



# Le biblioteche digitali: creazione, uso e condivisione degli spazi informativi delle comunità di ricerca

Donatella Castelli  
Pasquale Pagano

ISTI-CNR  
Pisa

# Contenuto



- ❖ Che cosa è una biblioteca digitale
- ❖ Breve storia delle biblioteche digitali
- ❖ Il sistema per la gestione di BD OpenDLib
- ❖ Una sperimentazione di OpenDLib nell'ambito del progetto ARTE
  - Introduzione
  - Dimostrazione

# Che cosa è una biblioteca digitale?



“ Una istituzione che fornisce (almeno) le funzioni di una biblioteca nel contesto di collezioni distribuite in rete di oggetti informativi in formato digitale”

Nicholas Belkin

Eight DELOS Workshop

Stockholm, 1998

# Oggetti digitali



❖ Non solamente descrizioni di oggetti (metadati) ma anche

**oggetti informativi** (testi, video, suoni, immagini 3D, ..... dati, programmi, ... mappe, .... )

❖ Non solo un **catalogo** ...

ma anche un **deposito**

# Un biblioteca digitale è quindi l'analogo di



- ❖ una biblioteca
- ❖ un museo
- ❖ un archivio
- ❖ una video, audio, foto - teca
- ❖ ...
- ❖ un centro di raccolta di dati geografici
- ❖ ...

# Le funzioni di una biblioteca



- ❖ selezionare, collezionare, preservare, organizzare, rappresentare, fornire accesso a, assicurare la conoscenza di, e disseminare oggetti **informativi**
- ❖ mediare e supportare l'interazione tra gli utenti dell'informazione e gli oggetti informativi

# Biblioteca come mezzo di comunicazione



- ❖ Creatori e fruitori di conoscenze “comunicano” attraverso la biblioteca
  - **Indiretta**: i documenti vengono raccolti per gli utenti futuri
  - **Mediata**: la biblioteca seleziona, rappresenta sinteticamente i documenti, ne organizza gli strumenti di reperimento

# Le origini



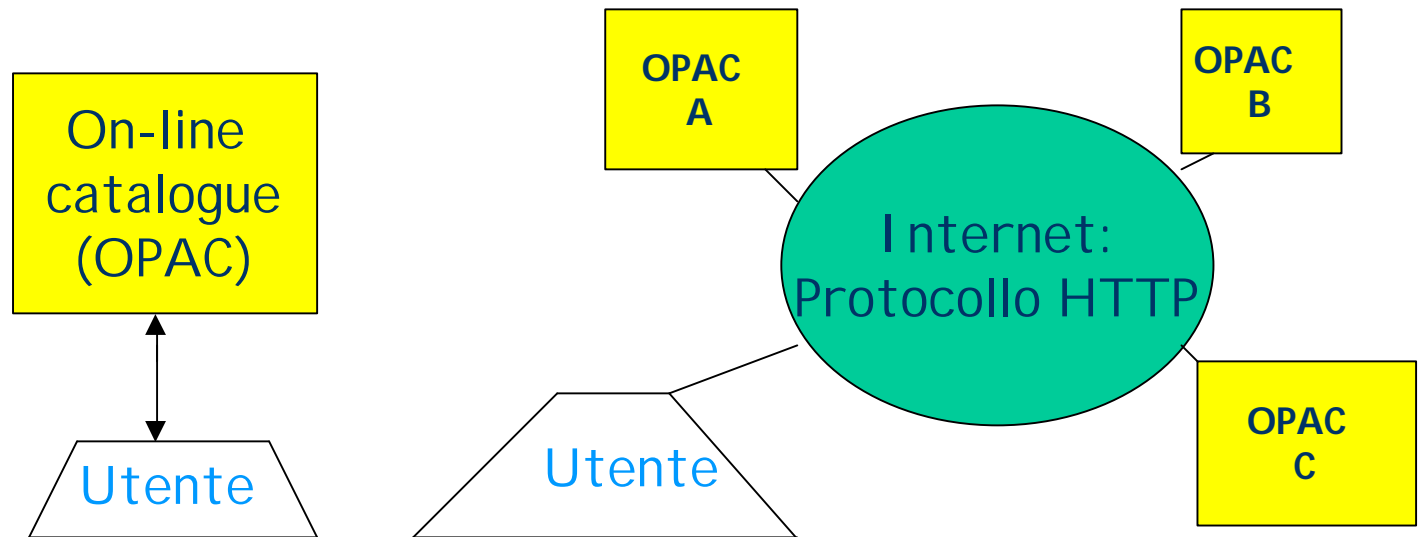
❖ I sistemi bibliotecari

❖ II Web

# Sistemi bibliotecari del passato

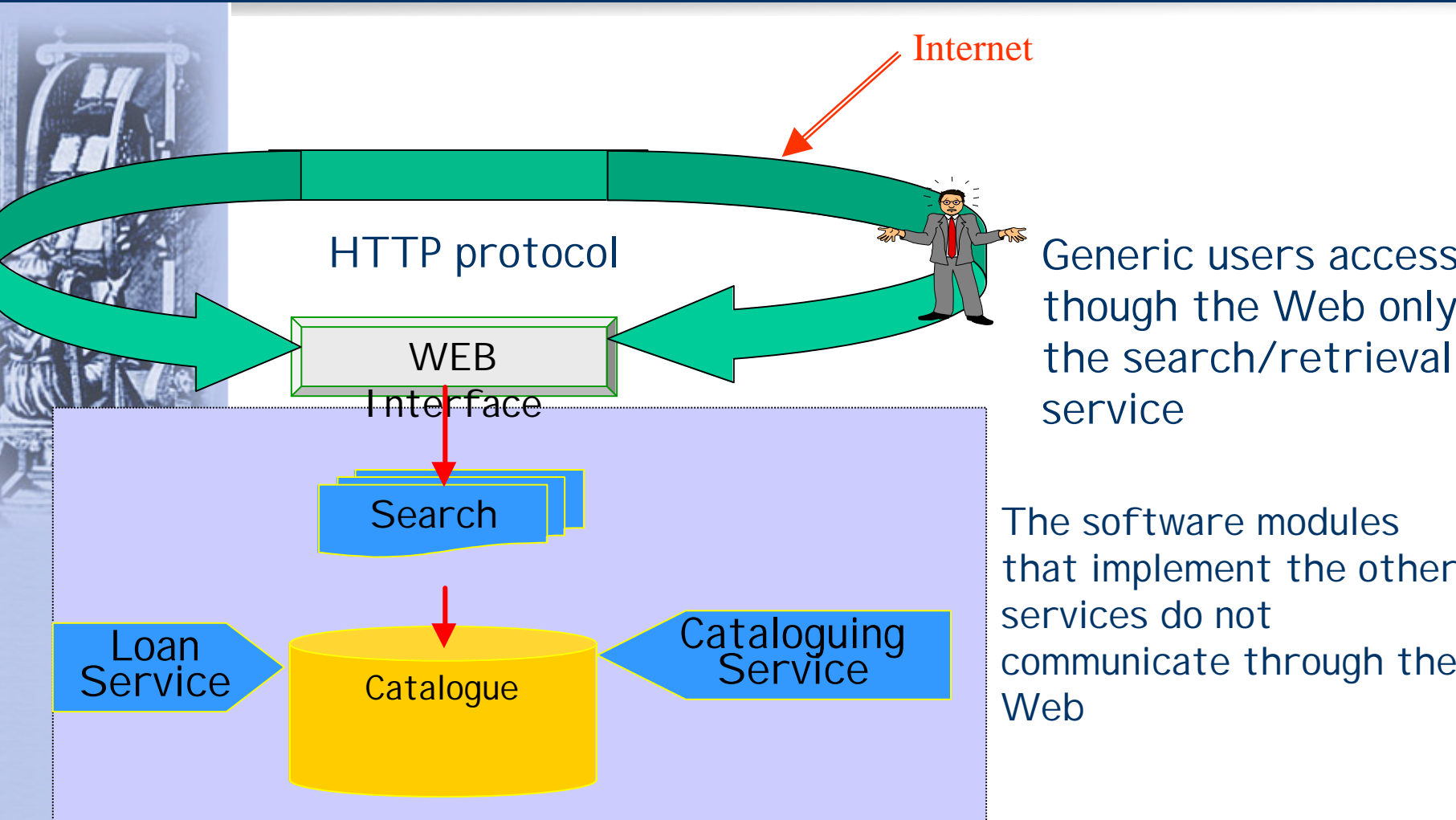


Da una comunicazione  
"diretta" (on-line) ....



.... ad una comunicazione su  
WEB: protocollo HTTP

# Accesso via Web agli OPAC



# Accesso via Web agli OPAC



- ❖ Ciascun catalogo è accessibile attraverso la propria interfaccia
- ❖ Le interfacce utente differiscono per:
  - punti di accesso
  - nomi dei punti di accesso
  - lingua
  - grafica
  - .....
- ❖ Gli utenti devono avere familiarità con molte interfacce
- ❖ Non si possono fare ricerche su più archivi contemporaneamente

An example of user interface

**Biblis WWW - Danmarks Biblioteksskoles Bibliotek**

Address: [http://ax.db.dk/biblis/biblis\\_u.htm](http://ax.db.dk/biblis/biblis_u.htm)

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

**Danmarks Biblioteksskoles Bibliotek**

**Udvidet søgning**

Vælg database: **Alle samlinger**

Indtast søgeord. Vælg felt og operatører. Trunkering med \$

som: **Titelord** mellemrum = **OR**

**AND**

som: **Ophav/Personer** mellemrum = **OR**

Årstal: **Alle** Sprog: **Alle**

Max. antal poster pr. side: **10** Vis: **Som Hypertekst** **Søg** **Slet alt**

**Danmarks Biblioteksskoles Bibliotek**  
Adresse: Birketinget 6, DK 2300 København S  
Tlf: (+45)3258 6066 lok 444/442 - Direkte: (+45)3258 7077  
Fax: (+45)3284 0201  
Email: [dbilaan@db.dk](mailto:dbilaan@db.dk)  
Åbent: man-tor 9-16, fre 9-15

Ansvarlig for denne side: **Bo Kristensen** ([bkr@db.dk](mailto:bkr@db.dk)),  
Sidst opdateret: 21-09-99

Internet zone

# Una soluzione: Z39.50

Protocollo standard per la ricerca e il recupero di informazioni

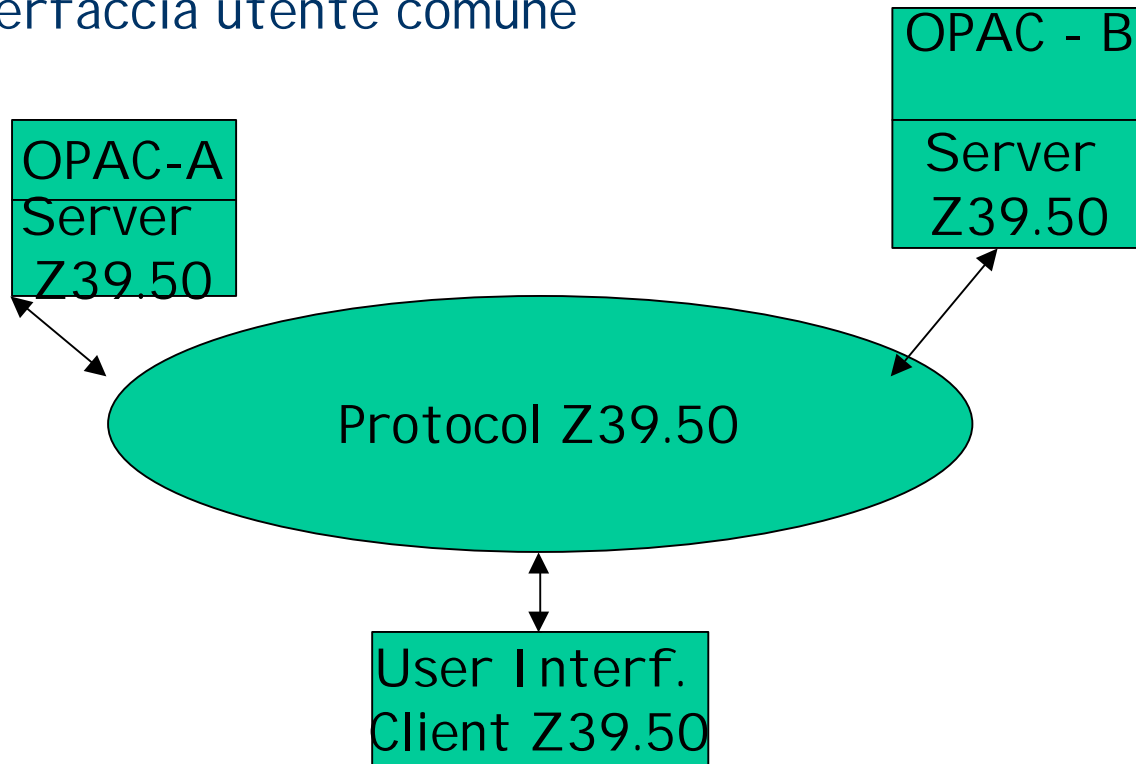
It establishes the rules that regulate the communication between the clients and the servers

(automatic catalogues)



# Z39.50

Permette la ricerca e il recupero di informazioni bibliografiche da differenti OPACs distribuiti attraverso la formulazione di una singola interrogazione attraverso una interfaccia utente comune



# Biblioteca virtuale



- ❖ Accesso in parallelo a OPAC Z39.50 selezionati
- ❖ Intefaccia utente comune
- ❖ WEB access with mapping HTTP-Z39.50 (and vice-versa)
- ❖ “Traditional” bibliographic search [ based on OPACs]

# Uso di Z39.50



- ❖ The use of Z39.50 is limited to catalographic resources
- ❖ No project implements the protocol functions for the management of electronic documents

# The Web



90-ies:

global access to information resources

- Different types of resources, stored on Internet distributed servers and accessible through the WWW-World Wide Web
- The Web allows to access a specific resource by specifying the network address (URL)

# Search engines



- ❖ Index the words in the Web pages
- ❖ Allow the resource discovery (without explicitly knowing the address)
- ❖ Implement their own resource selection policies and their own indexing techniques
- ❖ Offer statistics/probabilistic search
- ❖ Return Web addresses

# Resource discovery on the Web



"Finding relevant information on the World Wide Web has become increasingly problematic due to the explosive growth of networked resources. Current Web indexing evolved rapidly to fill the demand for resource discovery tools, but that indexing, while useful, is a poor substitute for richer varieties of resource description."

Dublin Core Metadata Initiative  
<<http://www.ietf.org/rfc/rfc2413.txt>>



- ❖ There is the need of some form of cataloguing of the resources available on Internet in order to achieve a good balance between “recall” and “precision”
- ❖ Descriptive rules must be suitable for all the types of information resources

## Dublin Core Metadata Format



- ❖ “Dublin Core metadata is used to supplement existing methods for searching and indexing Web based metadata, regardless of whether the corresponding resource is an electronic document or a "real" physical object.”
- ❖ “Dublin Core metadata provides card catalog-like definitions for defining the properties of objects for Web-based resource discovery systems.”

# What is the DC Metadata Element Set?



❖ “It is a set of **16** descriptive semantic definitions. It represents a core set of elements likely to be useful across a broad range of vertical industries and disciplines of study”

- |             |                   |
|-------------|-------------------|
| ? Title     | - Creator         |
| ? Subject   | - Description     |
| ? Publisher | - Contributor     |
| ? Date      | - Type            |
| ? Format    | - Identifier      |
| ? Source    | - Language        |
| ? Relation  | - Coverage        |
| ? Rights    | - <b>Audience</b> |

# First generation DLs: in the US



## 1994-1998 Digital Library Initiative – Phase I

### Funded by:

- National Science Foundation(NSF)
- Department of Defense Advanced Research Project Agency (DARPA)
- National Aeronautics and Space Administration (NASA)

### Obiettivo

“The focus is to dramatically advance the means to collect, store, and organize information in digital forms, and make it available for searching, retrieval, and processing via communication networks – all in user friendly ways”

# First generation DLs: in Europe

❖ 1996

ERCIM\* Technical Reference Digital Library



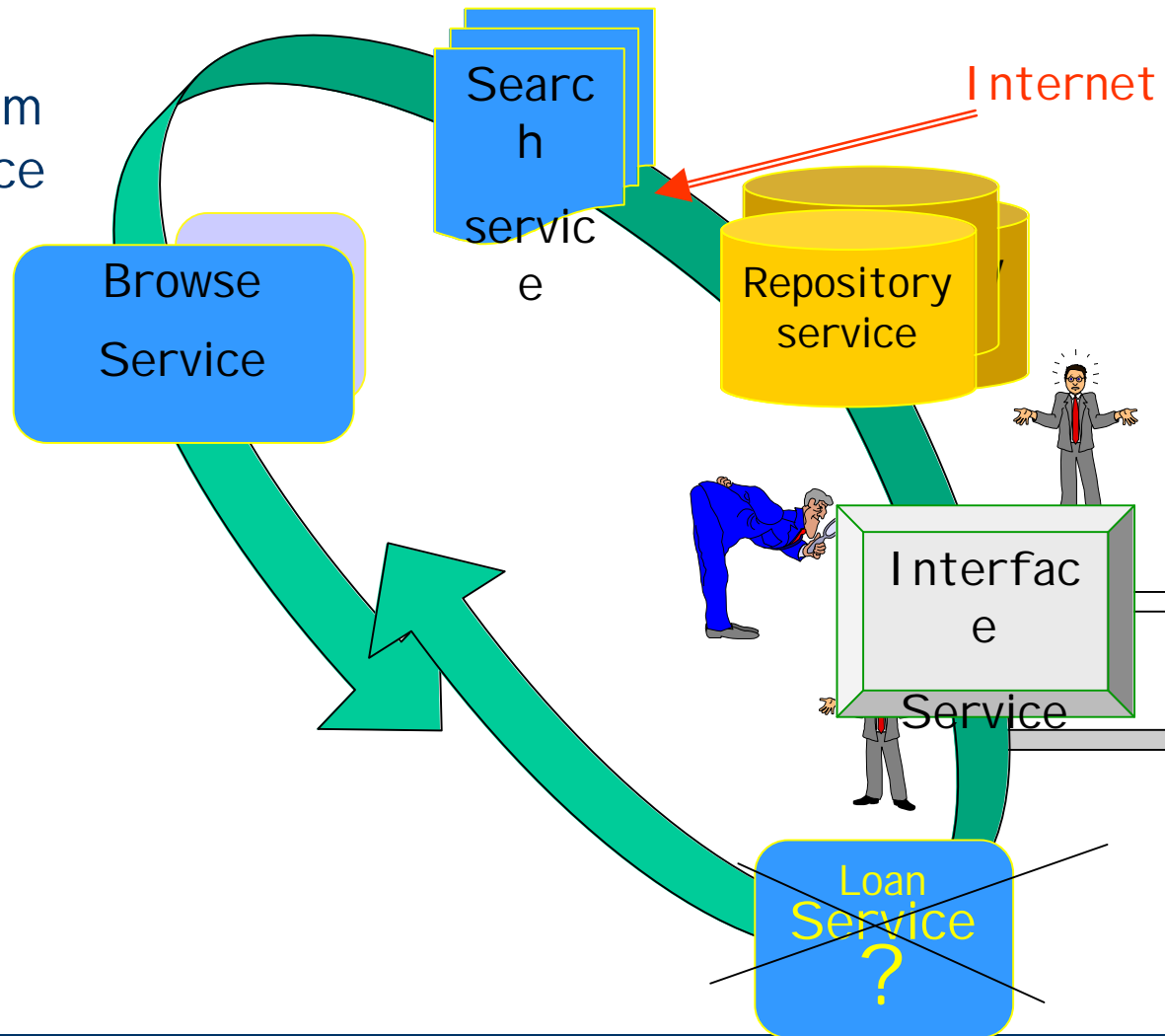
\*ERCIM =

European Consortium for Informatics  
and Applied Mathematics

# Servizi distribuiti

Users access the system through a Web interface

The services are distributed on the Internet.  
They communicate through an established protocol.



# Informazione multimediale

**CMU Informedia Video Library**

File Edit Navigate Options Data Window Help Comments!

Search for ANY of:  
fires floods earthquakes hurricanes tornadoes

Clear All History


Search Options Results Options

9 of 287 results: any of "fires floods earthquakes hurri  
tornadoes."

Click on a word to highlight it in yellow.

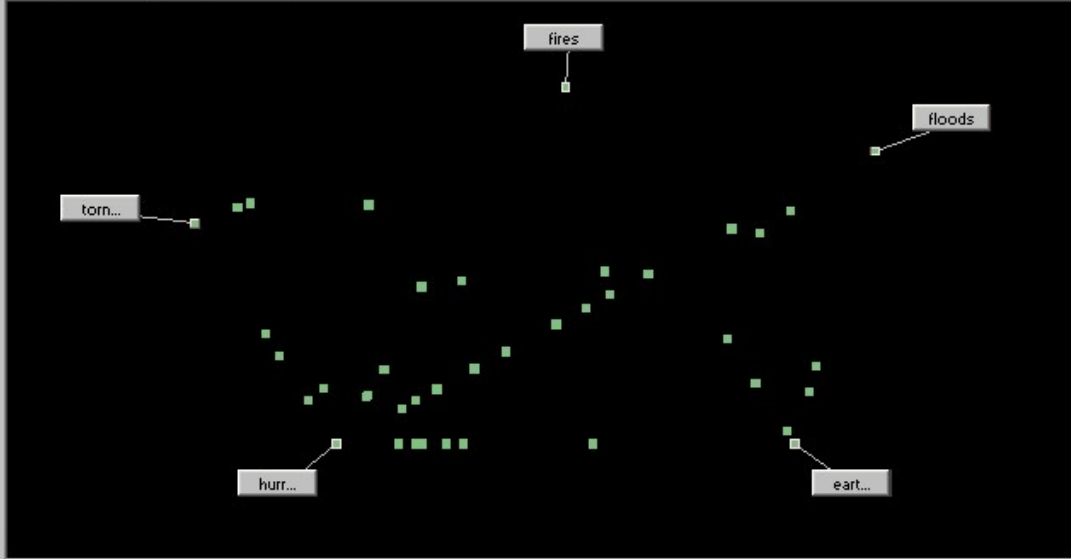
Prev. Page Next Page Go to Page... Visualize All...

Search Results (Page 1 of 32)



Visualization of search results set containing 287 documents

VIBE Timeline Map Topics



fires floods  
torn... hurr... eart...

- fires
- floods
- earthquakes
- hurricanes
- tornadoes

Display lines toggled. 287 visible documents.

Color Code By:  Minimum Value  Average Value  Maximum Value

Relevance (All: 0 - 100)  Size-Code  Color-Code

Within >

Date (All: 07/01/99 - 09/30/99)  Size-Code  Color-Code

Size (All: 0:02 - 30:30)  Size-Code  Color-Code

Map Hits (All: 1 - 125)  Size-Code  Color-Code

Topic Hits (All: 1 - 252)  Size-Code  Color-Code

Ellipse Re-Plot

Connect words

Show Doc. Info

Show Details

# The Informedia DL: example 2

CMU Informedia Video Library

Edit Navigate Options Data Window Help Comments!

Search for ANY of:  
fires floods earthquakes hurricanes tornadoes


Clear All History

Search Options Results Options

of 287 results: any of "fires floods earthquakes hurricanes tornadoes."  
*Click on a word to highlight it in yellow.*

Previous Page Next Page Go to Page... Visualize All...

Search Results (Page 1 of 32)




Visualization of search results set containing 287 documents

VIBE Timeline Map Topics

Search Zoom Hand Map Earth Refresh

Countries  States



TURKEY: 46/48 hits active

Document with no map associations.

Visible Invisible Inactive

<none>  
CHINA  
MEXICO  
TAMWAN  
TURKEY

Deactivate selected locations

Show Doc. Info  
Show Details

58 visible documents; 4 of 5 words ignored; 25 of 55 locations ignored.

Color Code By:  Minimum Value  Average Value  Maximum Value

Relevance (All: 0 - 100)  Size-Code  Color-Code

Within > [Progress bar]

Date (All: 07/01/99 - 09/30/99)  Size-Code  Color-Code

Within > [Progress bar]

Size (All: 0:02 - 30:30)  Size-Code  Color-Code

Map Hits (All: 1 - 125); Shown: 3.2 to 125  Size-Code  Color-Code

Within > [Progress bar]

Topic Hits (All: 1 - 253)  Size-Code  Color-Code

# From the first to the second generation DLs



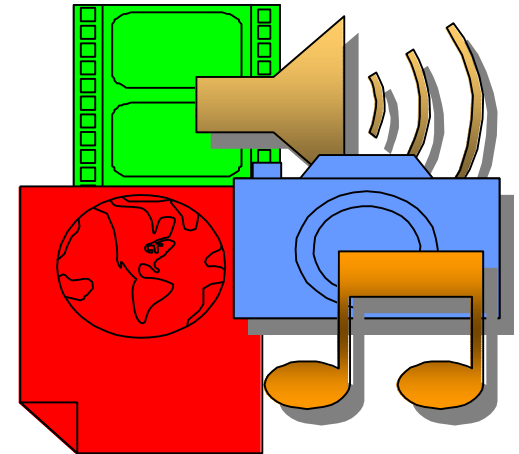
Rispetto alle biblioteche tradizionali le biblioteche digitali consentono una comunicazione più

- ampia
  - rapida
  - economica
- ❖ Gli oggetti digitali consentono altre forme di comunicazione
- ❖ Altre funzionalità possono essere realizzate creando del software opportuno

# New types of digital objects



- ❖ Multimedia
- ❖ Structured
- ❖ Annotated
- ❖ Multilingual



The new document types enrich the possible forms of remote collaboration among the members of a community of interest

# Multimedia documents

- ❖ video e lucidi di seminari, lezioni, presentazioni,
- ❖ corsi di formazione
- ❖ presentazione di progetti
- ❖ dimostrazioni

The screenshot shows a Microsoft Internet Explorer browser window displaying a multimedia document. The address bar shows the URL: <http://video server.id.pscn.it:2001/andilotti-1-MS.html>. The main content area features a video player showing a man in a suit speaking. To the right of the video is a diagram titled "COMUNICAZIONI NEI SISTEMI DISTRIBUITI" (Communications in Distributed Systems), which illustrates a layered communication model with levels (LIVELLO 0 to LIVELLO 5) and protocols (protocollo livello 0 to protocollo livello 5). Below the video player, there is a "Time value to seek to in object's timeline" field and "Click to seek" and "Restart" buttons. At the bottom, a navigation bar contains several icons, with one icon highlighted by a red box. The browser's status bar at the bottom indicates "Operazione completata" (Operation completed).

# Documenti strutturati



Journal

Article1

Article2

Article3



Thesis

Chapter1

SectionA

SectionB

Chapter2

SectionC



Video

Sequence1

SceneA

FrameA1

FrameA2

SceneB

FrameB1

Sequence2



Tutorial

Introduction

Part1

Part2

Part3



# Documenti annotati



- ❖ rating
- ❖ comment
- ❖ description
- ❖ link
  - agreement
  - disagreement
  - explanation

Untitled Document - Microsoft Internet Explorer

Annotation		ics.forth/ann00
Author	John Doe	
Project	SCHOLNET	
Group		
Program		
Subject	Digital Libraries, nice user interfaces	
Date	2002/08/23	
This book describes very well how to set up user interfaces in order to provide some search mechanisms. I think there are some important problems adressed, for instance, how to get the best tradeoff between advanced queries and usable interfaces.		
References		
Agreement	<a href="#">ics.forth/ann01</a> , <a href="#">ics.forth/ann02</a>	
Disagreement	<a href="#">ics.forth/ann04</a>	
Explanation	<a href="#">ics.forth/ann06</a>	

Annotate

- ❖ on the whole document or on its parts
- ❖ authored by different people
- ❖ public or restricted

# Documenti multimediali

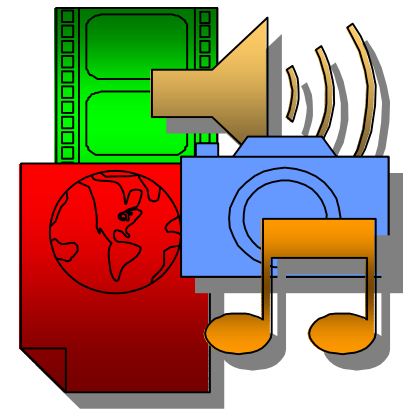
- ❖ Documents in different languages can be maintained in the same DL
- ❖ These documents can be accessed by querying in the language of the document and in any other supported language



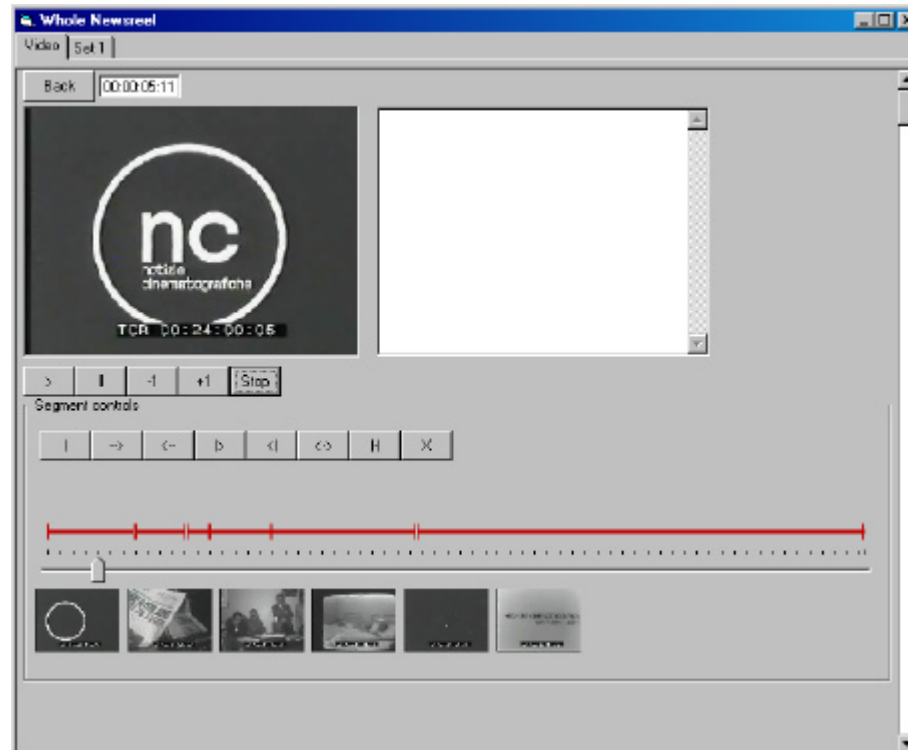
# New Services - (1)

❖ New document types impose a re-thinking of the “traditional” library services

- Submission
- Description
- Search
- Dissemination
- .....



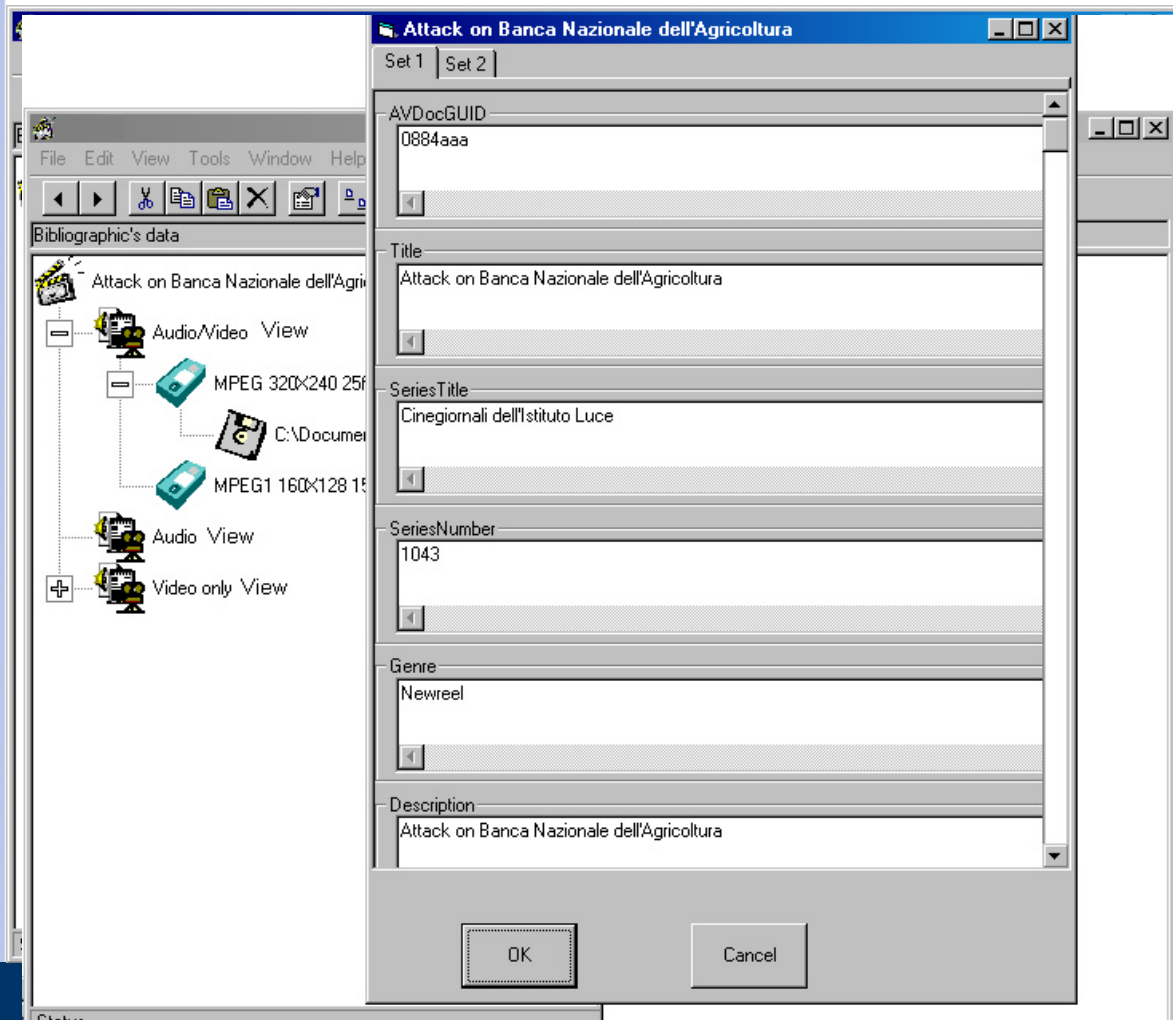
# the acquisition of video documents



It must be possible to structure the video into meaningful parts (sequences, scenes, frames) ...

# The description of video documents

... and describe the video and its parts separately



# Another example: the search

## Multiple search types options

- ❖ Free text search
- ❖ Fielded search
- ❖ Monolingual and cross language search
- ❖ Similarity search
- ❖ Search using the doc structure
- ❖ Search on annotations
- ❖ .....

# An example of complex query



All the *seminars*  
such that *contain a slide* such that  
*is about XML*  
and *contain a video* such that  
*there is an image of Serge Abiteboul*  
and *have been created after December 2000*  
and *have a "good" rating*

# New services – (2)



- ❖ New services (not necessarily document centered) enabled by recently developed technologies, can be included in a DL to improve its potential usage
  - Recommenders
  - Co-operative work services
  - Peer-reviewing supporting services
  - Authoring services
  - E-learning tools
  - ....

# An example: a collaborative environment

Collaborative Work

File Edit View Options GoTo

Home Comms Clipbd Waste Addr S & B

Your location:

catch up copy cut delete

Name	Shared	Note	Rating	Owner	Date	Events	Action
<b>Adaptive perceptual pattern recognition by self-organizing neural networks: Context, uncertainty, multiplicity, and scale</b>				cyclades	2003-01-14 18:32	★	▶
<b>Exploiting context when learning to classify</b>				cyclades	2003-01-14 18:32	★	▶
<b>Neighborhood Expansion Grammars</b>				cyclades	2003-01-14 18:32	★	▶
<b>Robust classification with context-sensitive features</b>				cyclades	2003-01-14 18:32	★	▶

BSCW © 1995-2001 GMD, © 2001 DrbitTeam

Mozilla

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop

http://calliope.ics.forth.gr:7007/Cyclades/UI/Login

Search Print

Home Bookmarks Google Pisa forecast RedHat rionegro.com Clarín.com RaiNews24

## Collaborative Work

File Edit View Options GoTo



Your location:

catch up send copy cut delete archive rate

agents

Name

Shared Note Rating Owner

Date

Events Act

<input type="checkbox"/>		<b>Recommendations</b>			Cyclades	2003-02-13 12:13		
<input type="checkbox"/>		<b>Design and Control of Large Collections of Learning Agents</b>			avancini	2003-01-20 16:16		
<input type="checkbox"/>		<b>Running agent-based simulations</b>			avancini	2002-12-03		

system

Mozilla

File Edit View **Bookmarks** Tools Window Help

Back Forward Reload Stop  Search Print

Home Bookmarks Google Pisa forecast RedHat rionegro.com Clarin.com RaiNews24

## Collaborative Work

File Edit View Options GoTo

Home Comms Clipbd Waste Addr

Your location:

catch up delete rate

Recommendations

	Name	Shared	Note	Rating	Owner	Date	Events	Action
<input type="checkbox"/>	<b>CYCLADES</b> CYCLADES super-collection				Cyclades	2003-02-05 18:31		
<input type="checkbox"/>	<b>New Community</b>				Cyclades	2003-02-06 17:41	★	
<input type="checkbox"/>	<b>Physics - Gravity</b>				Cyclades	2003-02-06 17:41	★	
<input type="checkbox"/>	<b>Physics - Mechanics</b> focussing on mechanics				Cyclades	2003-02-06 17:41	★	
<input type="checkbox"/>	<b>dian.tan</b>				Cyclades	2003-02-06 17:19	★	
<input type="checkbox"/>	<b>renda</b>				Cyclades	2003-02-13 12:13	★	
<input type="checkbox"/>	<b>TingLi</b>				Cyclades	2003-02-06 17:19	★	
<input type="checkbox"/>	<b>wirsam</b>				Cyclades	2003-02-06 17:19	★	

# Nuove direzioni di ricerca DL

- ❖ have been developed as “ad-hoc” systems











- ❖ These systems require a great investment in terms of man power and technologies for the implementation of the software and their maintenance
- ❖ Skilled personnell is needed
- ❖ Few communities can build their own DL

# ... to a digital library service system



## A digital library service system

Flexible and open DL system that offers DL services. It can be customized according to the characteristics of the context where the DL must operate and to the needs of its users

(like a DBMS)

# Example of customization dimensions

## ❖ Service specific parameters

- Metadata formats
- Document types
- Annotation model
- Controlled vocabulary
- Query language
- Formats of the results



# Customization is not enough...



- ❖ Ciascuna comunità richiede specifici servizi, in aggiunta ai servizi di base (basic services= e.g. search and browse)
- ❖ For example:
  - A community of physics may need a specific service to test the consistency of the results published
  - A worldwide community of medicine may need a translation service

# Openness is also required



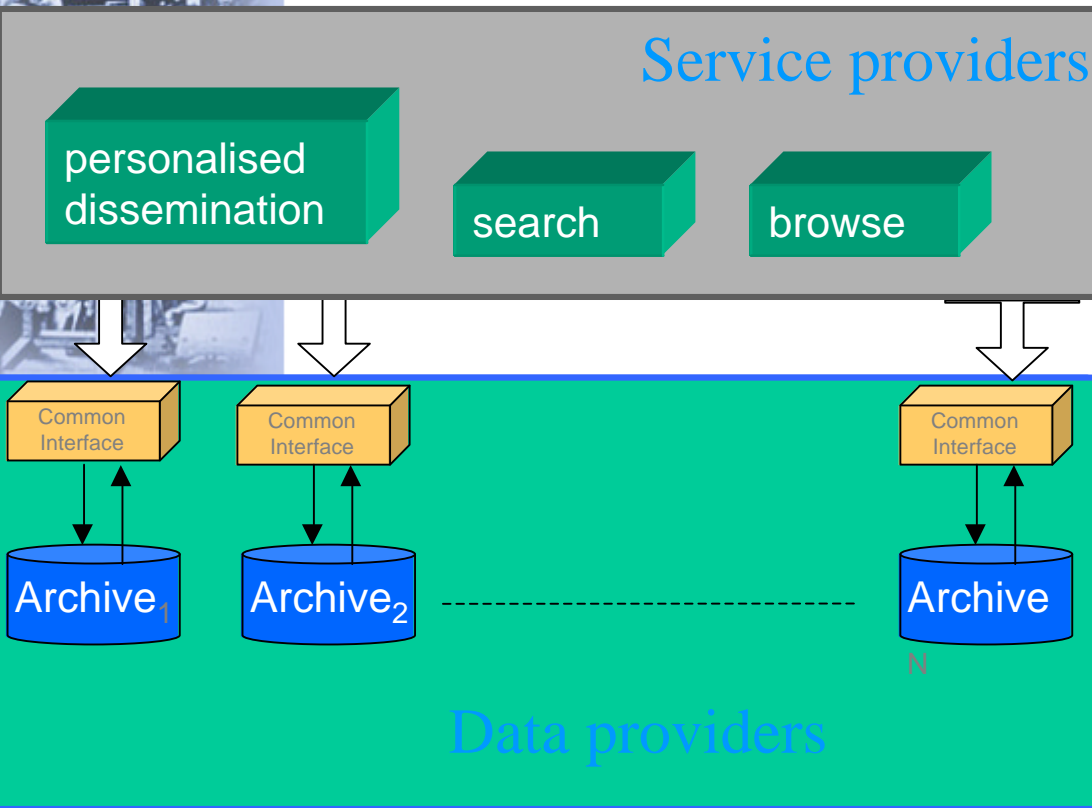
- ❖ “Open” means that new services can be easily added (expandability)
- ❖ Each community of users can add their own specific services
- ❖ The use of a DL may raise new requirements: other services can be added over the DL lifetime (dynamic expandability) to cover emerging needs

# Another trend: exploiting existing content



- ❖ The production of the DL content is a very expensive process
- ❖ DLs can also be built by exploiting content stored by existing distributed heterogeneous sources

# The reference model



Each service operates on data of multiple archives

data providers open their archives

# Data and service providers



- ❖ Advantages of this approach:
  - Third parties services that operate on existing data can be implemented
  - A source can be accessed through advanced services built by others

# Interoperability solutions



- ❖ Several solutions are possible:
  - The services apply schema mappings
  - Data providers implement a more or less complex protocol (e.g. OAI)
  - Automatic mapping generation (a current research topic)

# Il futuro



- ❖ Molte comunità di ricercatori vorrebbero avere una propria biblioteca digitale come strumento di comunicazione e collaborazione
  
- ❖ Queste comunità spesso:
  - hanno dimensioni contenute
  - sono distribuite geograficamente
  - non sono permanenti
  - non hanno le risorse materiali e umane per sostenere da sole la costruzione e gestione di una BD

# DILIGENT



- ❖ Progetto Integrato finanziato dalla Comunità Europea (Settembre 2004 - Agosto 2007)
- ❖ Partecipanti:

# Obiettivo



# Ruolo della SNS



# Fine storia



# OpenDLib



- ❖ Sistema che offre servizi di biblioteca digitale personalizzabili
- ❖ I servizi possono essere estesi

# Contenuto della biblioteca



- ❖ Il contenuto può essere
  - immesso direttamente dall'autore o dall'amministratore
  - raccolto automaticamente da "archivi" esistenti
- ❖ Testi, registrazioni video e audio, immagini, oggetti composti, ecc.

# I documenti

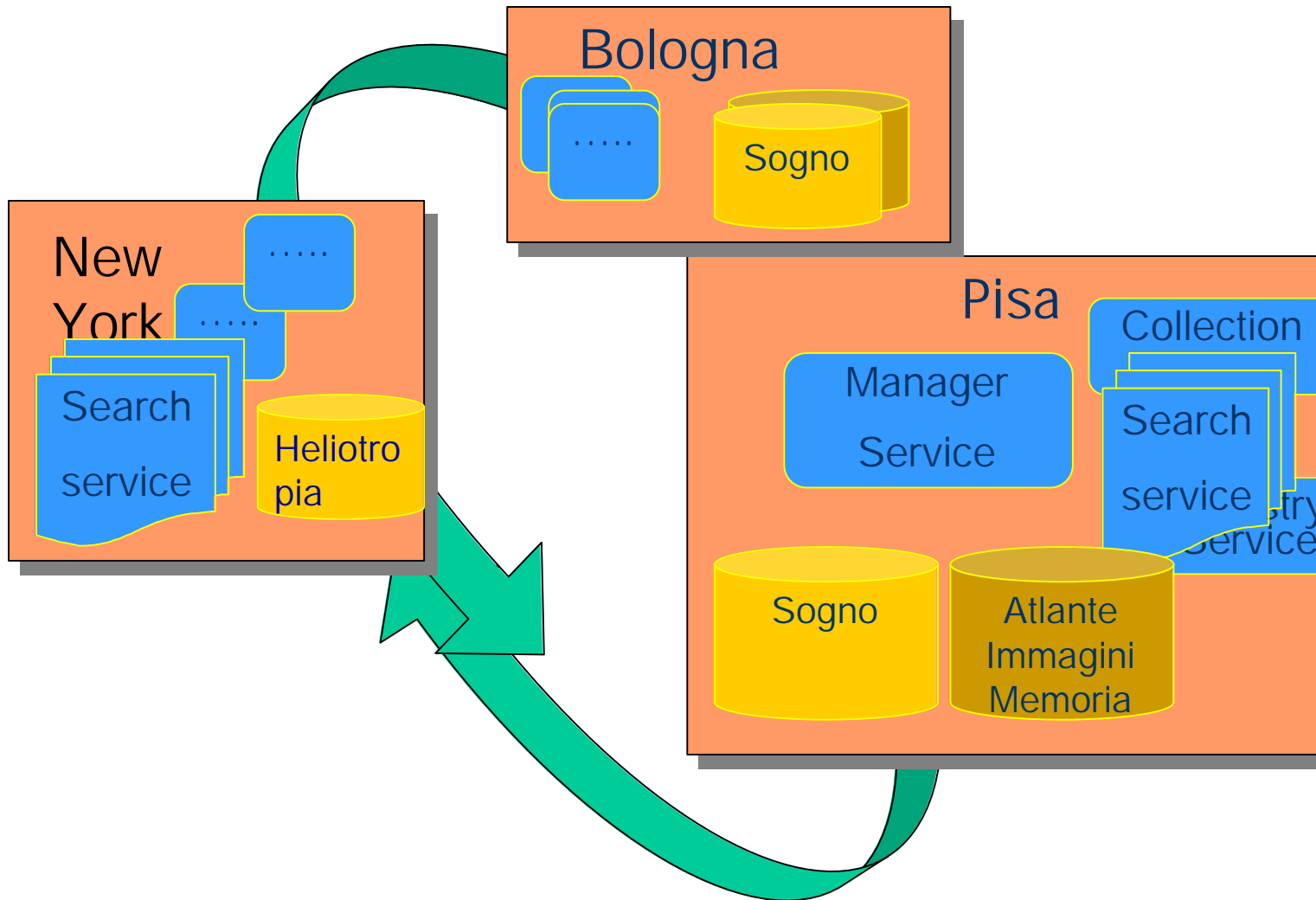


# Collezioni virtuali



- ❖ I documenti sono organizzati in collezioni virtuali
- ❖ Collezioni possono essere create da chiunque ne ha i diritti specificando un criterio
- ❖ L'accesso alle collezioni può essere ristretto ad un certo gruppo di utenti specificato dal creatore della collezione
- ❖ L'accesso all'insieme dei documenti è mediato dai servizi. Questo garantisce un certo livello di protezione su Internet

# Architettura federata



# Collezioni



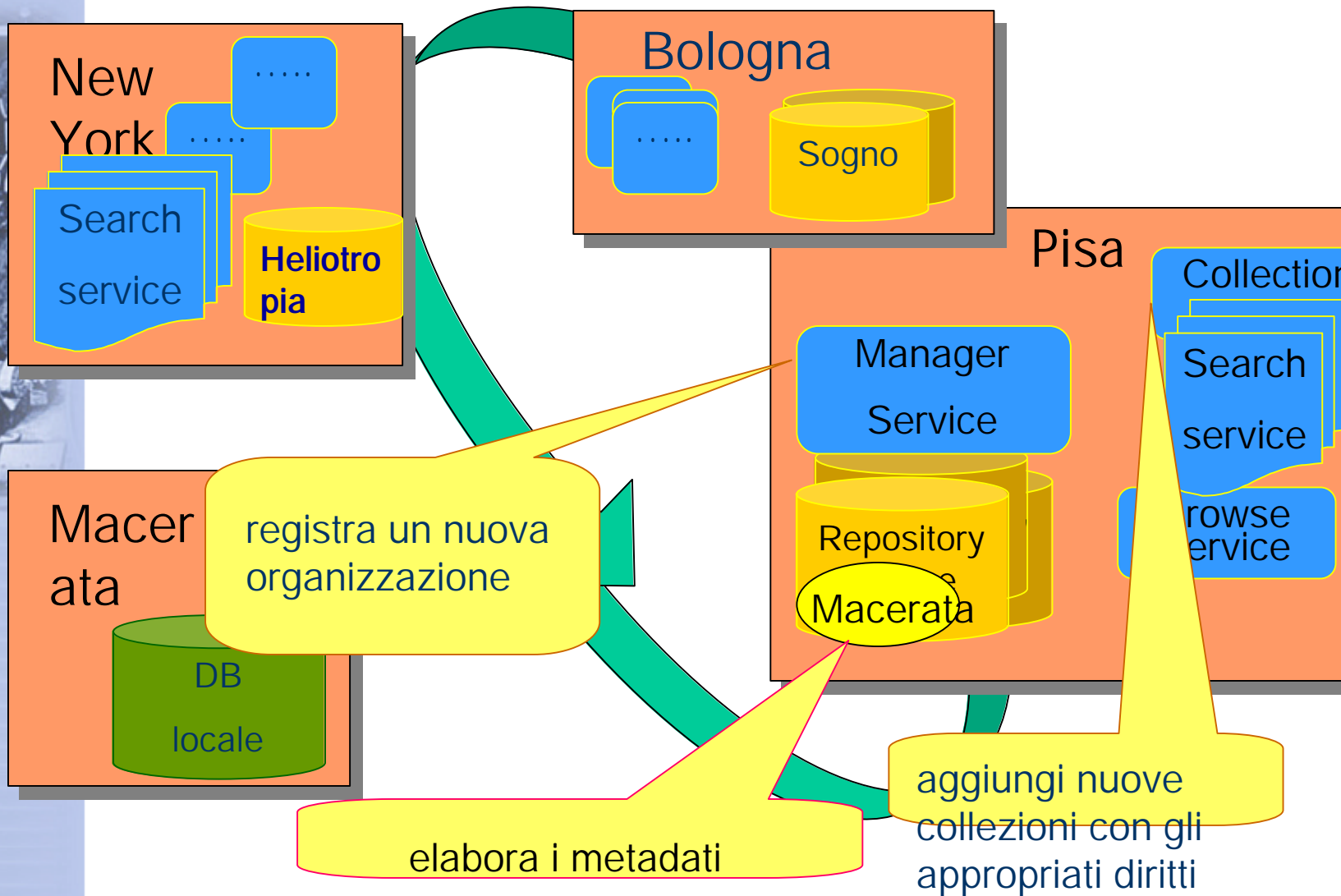
- ❖ Il Sogno raccontato nella letteratura moderna (unità del progetto)
- ❖ Arti della memoria (SNS, ....)
- ❖ Heliotropia (pubblica)
- ❖ Documenti relativi a Giovanni Boccaccio (studenti del corso "Letteratura Italiana" anno '03-'04 al Dip. Italian Studies at Brown University)
- ❖ Sogni descritti da Giovanni Boccaccio (studenti del corso "Progettare sul Web")
- ❖ ....

# Un possibile scenario



- ❖ L'unità di ricerca di Macerata produce materiale (testi ed immagini) in formato digitale
- ❖ Parte di questo materiale può essere reso pubblico, un'altra parte può essere condivisa tra i ricercatori del progetto "Sogno"
- ❖ Macerata vorrebbe poter accedere anche al materiale prodotto da altre unità di ricerca
- ❖ Macerata non ha abbastanza risorse umane e materiali per dedicare alla creazione e gestione di una biblioteca digitale

# Che cosa si può fare?



# Un altro scenario



- ❖ Macerata vuol mantenere sotto il proprio controllo i documenti con accesso limitato
- ❖ I dati e metadati prodotti dalla unità di Macerata possono essere elaborati automaticamente per estrarre altre informazioni
- ❖ Alcuni dei sogni sono descritti in lingue non ancora indicizzate dalla biblioteca esistente

# Che cosa si puo' fare?

