

# Building Digital Libraries on Open Archives

Donatella Castelli  
IEI-CNR  
Italy

# Open Archive

- **Archive**

repository of digital information

- **Open archive**

archive that provides a machine interface for making its content available to external services

# E-print archives

- Different scholarly communities “communicate” through electronic archives
  - **ArXiv** (Los Alamos National Laboratory Physics Archive -100.000 papers, 50.000 user daily)
  - **NCSTRL** (University of Cornell - papers on Computer Science from 120 Institutions)
  - **NTLDL** (electronic theses and dissertations)
  - **RePec** (papers on Economics)

# Cross-domain access

- Each archive has its own interface and its own services
- Cross-archives access is not possible
- Mechanisms for supporting interoperability are required

# Open Archives Initiative

Scientific communities that publish their pre-prints on electronic archives met in Santa Fe (New Mexico) on July 1999 and set up the **Open Archives Initiative**

- ArXiv
- NCSTRL
- NDLTD
- RePEc
- CogPrints

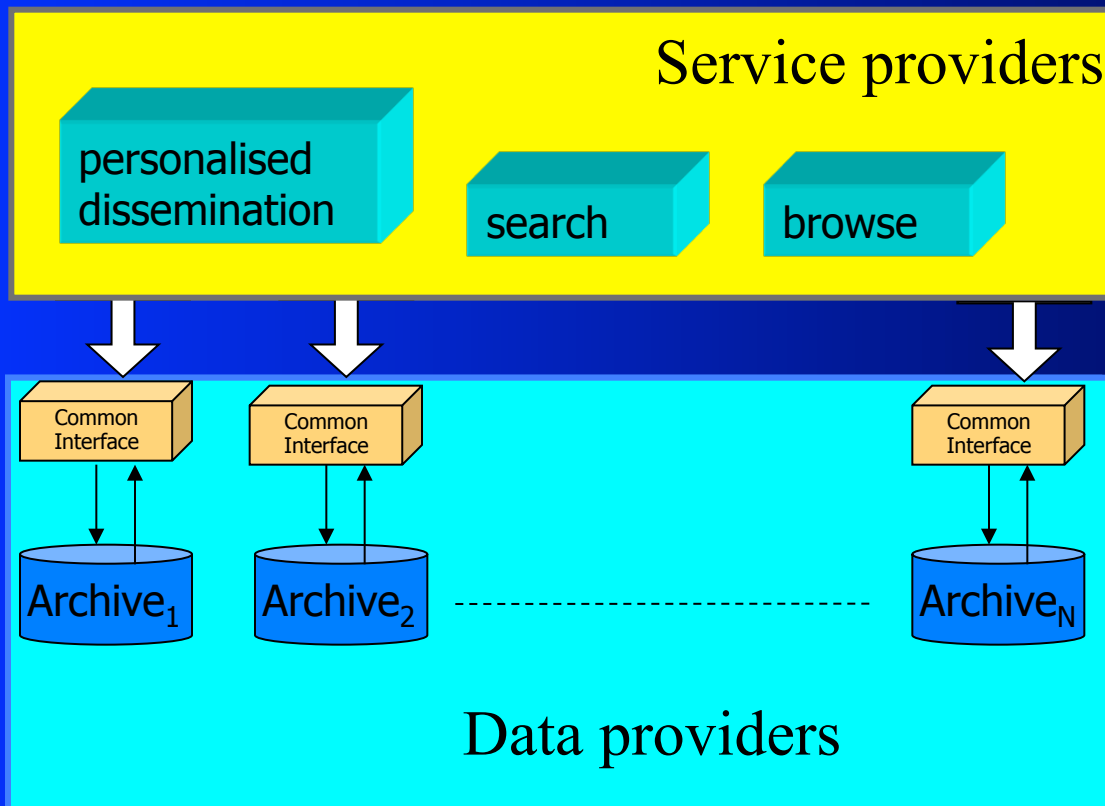
# OAI objective

To explore the co-operation among e-print archives as a way to contribute in a concrete manner to the transformation of the scholarly communication

# Key Issues

- To solve the problem of interoperability among the e-prints archives
- Very simple, low-barrier to entry interface that shifts implementation complexity and operational processing load away from the archives

# Solution proposed



OAI

Protocol for  
Metadata  
Harvesting  
(OAI-PMH)

# OAI Metadata Harvesting Protocol

service provider

H  
A  
R  
V  
E  
S  
T  
E  
R

- Identify
- ListMetadataFormats
- ListSets

- ListRecords
- ListIdentifiers
- GetRecord

- HTTP-embedded
- XML response format

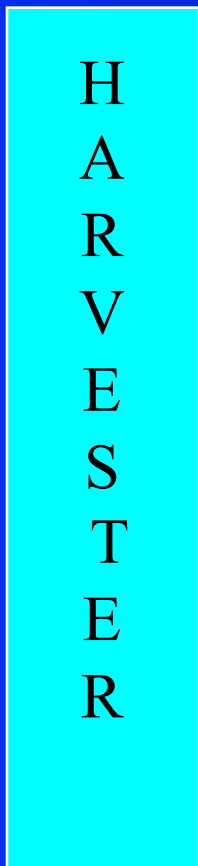
data provider

R  
E  
P  
O  
S  
I  
T  
O  
R  
Y

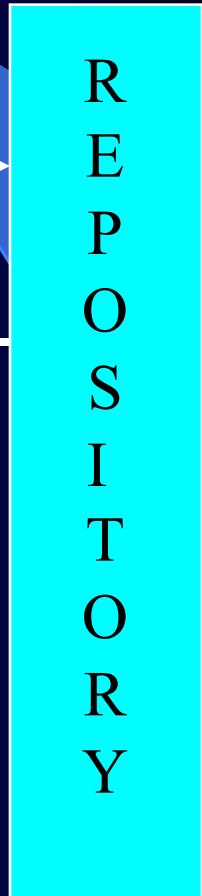
# Identify

service provider

data provider



Identify

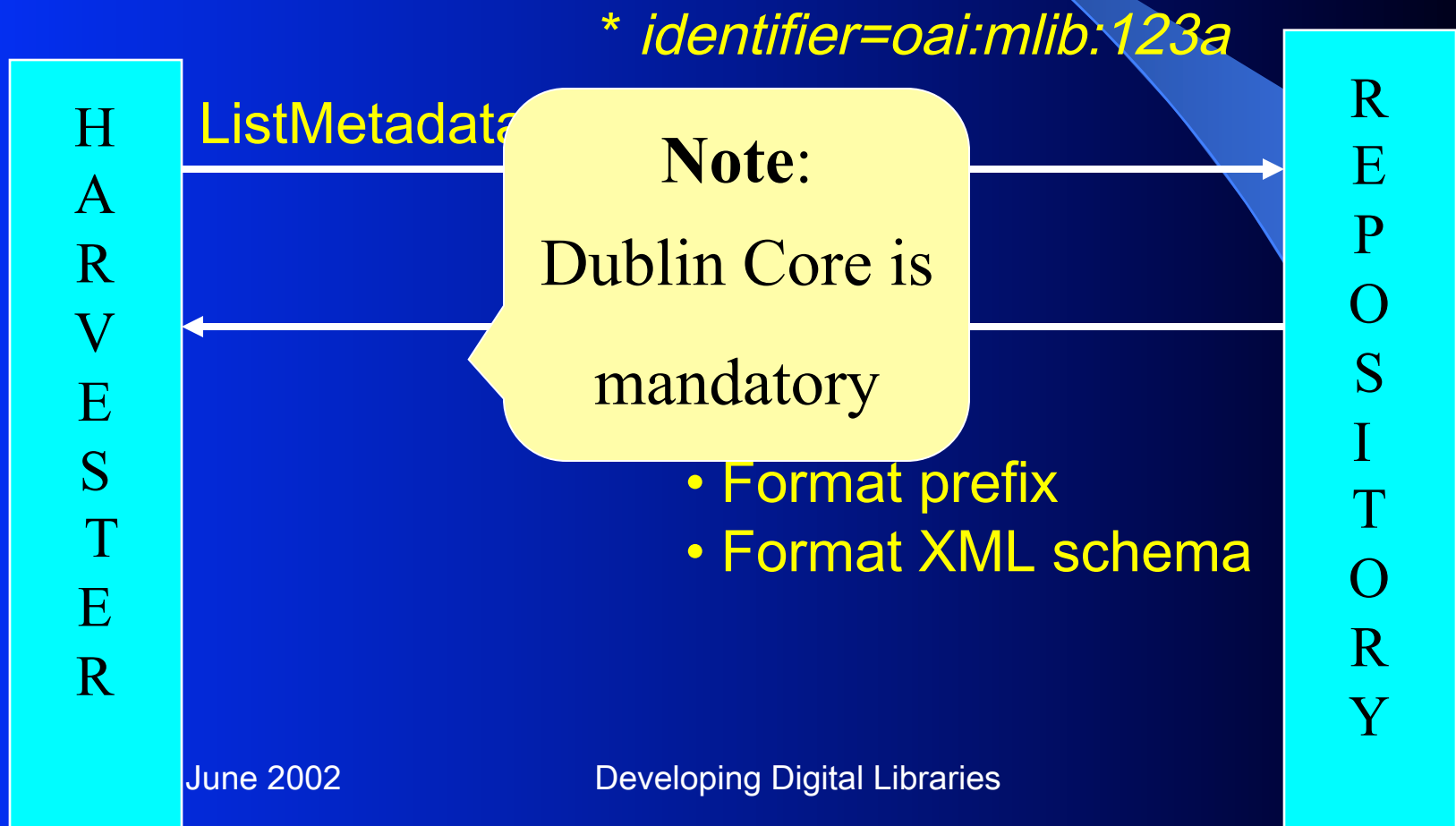


- Repository name
- Base-URL
- Admin e-mail
- OAI protocol version
- *Description Container*

# List Metadata Formats

service provider

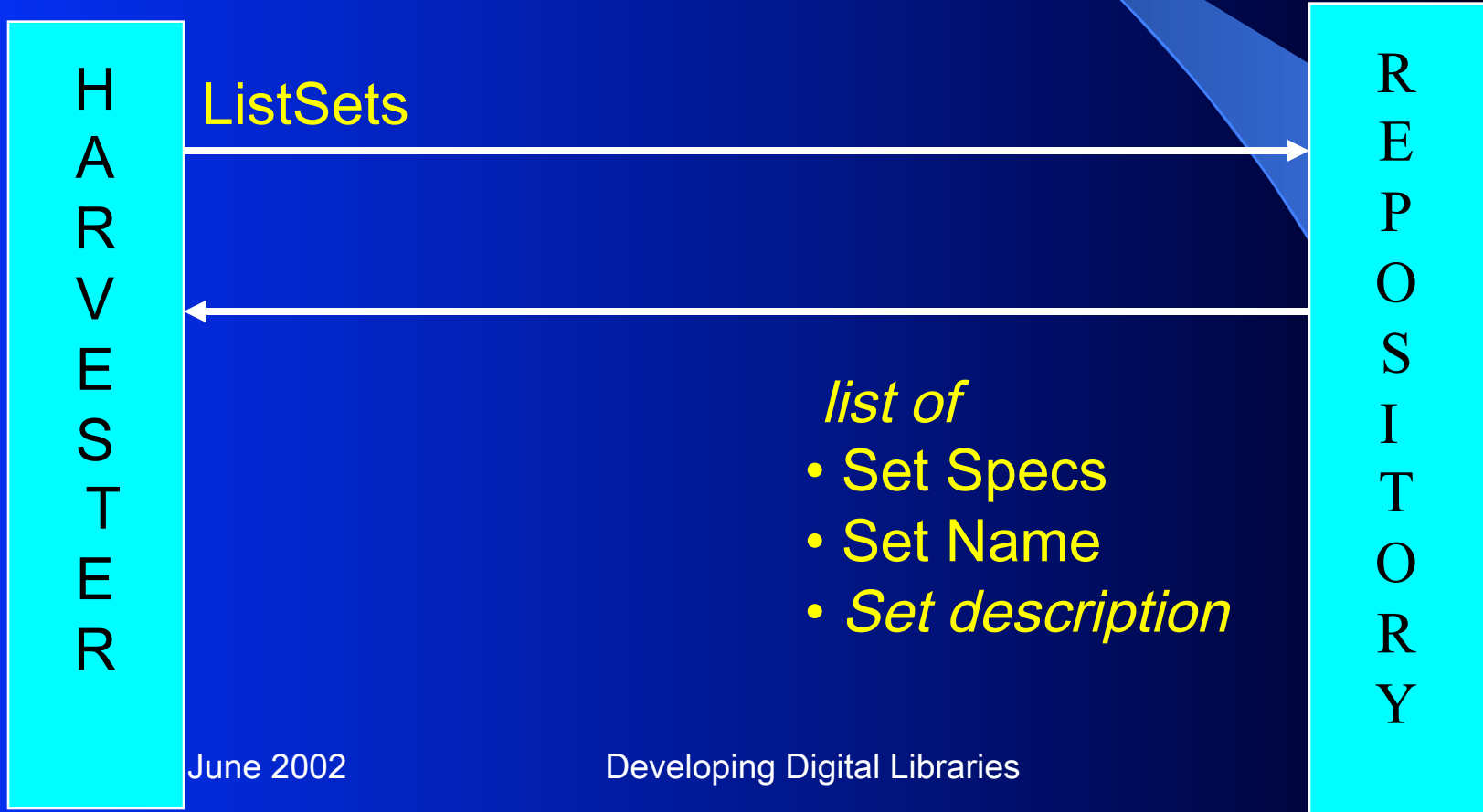
data provider



# List Sets

service provider

data provider



June 2002

Developing Digital Libraries

# List Records

service provider

data provider

*\* from=a*  
*\* until=b*  
*\* set=klm*  
*\* resumption token=XXX*  
*\* metadataPrefix=oai\_dc*

ListRecords

H  
A  
R  
V  
E  
S  
T  
E  
R

R  
E  
P  
O  
S  
I  
T  
O  
R  
Y

*list of*

- Identifier
- Datestamp
- Set spec
- Metadata
- *About Container*

# List Identifiers

service provider

data provider

- \* *from=a*
- \* *until=b*
- \* *metadataPrefix=oai\_dc*
- \* *resumption token=XXXX*
- \* *set=klm*

ListIdentifiers

H  
A  
R  
V  
E  
S  
T  
E  
R

R  
E  
P  
O  
S  
I  
T  
O  
R  
Y

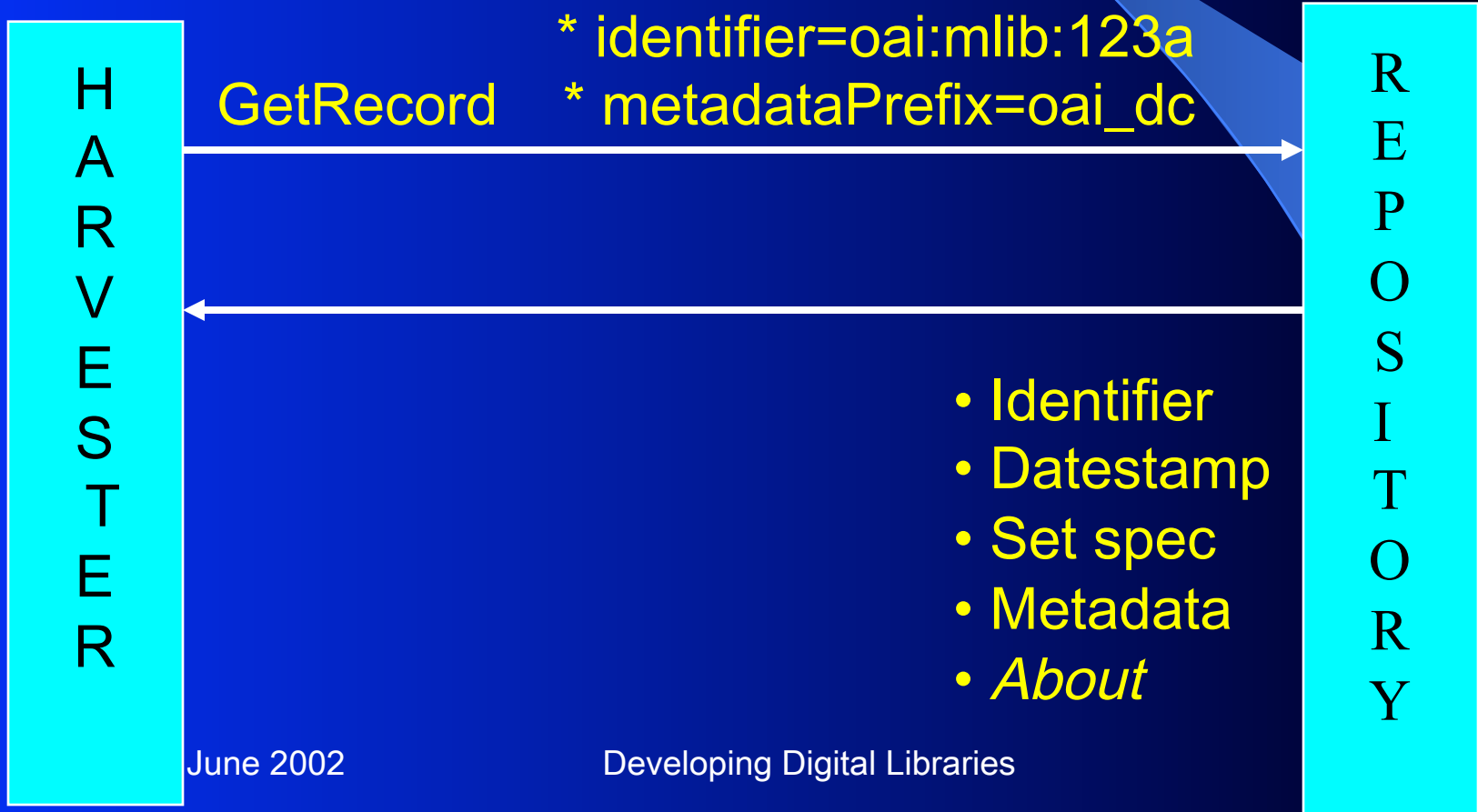
*list of*

- Identifier
- Datestamp
- Set specs

# Get Record

service provider

data provider



# OAI-PMH

Version 2.0 available since 14<sup>th</sup> of June 2002 at:

<http://www.openarchives.org/news/oaiv2press020614.html>

<http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>

# OAI compliant data providers

- Around 80 archives have implemented OAI-PMH
- Other communities
  - Libraries
  - Museums
- Budapest Open Archives Initiative  
(Open Society Institute – Soros Foundation)

# Service providers

- Cross-archives search

- Arc (<http://arc.cs.odu.edu/>)
- citebaseSearch (<http://citebase.eprints.org/cgi-bin/search>)
- my.OAI (<http://www.myoai.com/>)
- .....

- Other services

- Cyclades
- Scholnet



# Cyclades

E U V Framework Project – February 2001

- IEI-CNR (Italy)
- UNIVERSITY OF DORTMUND (Germany)
- FORTH (Greece)
- FRAUNHOFER FIT (Germany)
- ERCIM (France)

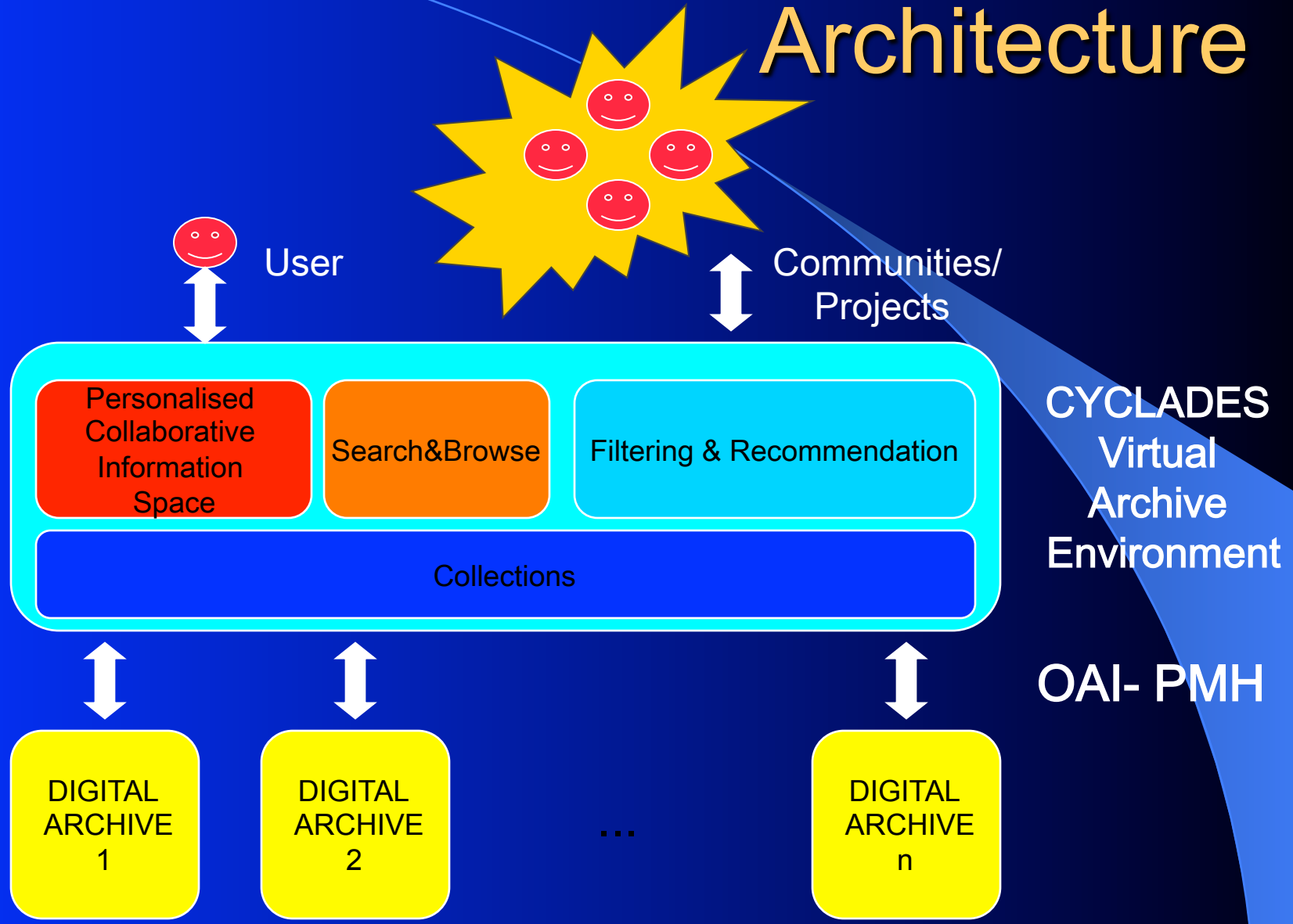
# Objectives

Develop a system which provides *an open collaborative virtual archive environment* for supporting single scholars or communities of scholars

# Functionality

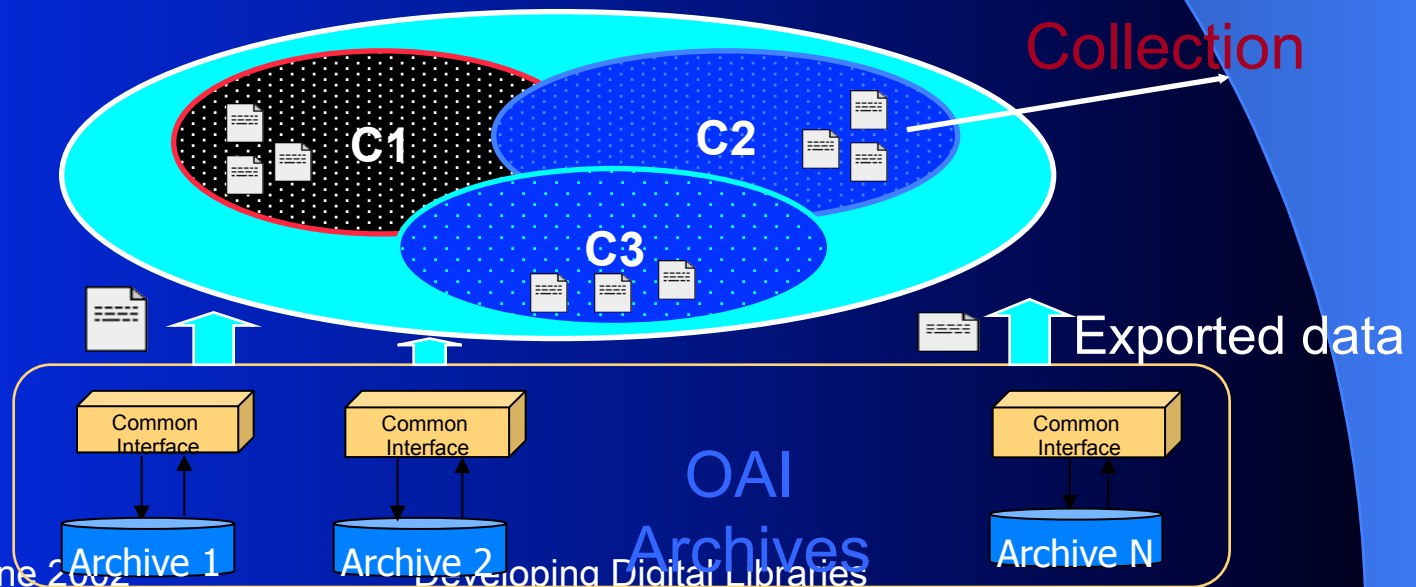
- Search in large, heterogeneous, multidisciplinary digital archives
- Personalised Information Space Organisation
- Support to collaboration
- Filtering and Recommendation

# Architecture



# Virtual Collections

- The information space is organised into virtual collections
- Users and Communities may define their own collections, e.g. by specifying criteria and by refinement of existing ones
- The OAI archives remain hidden to the users and communities



# Search & Browse

- Search using the collection's search fields
- Query formulation through browsing
- Multiple schema browsing allowed

## Search

Collection = "*Logic*",  
Author = "*Straccia*",  
Description = "*Fuzzy Logic*"

## Browse

### Schema

Author  
Description  
Title  
...

### Attribute

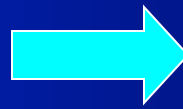
Castelli  
Straccia  
Thanos  
...

# Search & Browse

- Users may save the retrieved metadata records into their folders

QUERY:=

Collection = "*Logic*",  
Author = "*Straccia*",  
Description = "*Fuzzy Logic*"

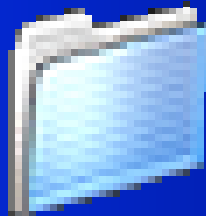


## Result List

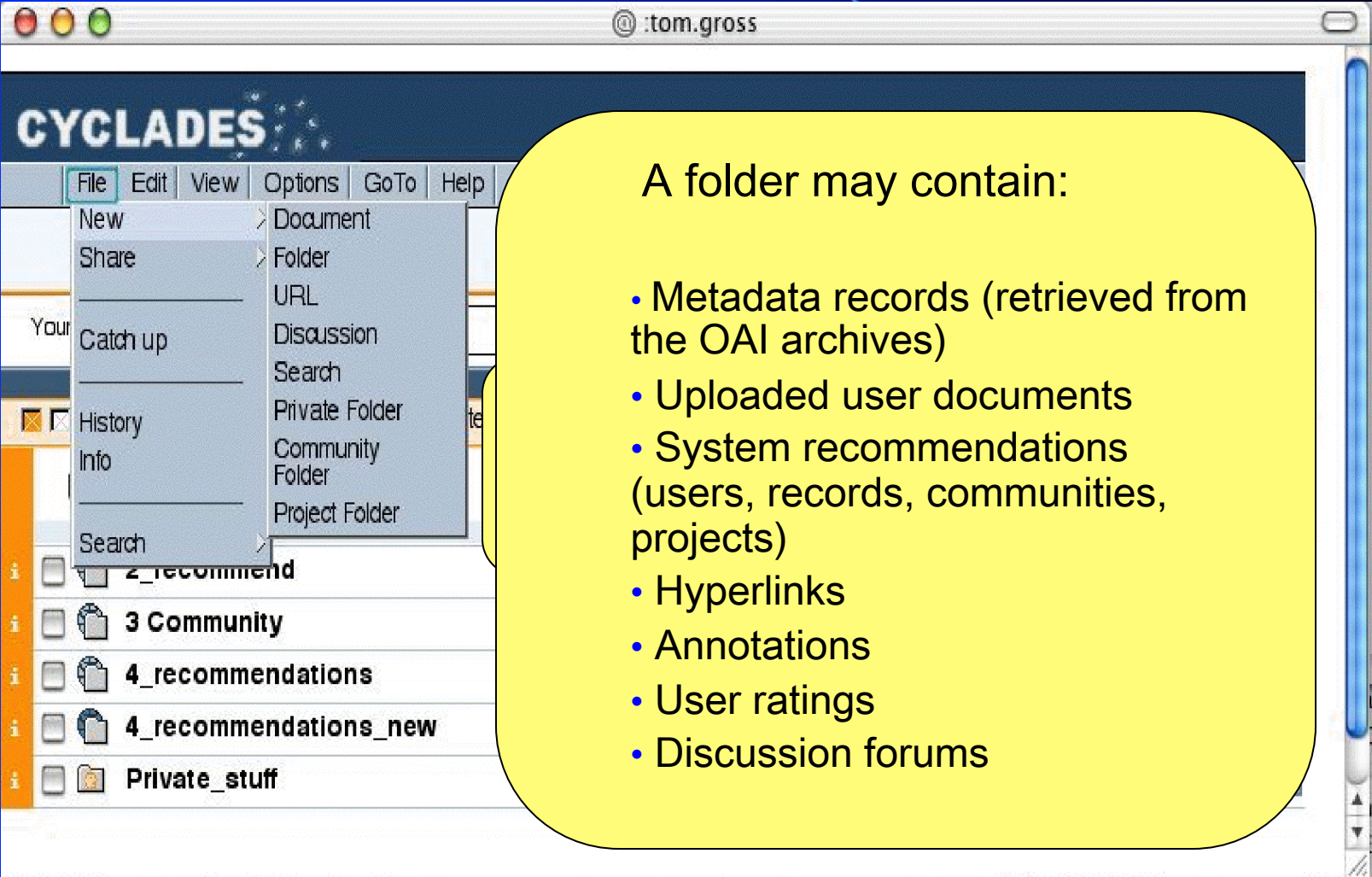
R1. A fuzzy Description Logic  
R2. A note on fuzzy IR

...

save



# Personalised, Collaborative Information Space



The screenshot shows a web browser window titled "@ :tom.gross" displaying the CYCLADES application. The application has a menu bar with "File", "Edit", "View", "Options", "GoTo", and "Help". The "File" menu is open, showing options: "New" (with a sub-menu for "Document"), "Share" (with a sub-menu for "Folder", "URL", "Discussion", "Search", "Private Folder", "Community Folder", "Project Folder"), "Catch up", "History", "Info", and "Search". Below the menu, a list of folders is visible: "2\_recommend", "3 Community", "4\_recommendations", "4\_recommendations\_new", and "Private\_stuff".

A folder may contain:

- Metadata records (retrieved from the OAI archives)
- Uploaded user documents
- System recommendations (users, records, communities, projects)
- Hyperlinks
- Annotations
- User ratings
- Discussion forums

# Filtering & Recommendation

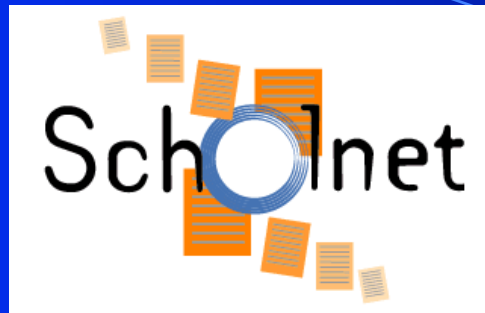
- Service learns the user information needs (**folder profile**) automatically from the user's folder content
- **Uses of folder profiles**
  - Filtering of metadata records  
Used by users to filter out irrelevant information during a search session
  - Recommendation of records, users, collections, communities  
Used by the system to automatically notify users about new documents relevant for them

# Recommendations

- Recommendations pertain to a user folder, i.e. user topic of interest
  - Document recommendation
  - Collection recommendation
  - User recommendation
  - Community recommendation

# Some considerations

- Cyclades is complex DL service that exploits data digitalised by others
- Existing archives can be accessed through the advanced Cyclades services
- Cyclades can be activated on selected OAI compliant archives



# Scholnet

- EU V Framework Programme project

CNR ( Italy)

FhG (Germany)

SICS (Sweden)

Univ. of Masaryk (Czech Republic)

INRIA (France)

FORTH (Greece)

ERCIM (France)

# Objective

SCHOLNET aims at developing a digital library infrastructure to support the communication and the collaboration within networked scholarly communities

# Functionality

- information acquisition, description, archiving, search, access, and dissemination of multimedia documents
- handling of annotations on documents
- multilingual access
- personalised information dissemination

# User communities

Scholnet must be able to serve the needs  
of any scholarly community

# Serving any community

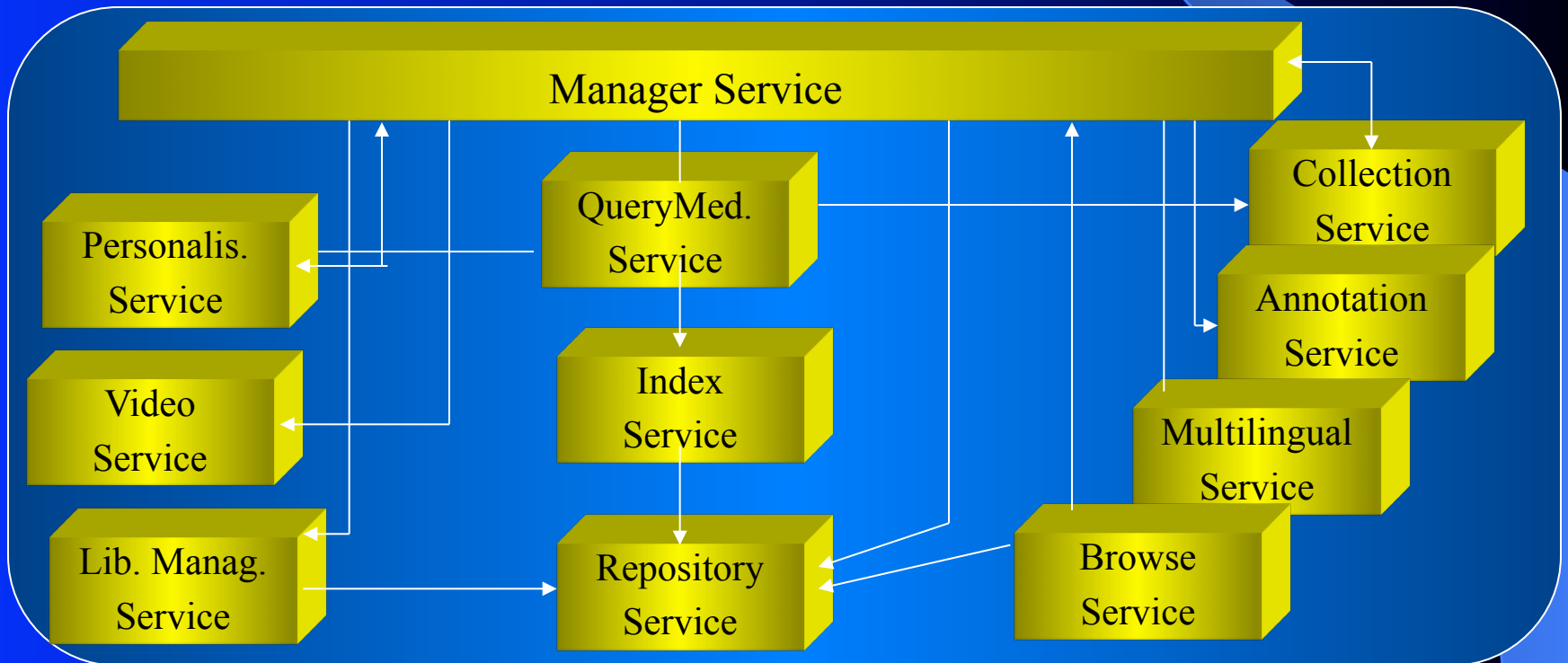
- Generic with respect to the DL content
  - Structure of the document
  - Organisation of the information space
  - Metadata format
  - Controlled vocabulary

# Serving any community

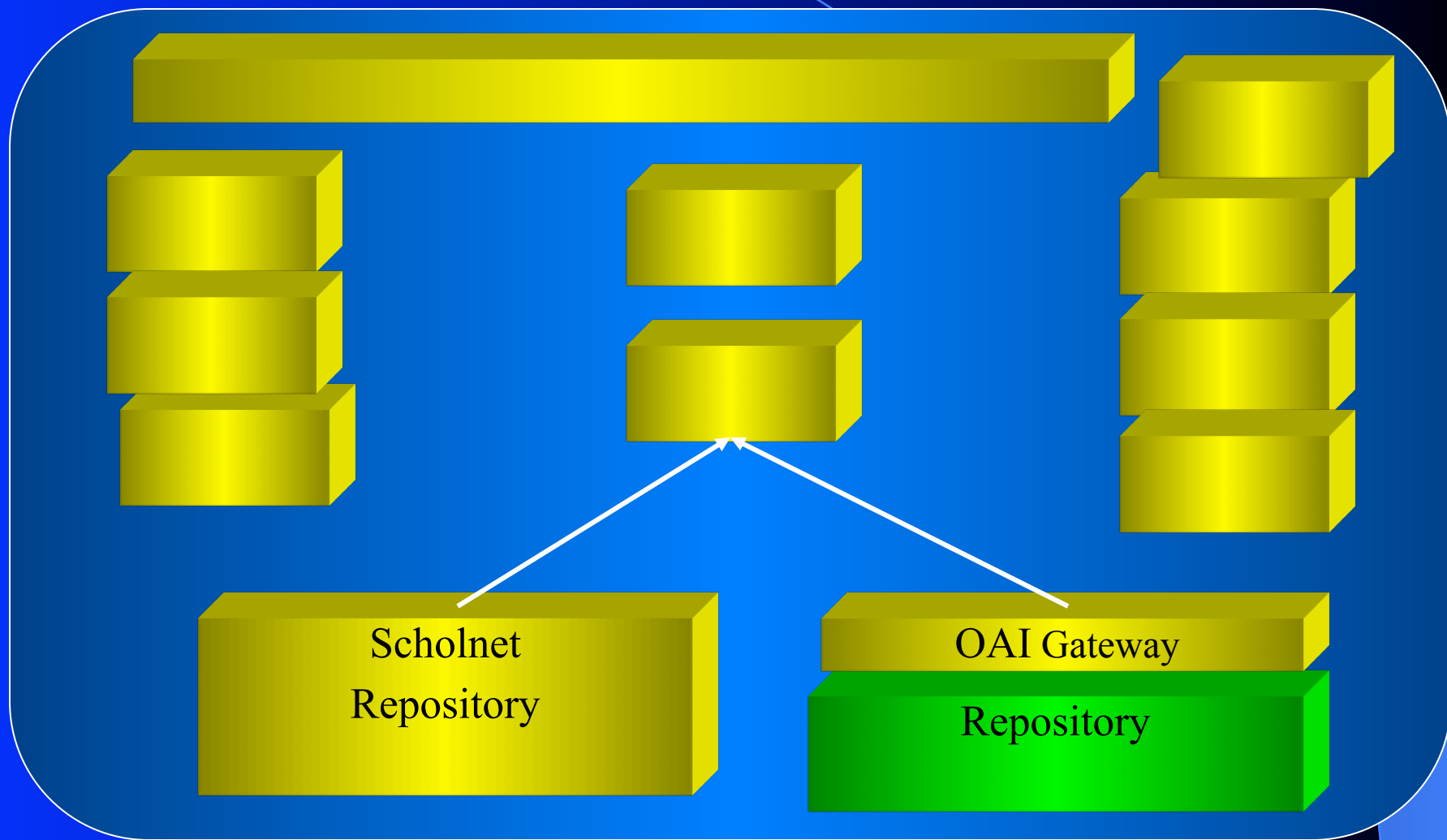
- Open
  - Easily extensible with other services that meet the specific needs of a user community

# Architecture

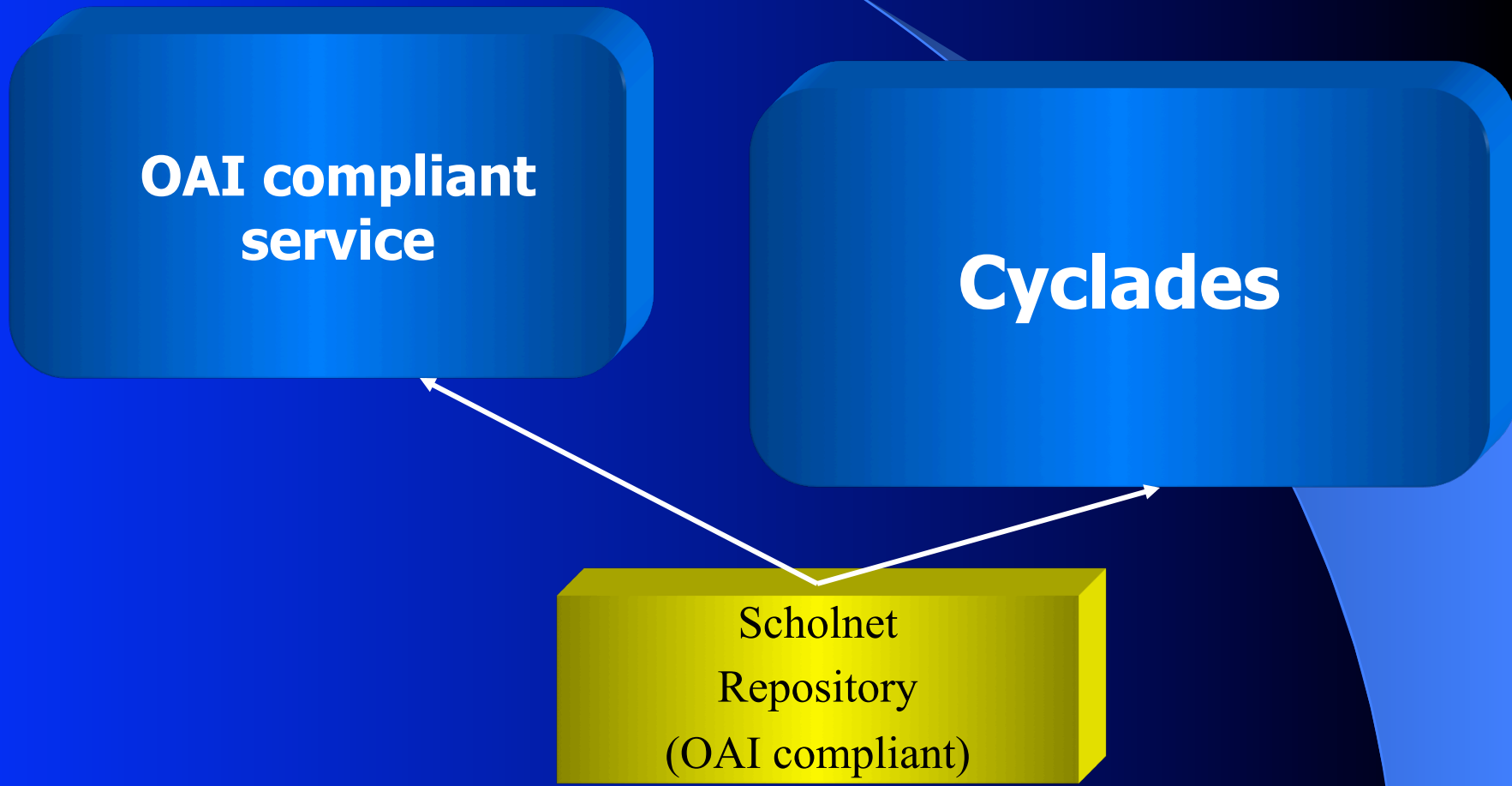
Federation of services (replicated and/or distributed) which communicate through an HTTP-based protocol



# Architecture (cont.)



# Interoperability





# Open Archives Forum

EU V Framework Accompanying Measure

## Objective

European forum for discussing issues related to the open archives and for disseminating information about the implemented solutions

open  
archives



6\7  
December  
2002

LISBON  
2nd workshop

to come

Programme

<http://www.oaforum.org>

Title to be decided

Open libraries and  
historical archives

# Some links

<http://www.openarchives.org>

<http://www.ercim.org/cyclades>

<http://www.ercim.org/scholnet>

<http://www.oaforum.org>