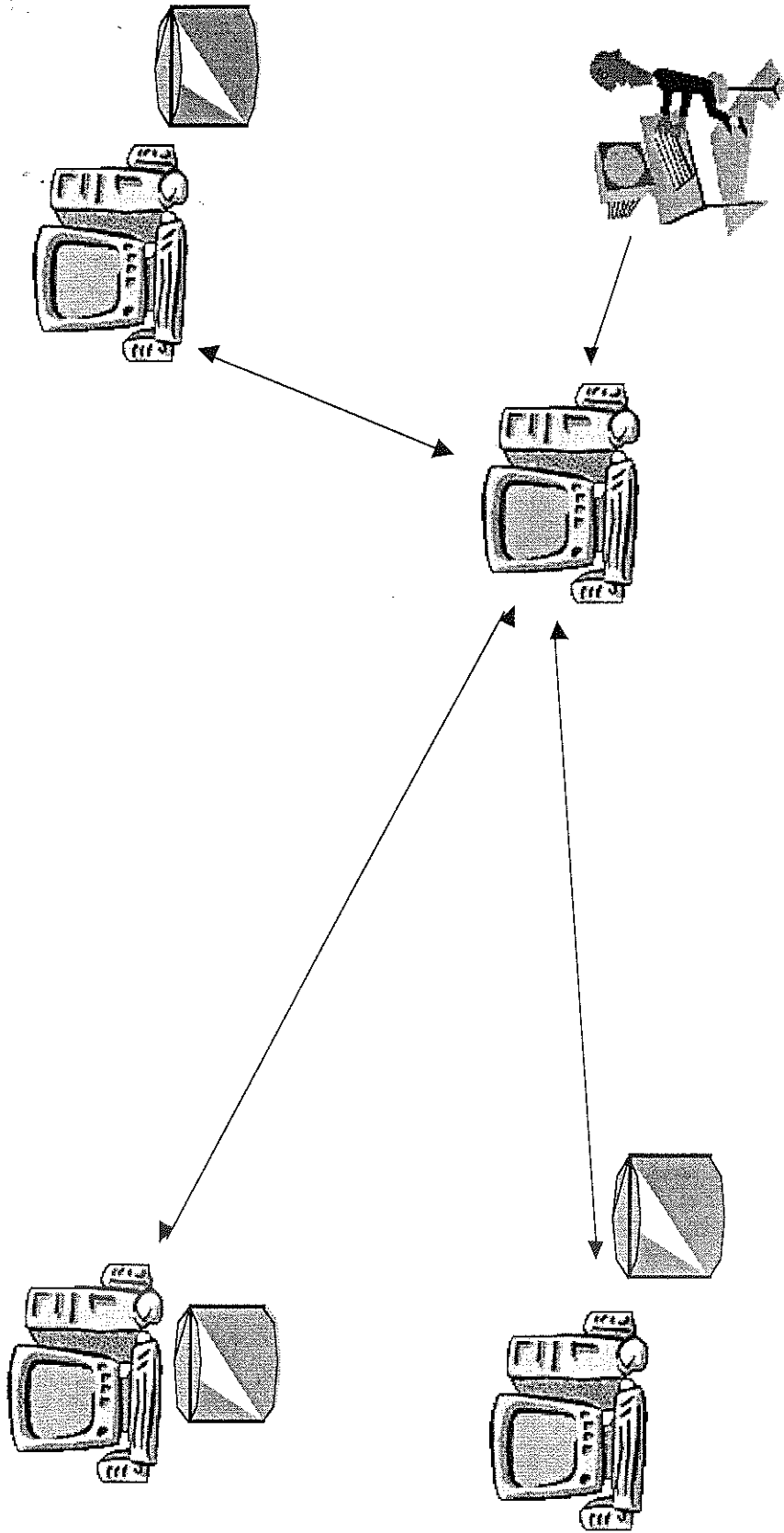- **ONE Profile: the profile agreed by OPACs in Europe**

- **CIMI Profile: the profile agreed for Cultural Heritage Information applications**

The virtual catalog

■ In spite of their possible differences, catalogs searchable with an agreed profile of the Z39.50 protocol realize a unique virtual catalog that any user can search using the interface he/she is familiar with.

Il catalogo virtuale

Indietro Avanti Ricarica Home Ricerca Naviga Stampa Sicurezza Stop

Segnalibri Indirizzo: 52font%253e%2520contiene%2520%2520%253cFONT%2520color%253d%2522red%2522%2522%253epicasso%253c%2522font%253e

Internet Ricerca Novità e curios

Istituto Centrale per il Catalogo Unico
Indice SBN On-Line

Help | Altri Cataloghi | Indice SBN | Ricerca Semplice | Ricerca Avanzata

Catalogo: Fondazione Peggy Guggenheim
Ricerca effettuata: Autore contiene picasso
Risultati: da 1 a 7 di 7

Risultati Sintetico Analitico SUTRS

Titolo	Pipe, verre, bouteille de Vieux Marc/ Pipa, bicchiere, bottiglia Vieux Marc/ Pipe, glass, bottle of Vieux Marc/ Lacerba
Autore	Picasso Pablo 1881/ 1973
Oggetto	Dipinto
Datazione	XX 1914 1914

File Modifica Visualizza Vai Finestra ?

Indietro   Avanti   Ricarica   Home   Ricerca   Naviga   Stampa   Sicurezza   Stop

Segnalibri   Indirizzo: 52/font%253e%2520contiene%2520color%253d%2522red%2522%253epicasso%253c%252ffont%253e

Internet   Ricerca   Novità e curios

Oggetto   Dipinto

Titolo   Pipe, verre, bouteille de Vieux Marc/ Pipa, bicchiere, bottiglia Vieux Marc/
               Pipe, verre, bouteille de Vieux Marc/ Pipa, bicchiere, bottiglia Vieux Marc/
               Pipe, glass, bottle of Vieux Marc/ Lacerba

Documento: Eseguito

Hybrid Libraries [1]

- The hybrid library is on the continuum between the "conventional" and digital library where electronic and paper-based information sources are used alongside each other.

Hybrid Libraries [2]

A very simple model of hybrid library:

Electronic documents are the electronic version of paper-documents

e.g.:

- Electronic version of paper journals
- Electronic journals

Hybrid Libraries   [3]

However, documents in the DL may be of totally new types, that have no print-on-paper equivalent

On-going projects have the objective of allowing

integrated access to different types of documents and media

Hybrid Libraries   *[4]*

Hybrid libraries have a number of significant similarities with **digital libraries**, thus we may turn back to our original objective of understanding what a digital library is…………

Digital versus traditional libraries [1]

- "A digital library can be regarded as a system that performs or supports at least the functions of a library in a context of distributed networked collections"

- *At this point, the above definition should have become a very intuitive one, but perhaps it ties us too much to traditional concepts of library functions and technologies whereas substantial changes are implied when we introduce the idea of digital libraries*

- *In fact, systems managing libraries of digital documents must be capable of handling many issues unknown to traditional library systems: security, version control, copy rights, etc.*

Digital versus traditional libraries [1]

▪ Exploring the differences

- Traditional libraries are stable and slowly evolving; digital libraries are highly dynamic, ephemeral and versioned

- Traditional (academic) libraries hold objects which are authored through a ponderous publishing stream; digital libraries allow anyone to publish in a lightweight way documents of very different types

- In traditional libraries the objects are physically and logically co-controlled; in digital libraries the physical and logical organisations can be separated (virtual collections)

Digital versus traditional libraries [2]

- The tradition of public libraries is universal access and free; digital libraries could be similar in this regard, or digital libraries could support rich layers of access control and management of terms and conditions

- In traditional libraries structured text queries (and some browsing) are used to aid intellectual access; in digital libraries complex interactions of query, navigation/browsing, and social filtering can be used

- Traditional libraries are based upon a model of one-way search (a consumer looking for an object); digital libraries support symmetric search (consumer looking for an object and producer of the object looking for a consumer)

Digital versus traditional libraries   [3]

There are a lot of differences between

the digital library as a simulation of a traditional library and

a digital library with entirely new modes of supporting the life cycle of information creation, distribution, use, and preservation

- The new physical state of documents not only affects all the procedures operating upon documents as physical objects [collections organisation, collocation, loan, etc] but also the administrative procedures concerning publishers and users, whereas

- the new physical state has a lower impact on procedures treating the information contents of documents [i.e.,cataloguing, indexing classifying]

Digital versus traditional libraries [4]

- Even though the semantic contents of digital documents can be treated with concepts and technologies of "traditional" information search and retrieval techniques, documents in the DL may be of totally new types, that have no print-on-paper equivalent; therefore, new mechanisms are required for

- cataloguing and discovering network information resources
- managing terms and conditions to access documents
- payments of document copies, etc.

New terms for new concepts [1]

The need to distinguish between old and new concepts and related technologies is revealed by the appearance of new terms

■ Digital object means any of the digital document types, be it a text, a computer program, a video, a figure......

■ Electronic resource means an Internet accessible digital object

■ Repository means the "container" of digital objects; it provides the array of services that make digital objects accessible

New terms for new concepts [2]

- Metadata means "structured data about data" where the term "data" includes an almost limitless spectrum of possibilities: from human generated textual description of a resource to machine-generated data useful only to software applications.

  - *In the traditional library world, "metadata" is not other than a record of cataloguing data [i.e, the set of bibliographic elements and their associated values that describes/represents a document]. In the Internet world, however, there has been the need of extending the concept of "cataloguing data" to objects different from the traditional library objects.*

New terms for new concepts   *[3]*

- Dublin Core Metadata Set. This term indicates a set of elements that a broad interdisciplinary consensus judged to be capable to represent/describe any Internet accessible resource, independently of the medium of the resource

- *Considerable attention has been invested in making the Dublin Core sufficiently flexible to represent resources (and relationships among resources) that are both digital and exist in traditional formats as well. A metadata record can thus represent a digital object directly accessible through the network as well as an OPAC whose records are only searchable through the OPAC Interface*

- *Dublin Core metadata is specifically intended to support resource discovery*

New terms for new concepts [4]

- Resource discovery means searching for those Internet accessible electronic resources potentially useful for satisfying information needs

- *In the traditional library world, "resource discovery" doesn't mean other than "information search and retrieve". In the Internet world, however, there has been the need to limit this concept to the case where an information resource is discovered, but retrieval is only possible with further steps0*

New functions

[1]

- Another significant example of the transformations induced by digital libraries is how traditional library functions such as acquisition, location, loan, etc. are absorbed into one of the major components of a digital library, i.e., the *repository* of digital objects

- On the contrary, the repository enriches the access services with a set of new functions such as *authentication, security, protection of intellectual property, payment,* etc. required by the different nature of documents

52

New functions [2]

■ Also the user interfaces must provide new functions: for example, interfaces have to

▌ allow users to access digital libraries playing different roles - for example, as "authors" who provide documents, as "information seekers", or as "information managers"

▌ provide functions to "manipulate" the digital objects – for example, to overview or browse the entire document, select and download some parts of its contents, etc.

53

Final remarks

- Although digital libraries allow users to search with interface operations similar to those in traditional OPACs, they are characterized by a lot of new functions and services

- The new functions and the new services are supported by the new architectures and technologies that underlie digital libraries