



# Introduction to DL.org, digital library modelling and the DELOS DL Reference Model

*Donatella Castelli, Yannis Ioannidis and Seamus Ross*

**4 October 2010**

DL.org Autumn School – Athens, 3-8 October 2010



# Outline

**The DL.org project**

**Digital Libraries: a brief overview**

**The DELOS Manifesto**

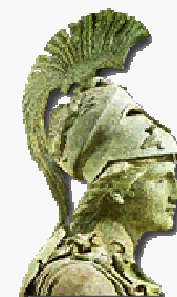
**Introduction to the Digital Library Reference Model**

.....

.....

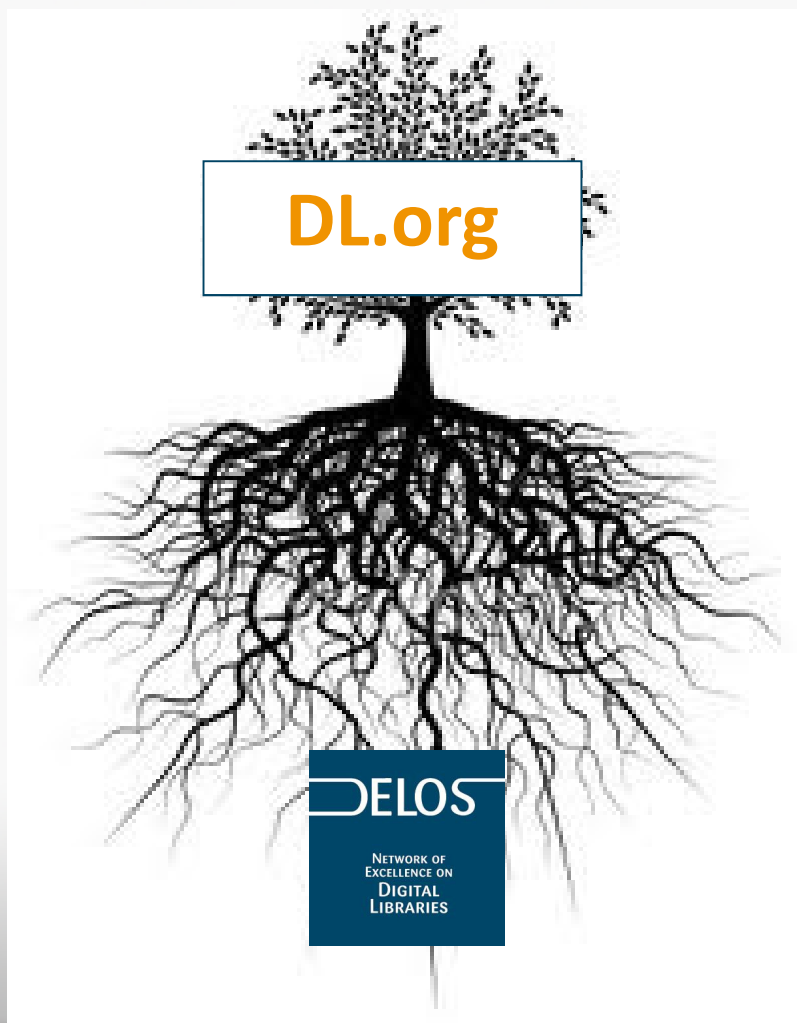
## DL.org Consortium

- CNR-ISTI (coordinator, Italy)
- NKUA (Greece)
- HATII (United Kingdom)
- Trust-IT (United Kingdom)



# DL.org

## *Coordination Action on Digital Library Interoperability, Best Practices and Modelling Foundations*



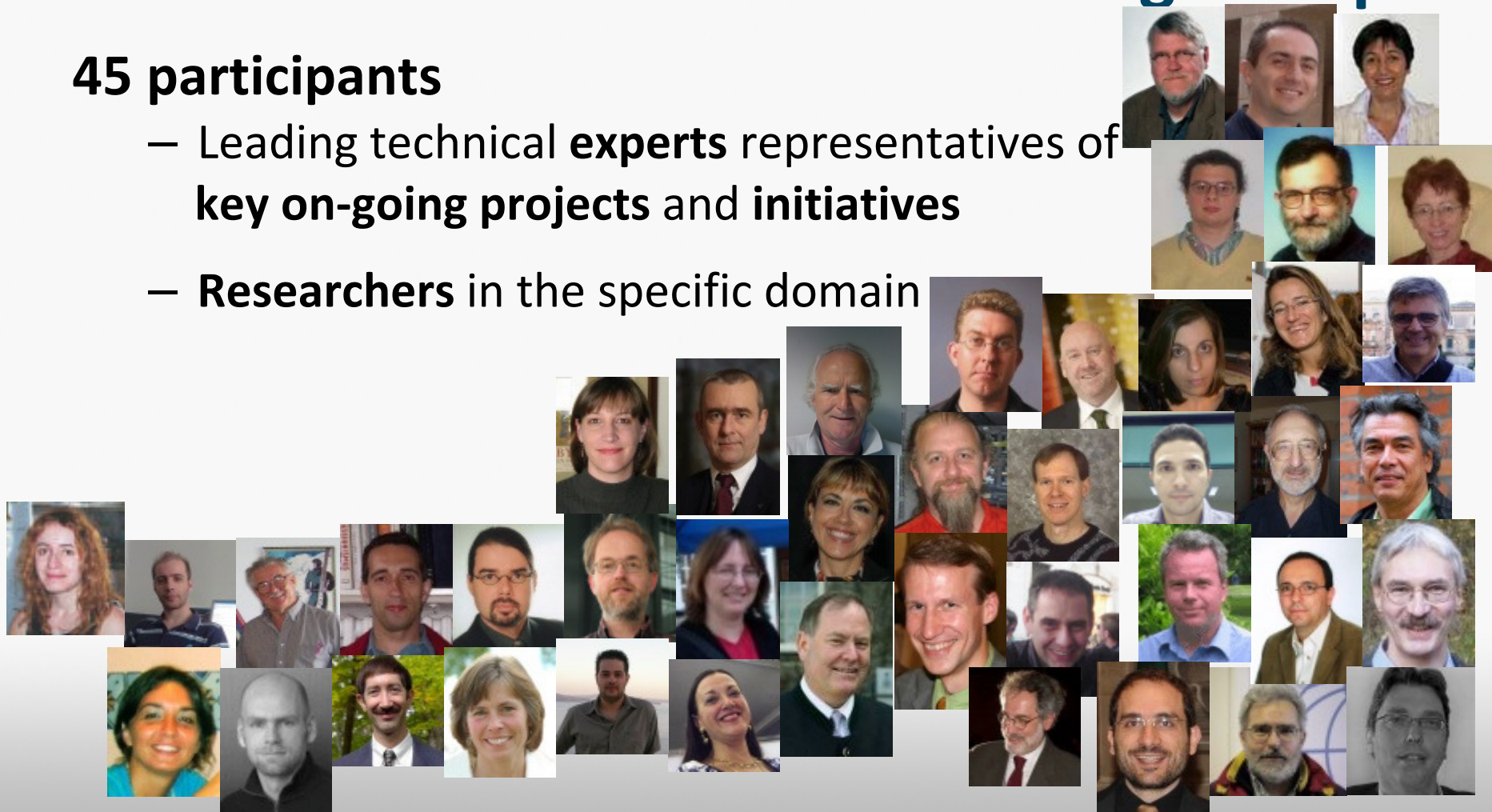
# DL.org objectives

- 1. Promote interoperability between current and future DL initiatives**
  - by adopting a systematic approach
  - by mobilizing key people and projects
- 2. Deliver a DL Technology and Methodology Cookbook**
  - by identifying most common best practices and technology patterns
- 3. Consolidate and enhance the DELOS Reference Model**
  - by exploiting feedback collected by members of different communities
- 4. Disseminate project outcomes**
  - via Web site, Workshops, Schools

# Six Thematic Working Groups

**45 participants**

- Leading technical **experts** representatives of **key on-going projects and initiatives**
- **Researchers** in the specific domain



## 18 members:

- Stakeholders of **DL organizations** and **coalitions**
- Well known **experts** in the DL field

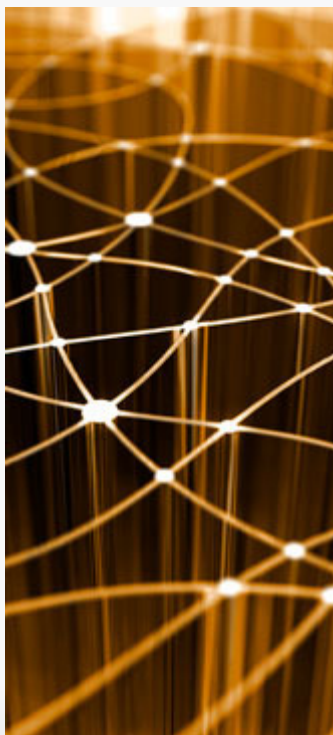




# Meetings and Events

1. ***1st Face-to-Face meetings, late Spring-Summer, 2009***
2. ***2nd Face-to-face meeting, Rome, 30 November 2009***
3. ***DL.org 1st Workshop: Digital Libraries: Interoperability, Best Practices & Modeling Foundations, 1 October 2009, Corfu, Greece***
4. ***DL.org 2nd Workshop: Making Digital Libraries Interoperable: Challenges and Approaches, 9-10 September 2010***





[www.dlorg.eu](http://www.dlorg.eu)

# DIGITAL LIBRARIES OVERVIEW

# “Digital Library” a name for many entities

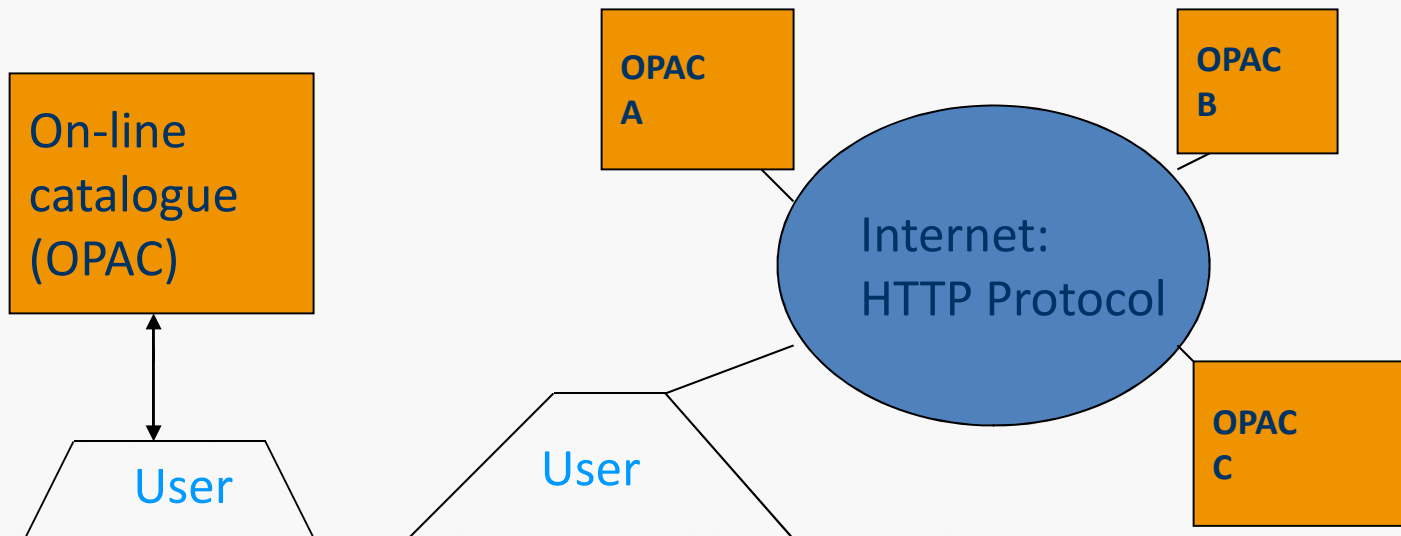
## From OPACs

- .....
- .....
- .....

## to Data Infrastructures

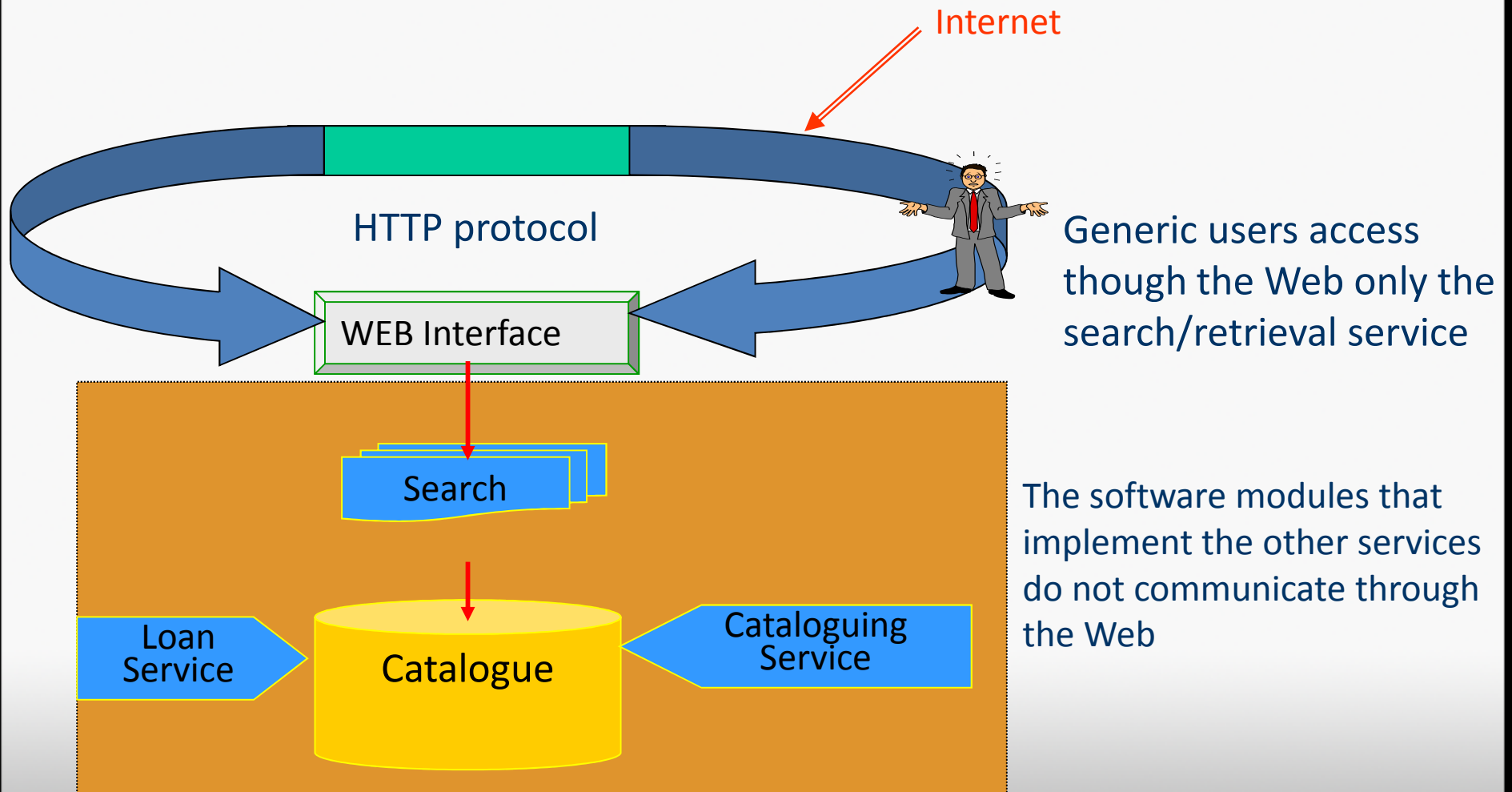
# Library systems in the past

from a “direct” communication  
(on-line) ....



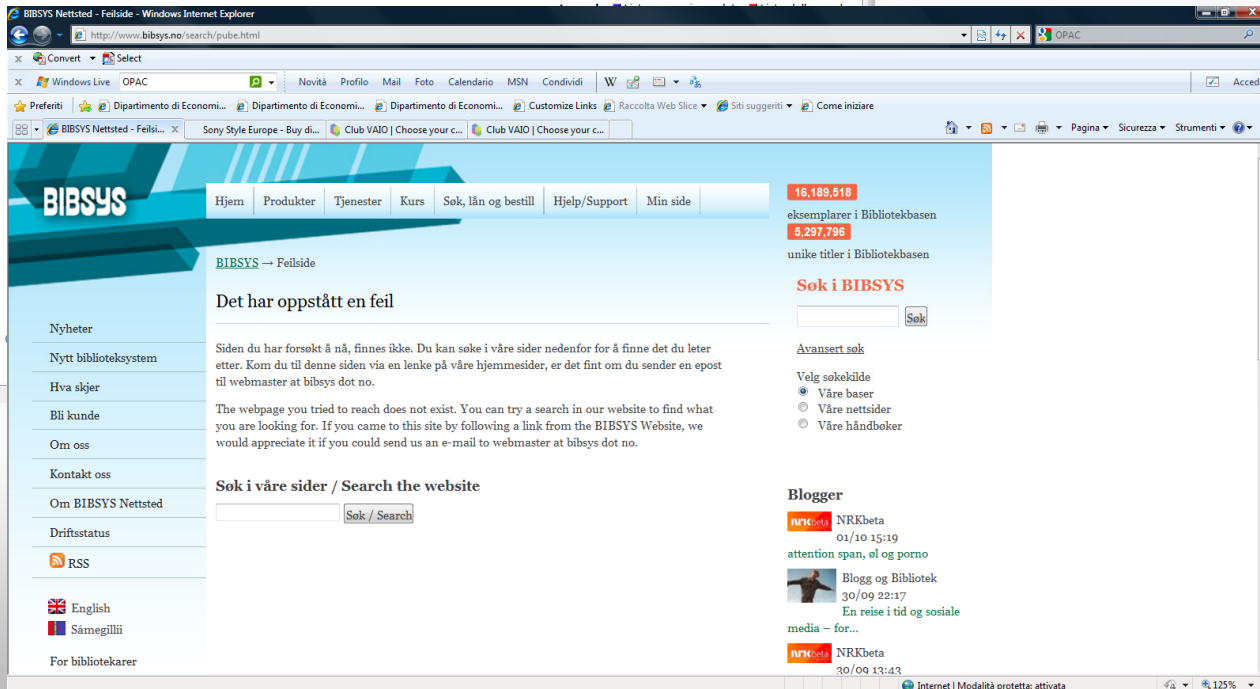
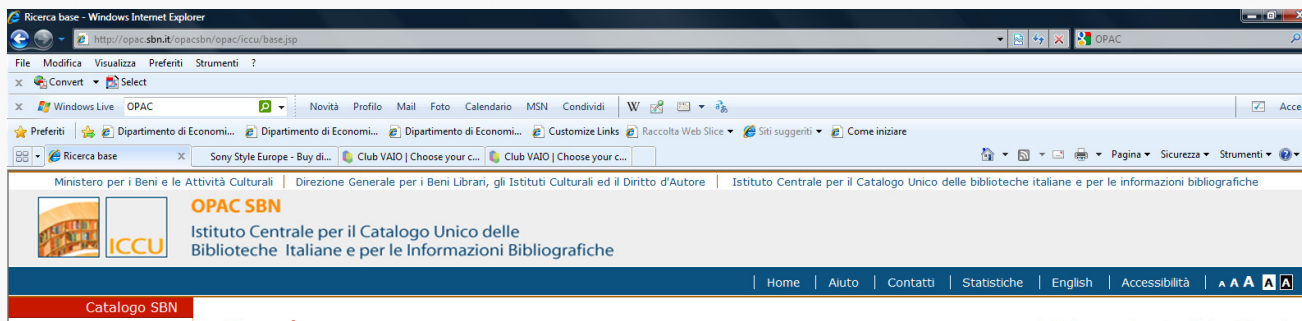
.... to a communication via  
WEB: HTTP protocol

# Web Access to OPACs



## Web Access to OPACs

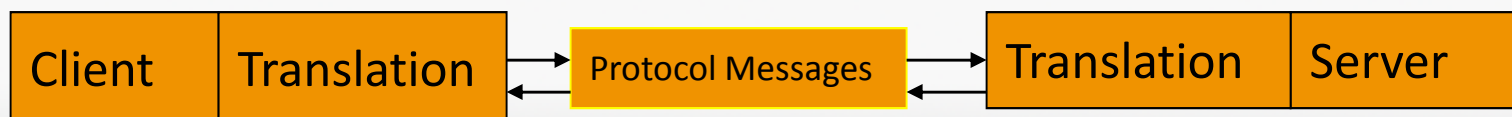
- Each catalogue is accessible through its own user interface
- User interfaces differ for:
  - access points
  - names of the access points
  - language
  - graphics
  - .....
- Users must be familiar with many interfaces
- No cross-searches are possible



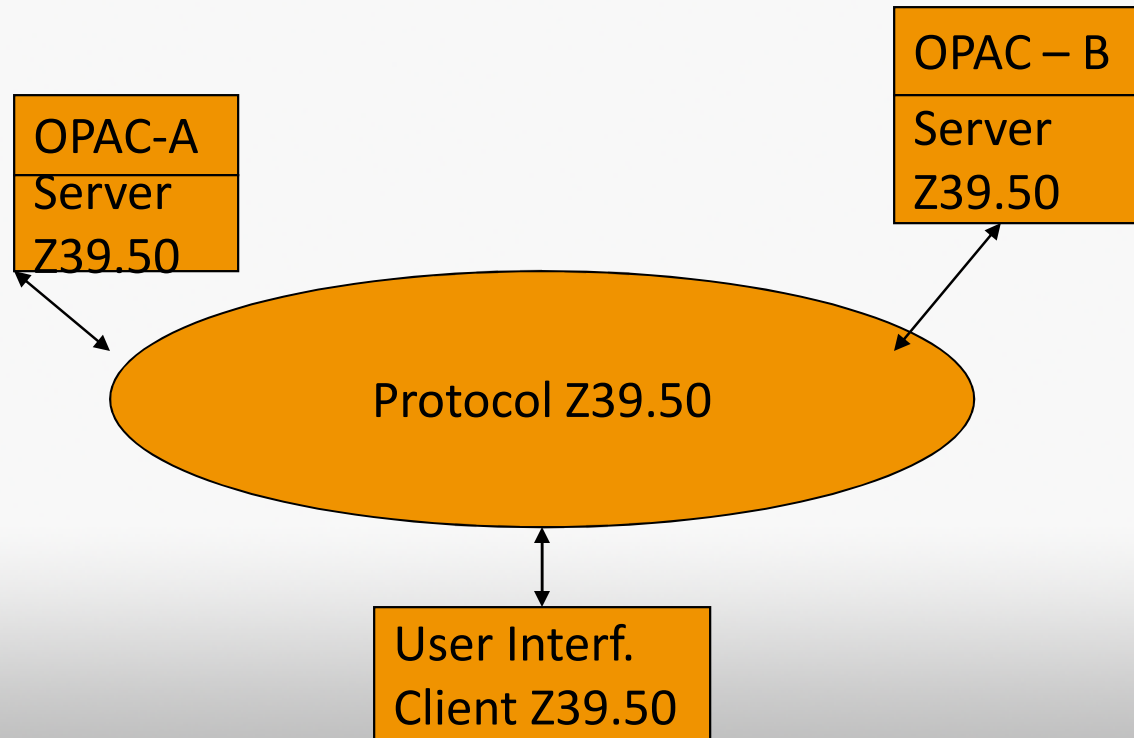


## Standard communication protocol for information search and retrieval

It regulates the communication between the clients and the servers (automatic catalogues)



It allows the search and retrieval of bibliographic information from different distributed OPACs by issuing a single query through a common user interface



- **SRW (Search & Retrieve Web Service)** draws heavily on the abstract models and functionality of Z39.50, but removes much of the complexity. SRW is built using common web development tools (WSDL, SOAP, HTTP and XML)
- **SRU (Search & Retrieve URL Service)** a URL-based alternative to SRW

# Digital Libraries come into life

## 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)

### Objective

*“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”*

## What is a DL ? (definition in 1998)

*“An institution which performs and/or supports (at least) the functions of a library in the context of distributed, networked collections of information objects in digital form”*

Nicholas Belkin  
Eight DELOS Workshop  
Stockholm, 1998

## Digital objects

Not only **document descriptions** (metadata)  
but also

**documents** (texts, videos, sounds, 3D  
images, ..... data, programs, ... maps, .... )

Not only a **catalogue....**

but also a **repository**

# DL basic functionality

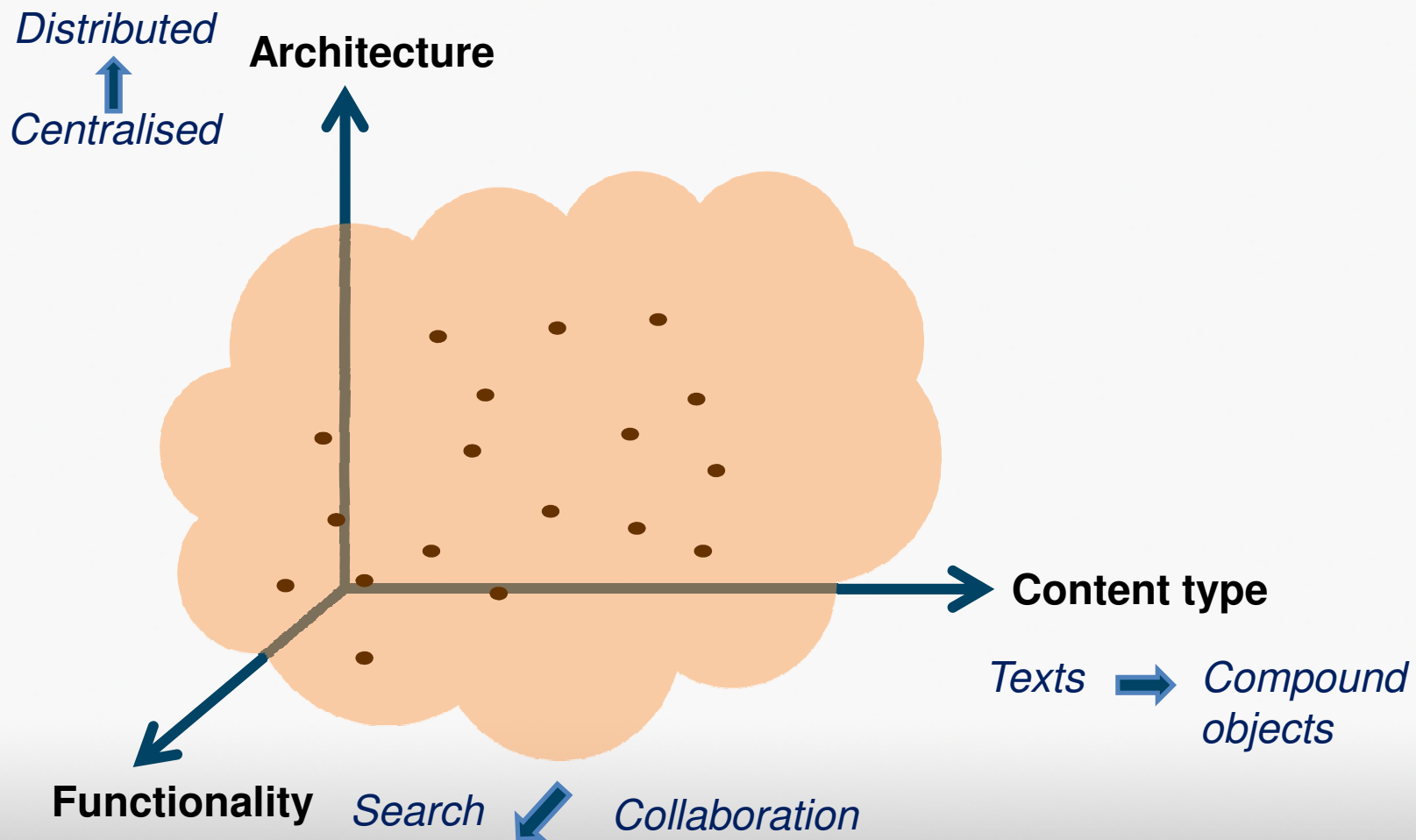
- Acquisition
- Submission
- Repository
- Search/Browsing/Retrieval
- Access
- .....

## A digital library is thus the analogous of...

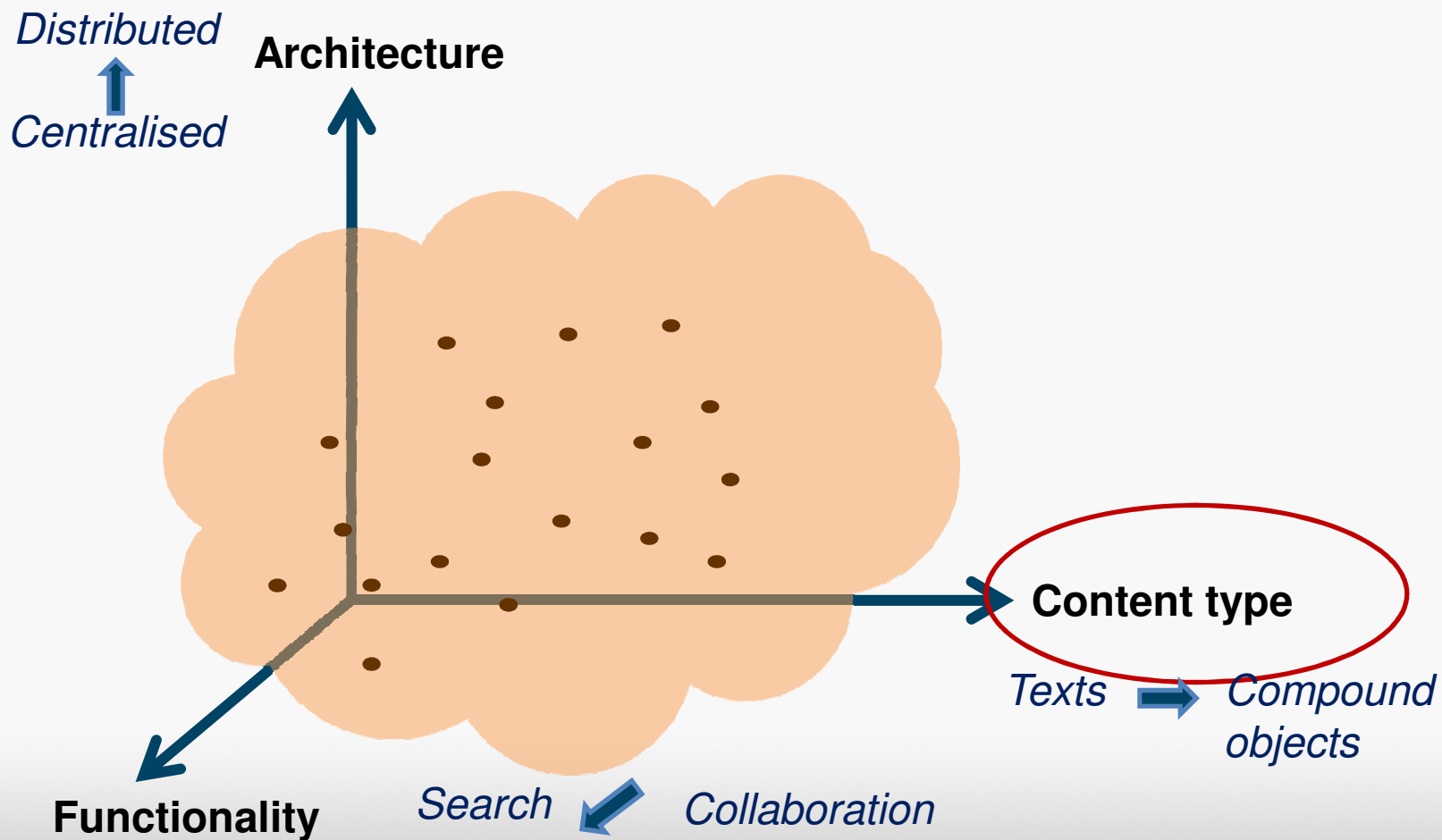
- A digital library
- A digital museum
- A digital archive
- A digital audio-video archive
- ...
- A data center
- .....



# Evolution Axes

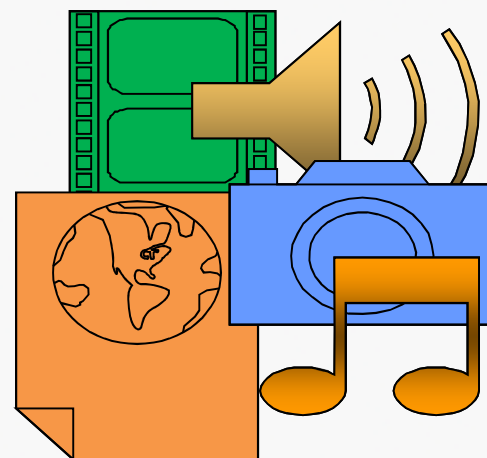


# Evolution Axes



## Content type

- Text objects
- Multilingual objects
- Multimedia objects
- Annotated objects
- Compound objects
- Enhanced publications
- .....



New document types empower novel forms of communication and remote collaboration among the members of a community of interest

# Multilingual documents

- 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



## Informedia DL system

**1994 – 2007** Centralized **audio-video DL system**

- **Documents**
  - Audio-video resources (mainly News)
- **Metadata**
  - Terms extracted from the transcript and from the image captions
  - Locations
  - Keyframe
  - Faces
  - Name of the speakers
  - Video abstract
  - .....

# Informedia DL system

- **Automatic content metadata extraction** through the integration of various technologies:
  - Speech understanding for automatically derived transcripts
  - Face, text and object recognition
  - Key frame extraction and indexing
  - Geo-coding
  - Topic assignment

## Informedia DL system (cont.)

- **Functionality**
  - Search based on
    - Free text
    - Image similarity
    - Face and object similarity
    - Geographical information
  - Multiple presentation styles of query results

# Informedia DL system (cont.)

**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.*

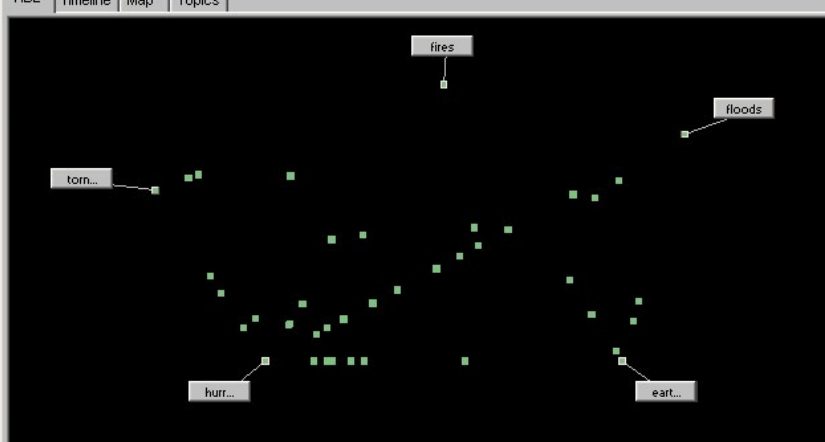
Prev. Page Next Page Go to Page... Visualize All...

Search Results (Page 1 of 32)



**Visualization of search results set containing 287 documents**

VIBE Timeline Map Topics



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 - 253)  Size-Code  Color-Code

fires  
 floods  
 earthquakes  
 hurricanes  
 tornadoes

Ellipse Re-Plot

Connect words

Show Doc. Info

Show Details...



# Informedia DL system (cont.)

**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 hurr  
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

Countries States



Documents with no map associations.

Visible Invisible Inactive

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 >

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

Within >

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 >

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

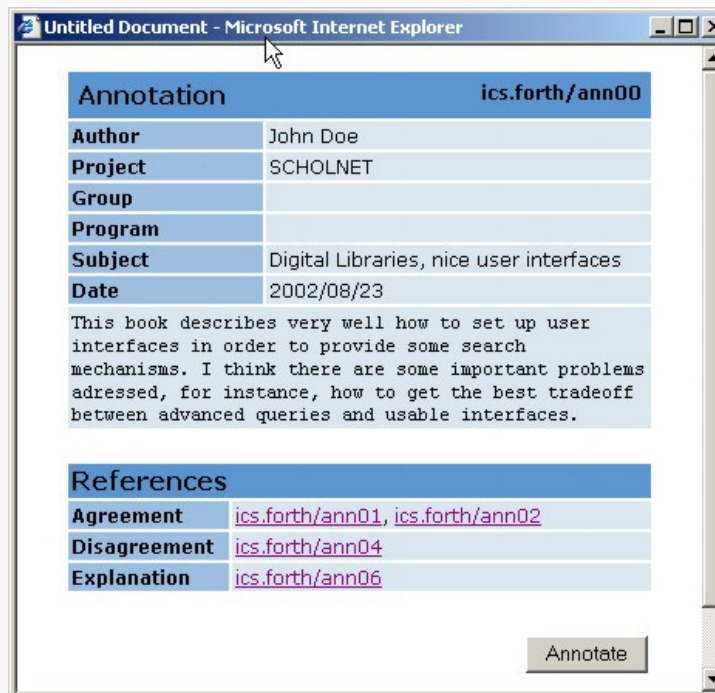
Deactivate selected locations

Show Doc. Info

Show Details...

# Annotated objects

- comment
- rating
- description
- link
- .....



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 addressed, 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

# Compound documents

## Journal

Article1  
Article2  
Article3



## Thesis

Chapter1  
    SectionA  
    SectionB  
Chapter2  
    SectionC



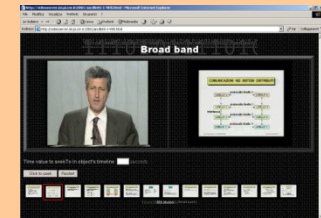
## Video

Sequence1  
    SceneA  
        FrameA1  
        FrameA2  
    SceneB  
        FrameB1  
Sequence2



## Tutorial

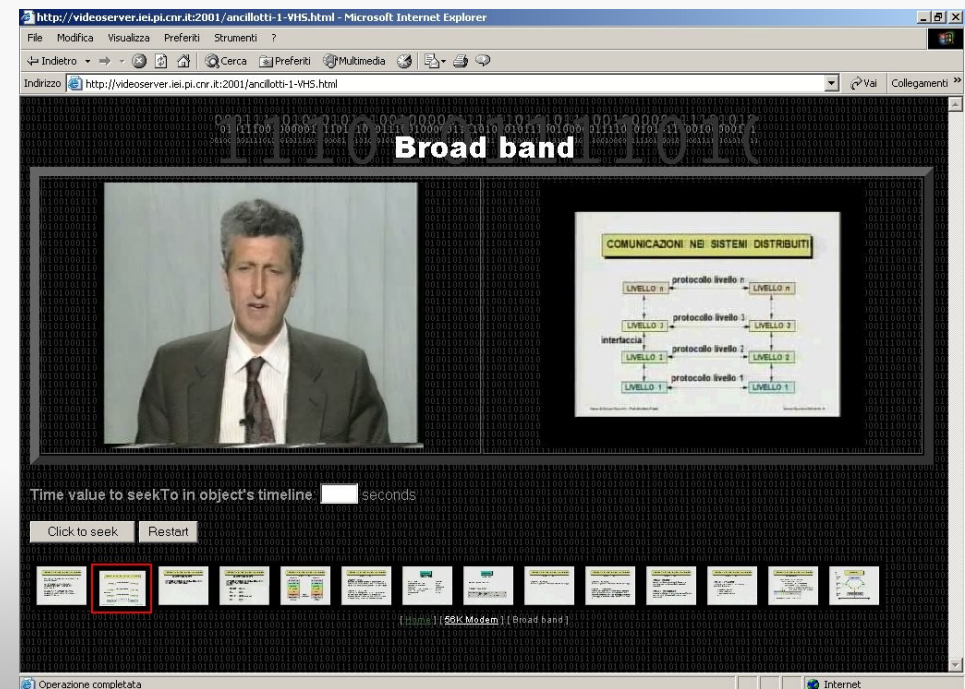
Introduction  
Part1  
Part2  
Part3



# Multimedia compound objects

e.g. DL.org Autumn school – October 2010

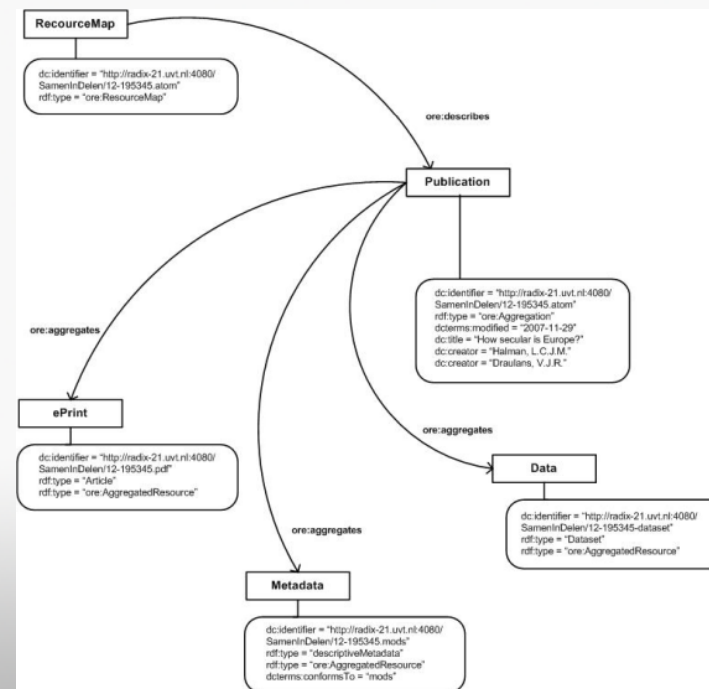
- Sequence of lectures
  - Videos of the lecture
  - Slides
  - Demos
  - Related documentation



# Enhanced publications

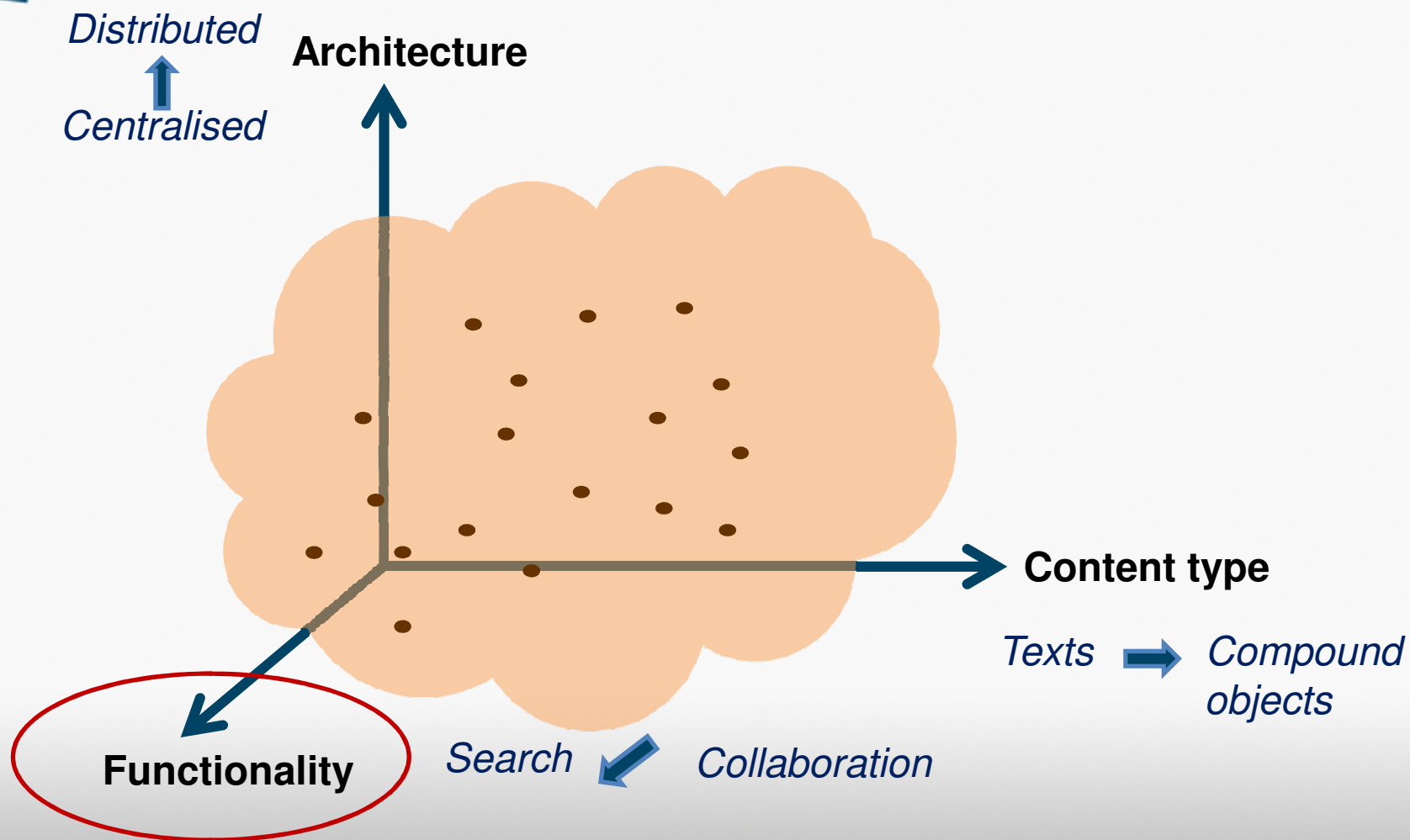
An enhanced publication is a publication that is enriched with three categories of information:

- **research data** (evidence of the research)
- **extra materials** (to illustrate or clarify)
- **post-publication data** (commentaries, ranking)





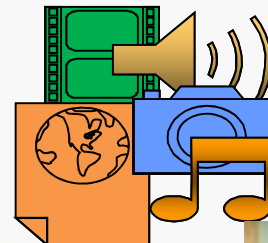
# Evolution Axes



# DL functionality (1)

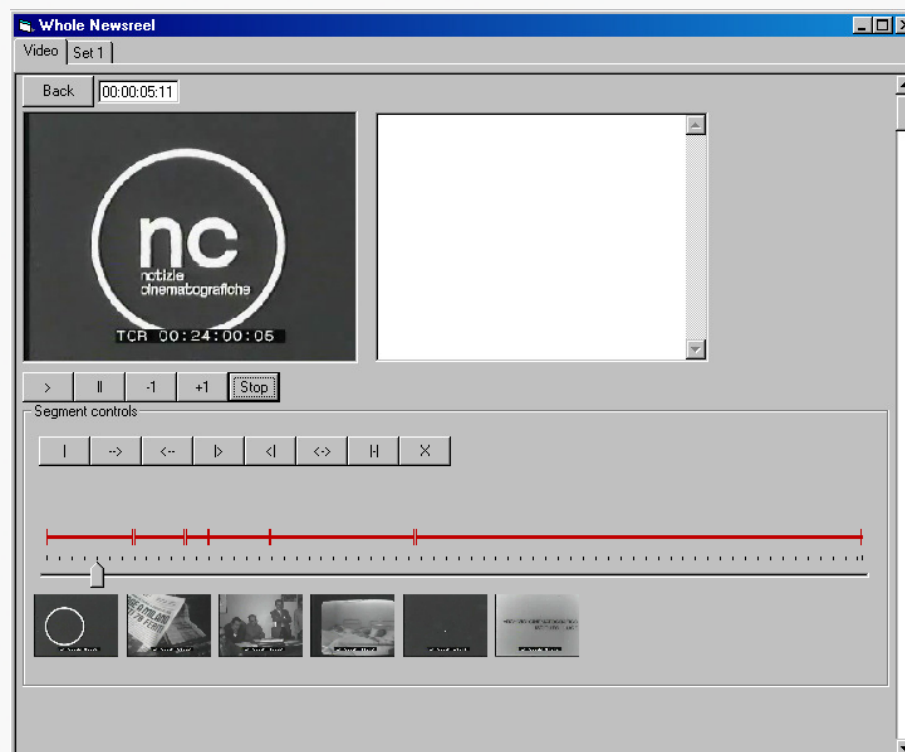
- New document types impose a re-thinking of the “traditional” library services

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





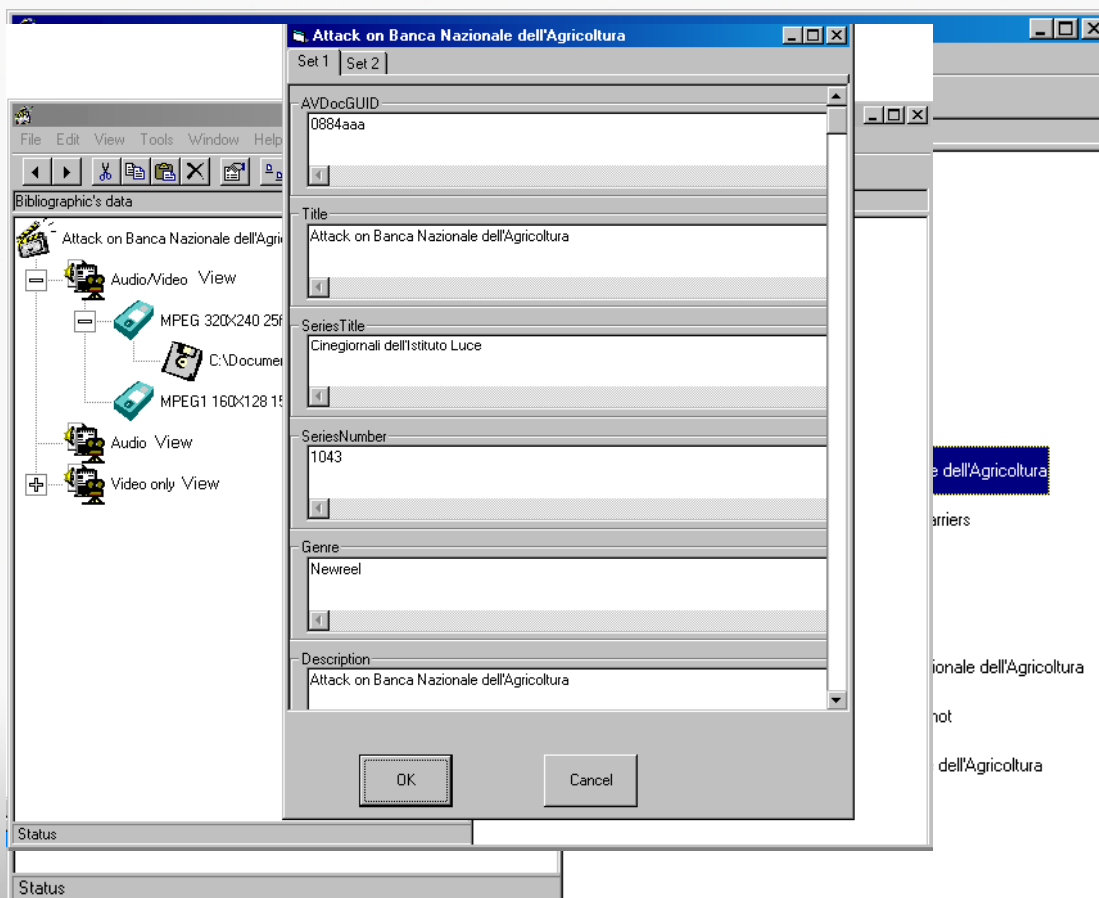
## e.g. the acquisition of video documents



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

# e.g. the description of video documents

... and describe the video and its parts separately



**e.g. search**

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

## DL functionality (2)

A DL is not only a instrument for a

- wider
- cheaper
- faster

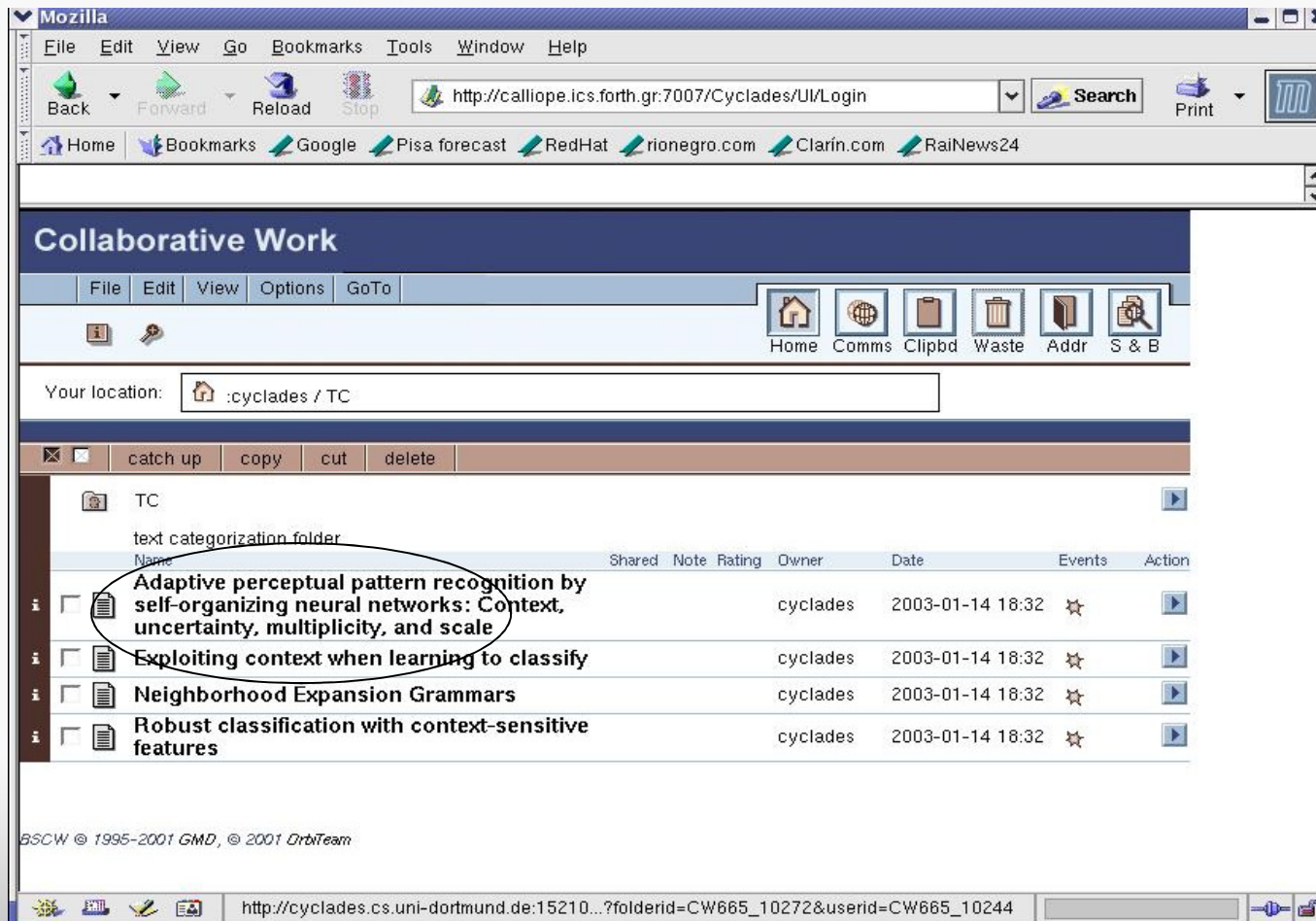
dissemination of information but it can also be a mean for supporting the **communication and collaboration** between the members of a community of interest

## Other DL Services

New services (not necessarily document centered) can be included in a DL to improve its potential usages

- Recommenders
- Co-operative work services
- Peer-reviewing supporting services
- Shared authoring services
- Social networking
- ....

# Cyclades collaborative environm.



Collaborative Work

File Edit View Options GoTo

Home Comms Clipbd Waste Addr S & B

Your location:

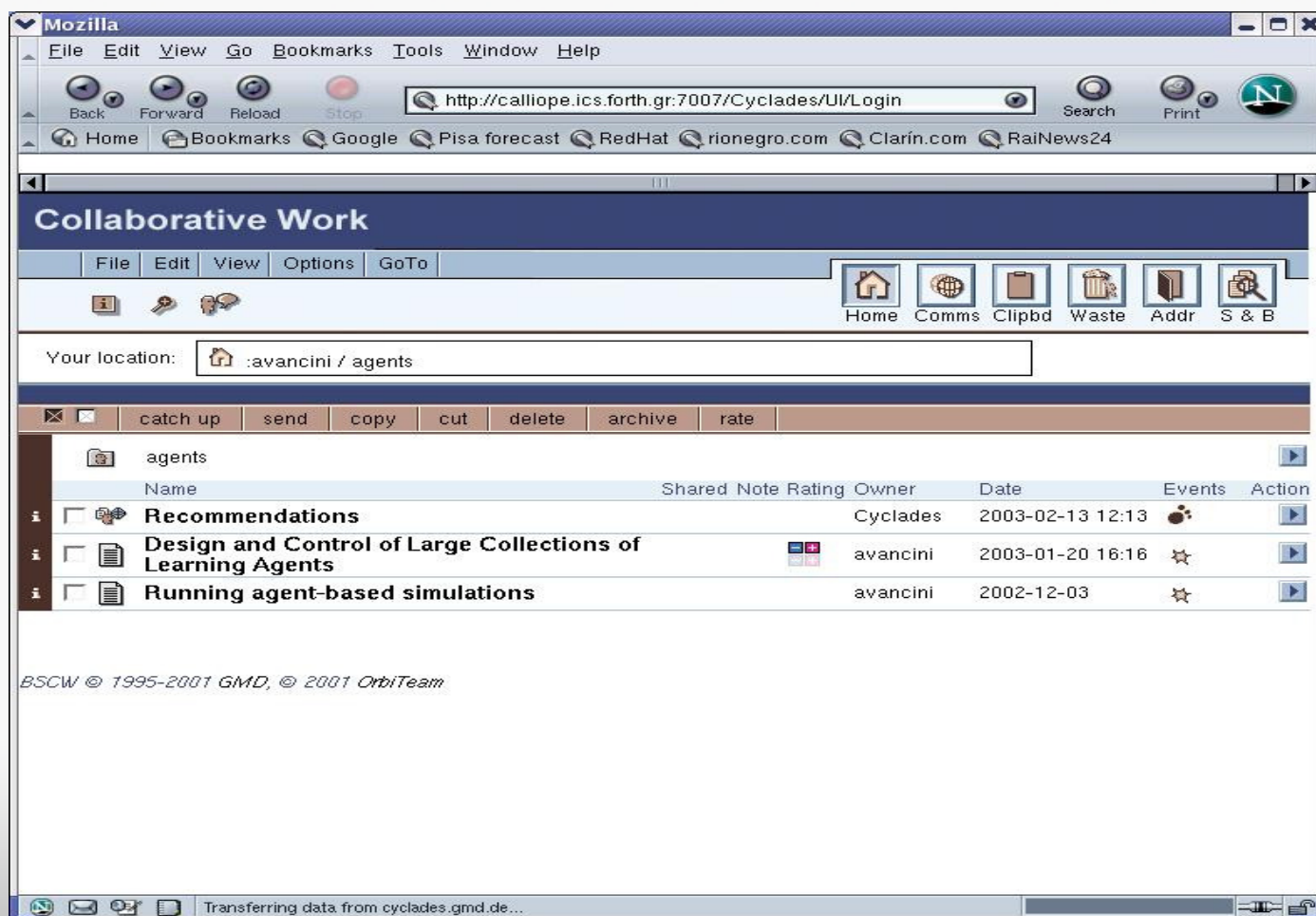
catch up copy cut delete

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

BSCW © 1995-2001 GMD, © 2001 OrbiTeam

http://cyclades.cs.uni-dortmund.de:15210...?folderid=CW665\_10272&userid=CW665\_10244

## Collaborative environm.(cont)



Collaborative Work

File Edit View Options GoTo

Home Comms Clipbd Waste Addr S & B

Your location:

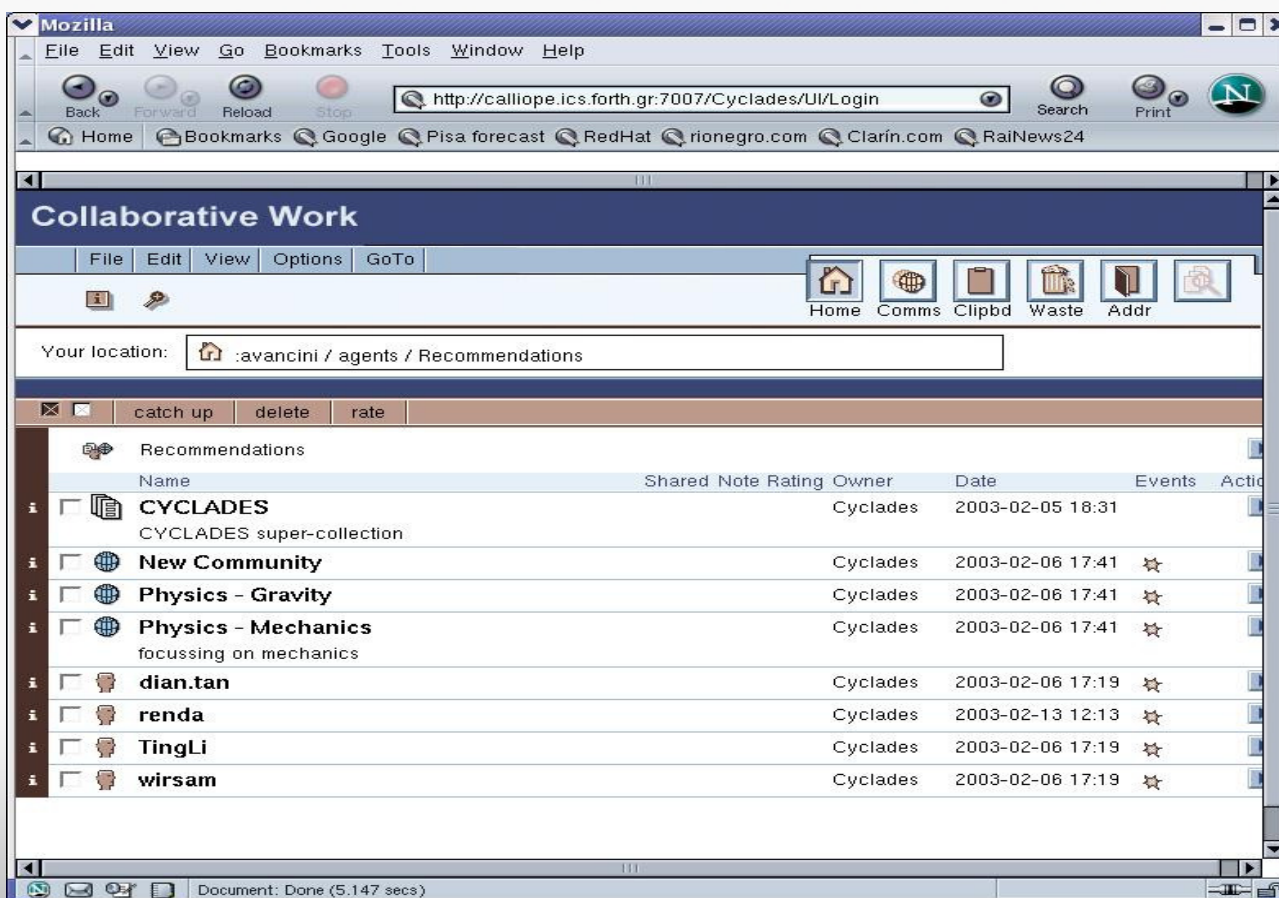
catch up send copy cut delete archive rate

Name	Shared	Note	Rating	Owner	Date	Events	Action
<b>Recommendations</b>				Cyclades	2003-02-13 12:13		
<b>Design and Control of Large Collections of Learning Agents</b>				avancini	2003-01-20 16:16		
<b>Running agent-based simulations</b>				avancini	2002-12-03		

BSCW © 1995-2001 GMD, © 2001 OrbiTeam

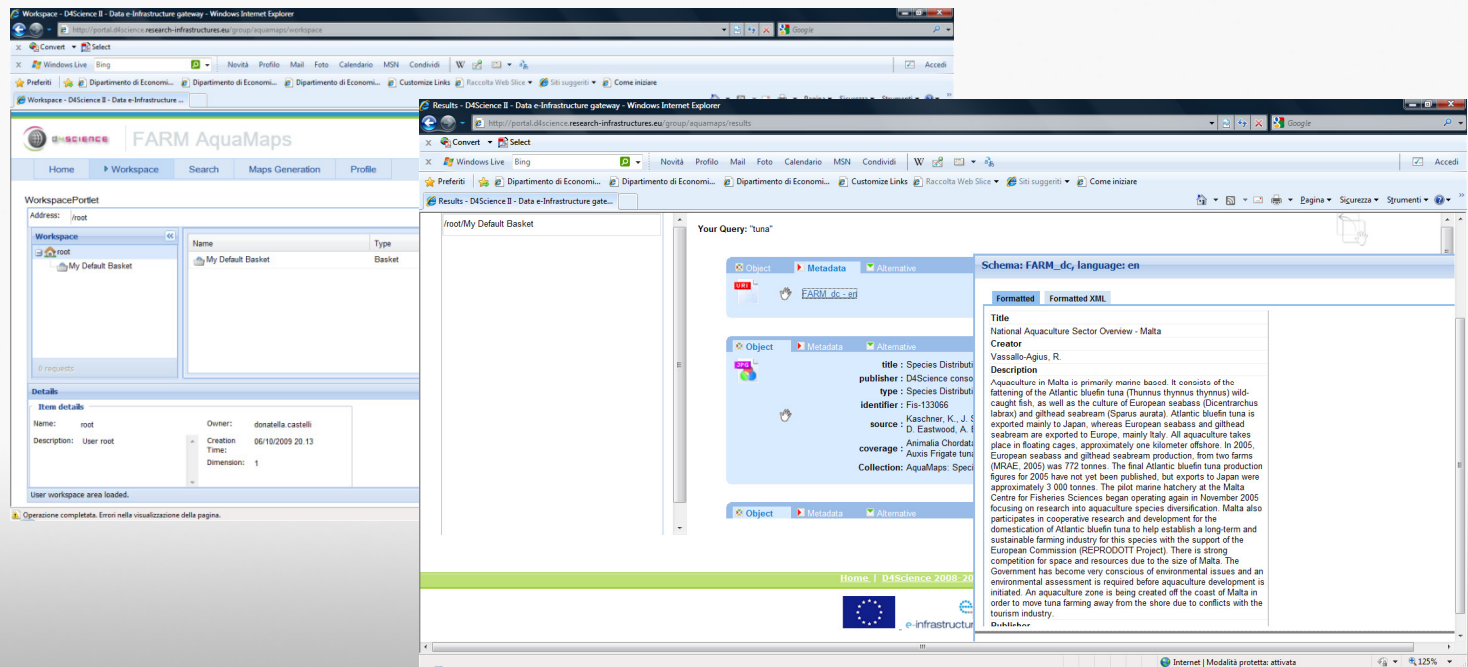
Transferring data from cyclades.gmd.de...

# Cyclades Recommender system (cont.)



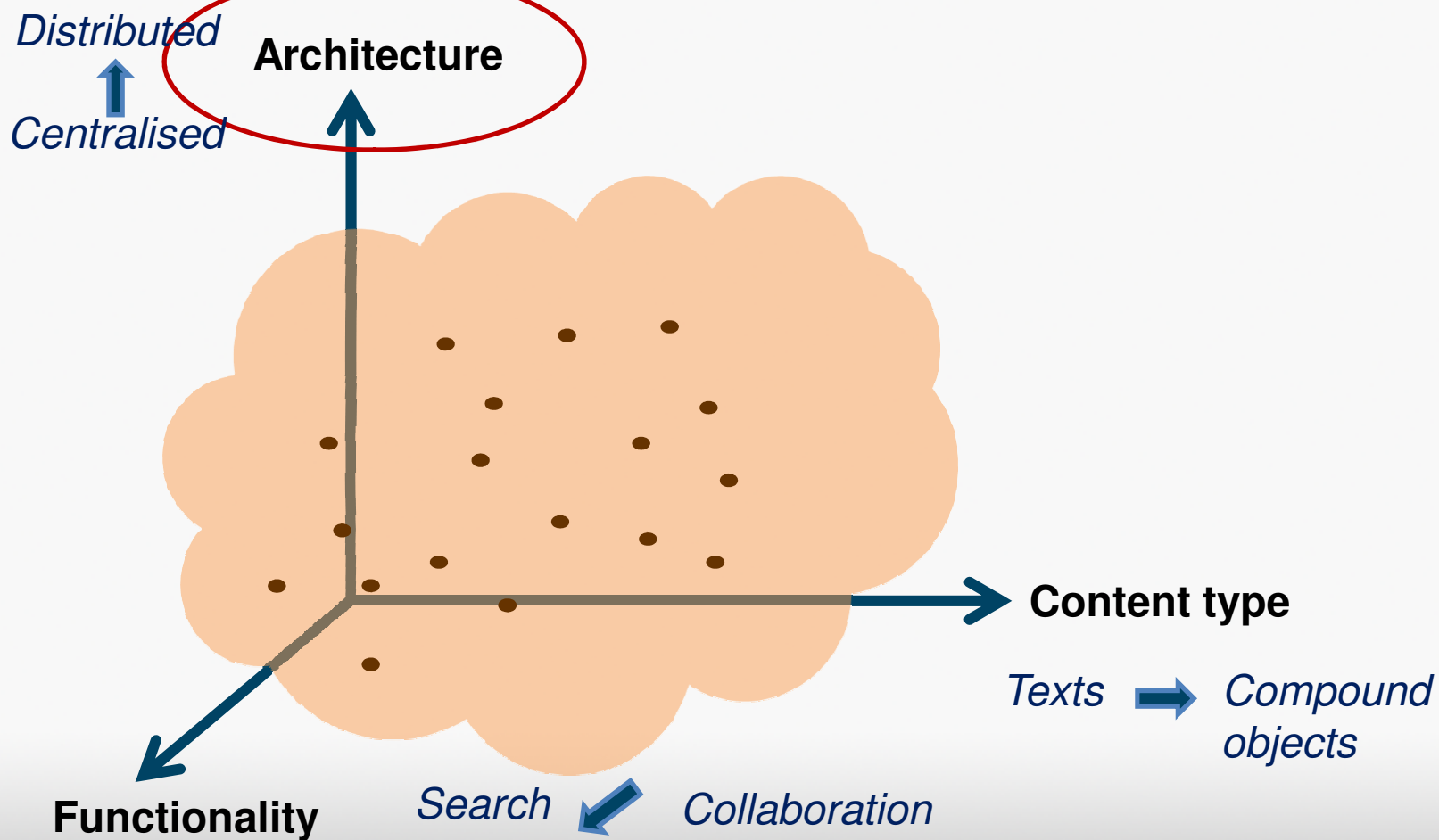


Integrated environment for the generation of products (through data processing), sharing and generation of new documents



The image displays two screenshots of the D4Science FARM AquaMaps web application. The left screenshot shows the 'Workspace' interface, which includes a 'WorkspacePortlet' with a tree view showing a 'root' directory and a 'My Default Basket'. Below this, 'Item details' are shown for the 'root' item, including the owner 'donatella.castelli', creation time '06/10/2009 20:13', and dimension '1'. The right screenshot shows the 'Results' page for a query 'tuna'. It features a search bar, a list of results with columns for 'Object', 'Metadata', and 'Alternative', and a detailed view of the selected result. The detailed view includes a 'Schema: FARM\_dc, language: en' and a 'Formatted XML' section with fields like 'Title', 'Creator', 'Description', 'title', 'publisher', 'identifier', 'source', 'coverage', and 'Collection'. The description text discusses the aquaculture of Atlantic bluefin tuna in Malta, mentioning the Malta Centre for Fisheries Sciences and the REPRODOTT Project.

# Evolution Axes



# Centralised systems



- Store
- Curate
- Search
- Access



- Submit
- Annotate
- Recommend
- Collaborate
- Specialised analysis & processing
- .....

➤ Single-type document

## Cost of centralised DLs

- Digital content production & curation
- System development
- Hw & sw maintenance
- Workload peaks coverage



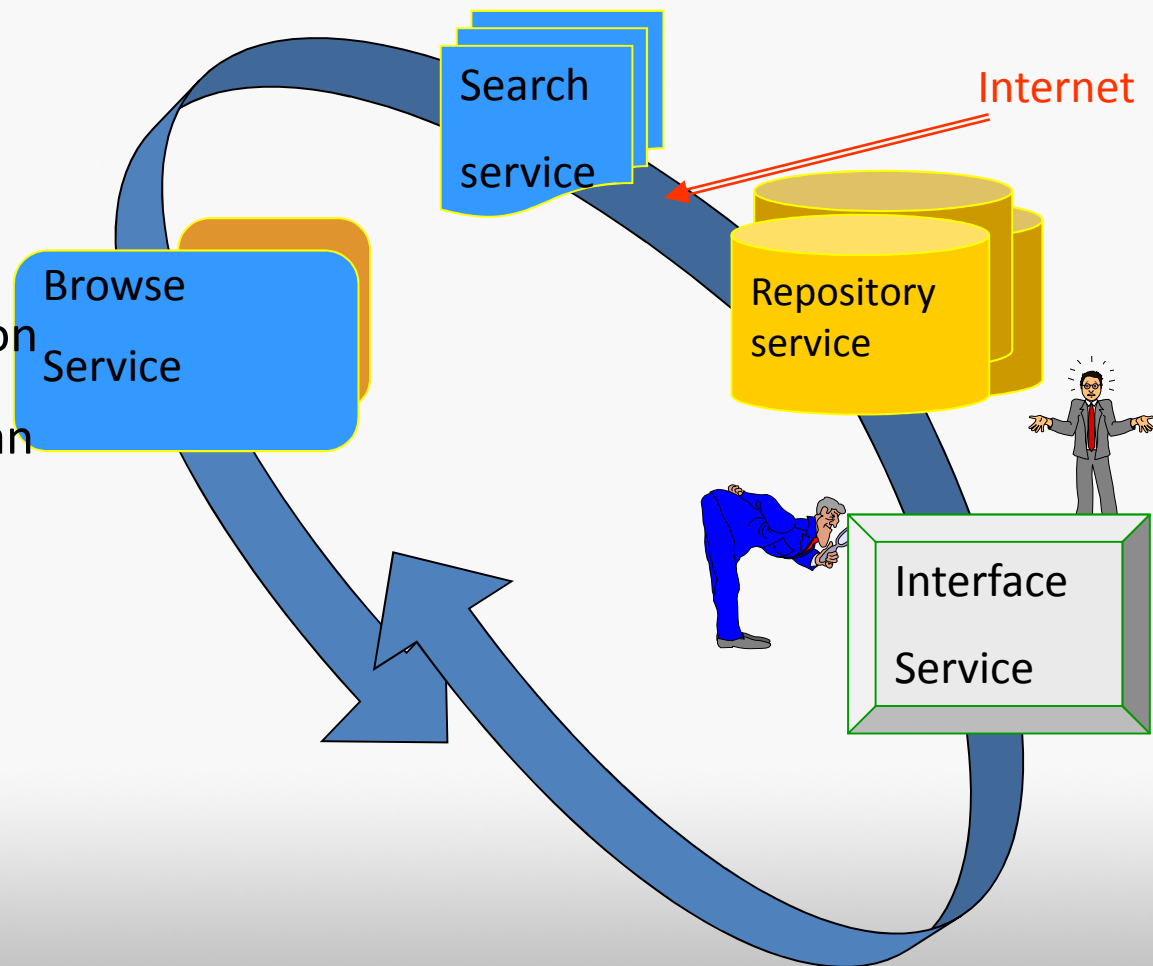
# Centralised vs distributed

## Distributed systems

The services are distributed on the Internet.

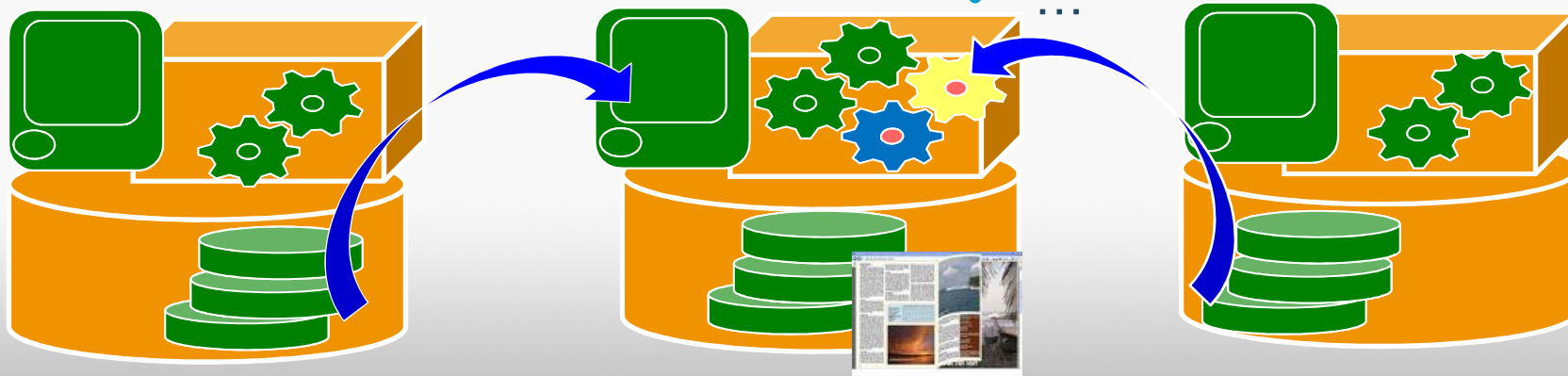
They communicate through an established protocol.

Users access the system through a Web interface



# Federated DL

- Cross-access
- Metadata & document mapping & harmonization
- Metadata cleaning
- Policy control



## Cost of federated DLs

- Digital content production & curation (reduced)
- System development
- Hw & sw maintenance
- Workload peaks coverage
  
- Interoperability
  - Metadata & object structure
  - Contextual information
  - Provenance
  - Policy
  - Quality
  - ...



1996 - ERCIM\* Technical Reference DL

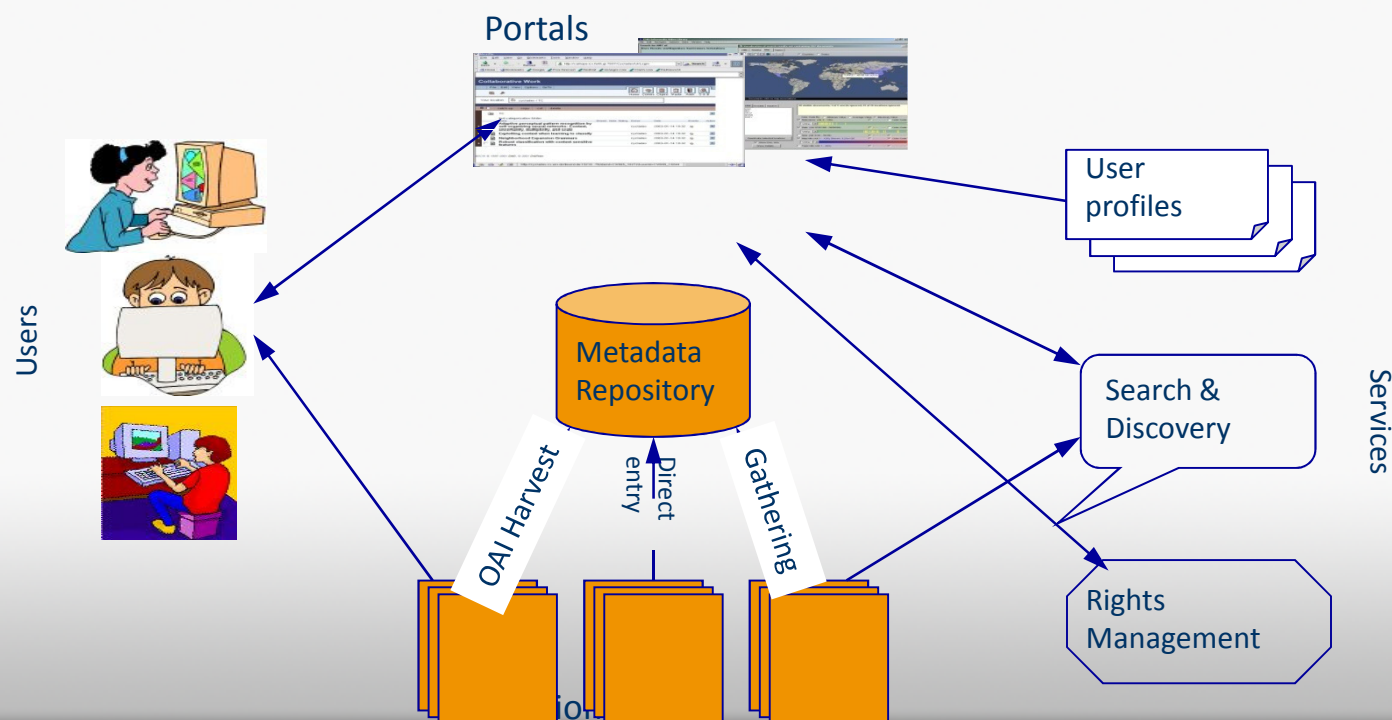
Access to the content of ERCIM institutional repositories



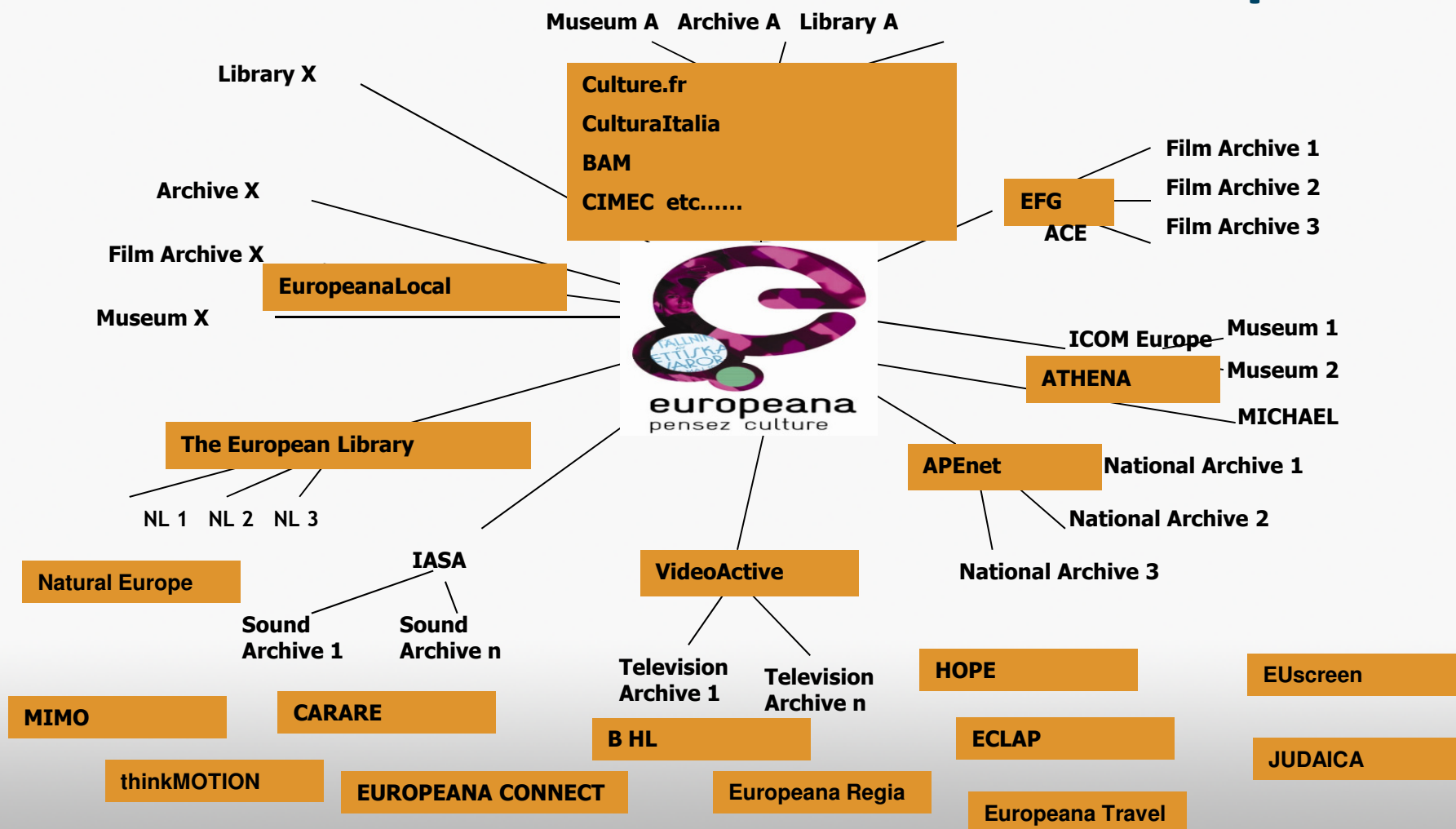
\*ERCIM =  
European Consortium for Informatics  
and Applied Mathematics



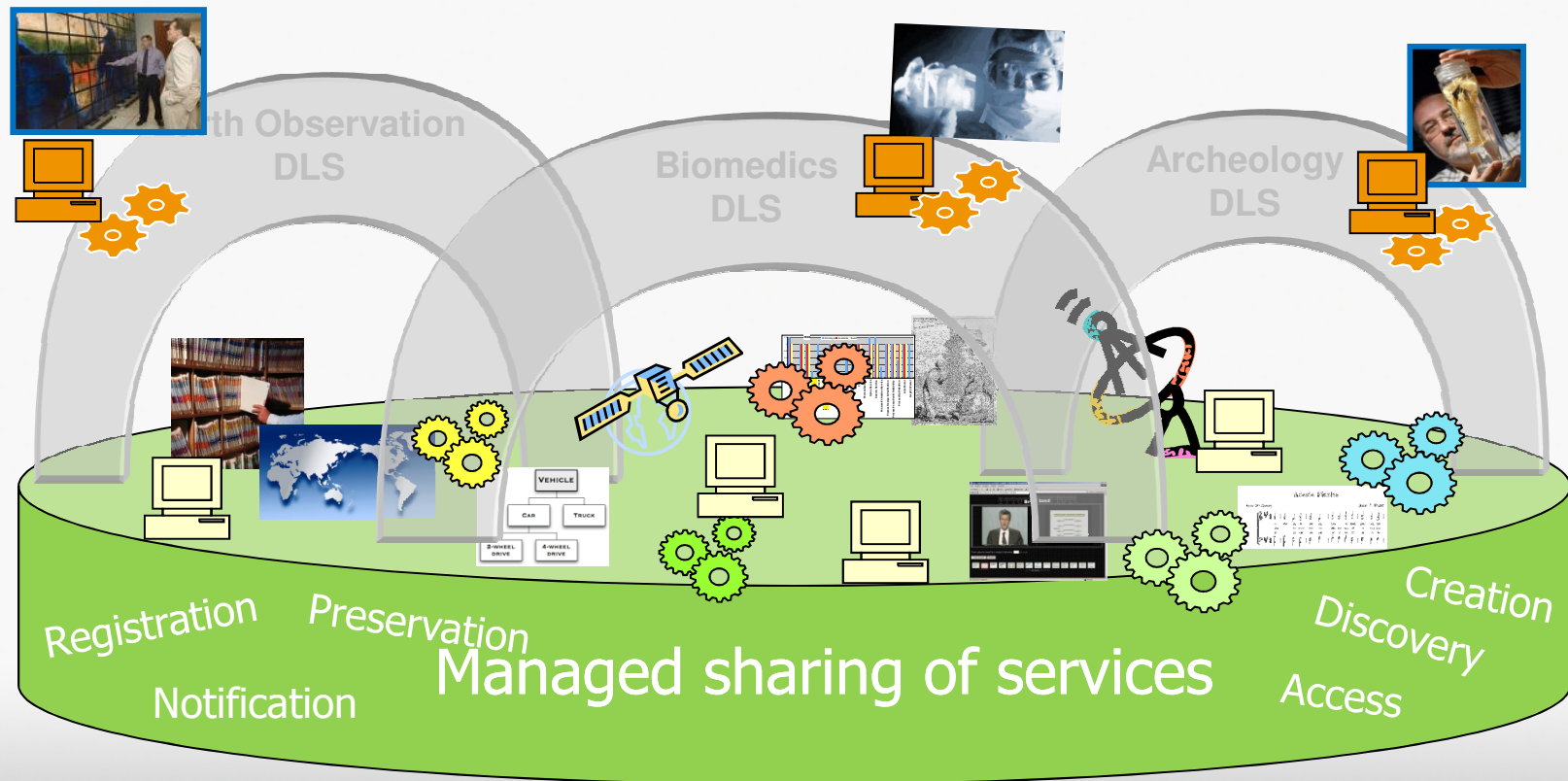
## US Nation's online library for education and research in Science, Technology, Engineering, Mathematics



# Large federations: Europeana



# Data Infrastructures



## Exemplars

- DRIVER
- OpenAIRE
- EFG
- HOPE

**D-Net**  
enabled



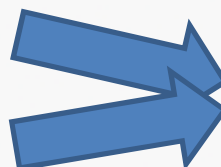
- D4Science
- gMan

**gCube**  
enabled



## Data infrastructure cost

- Digital content production & curation
- System development
- Hw & sw maintenance
- Workload peaks coverage



Covered by the e-Infrastructure



E-Infrastructure set-up and mgmt cost



# The DL Universe

DIENST

NDLTD

NSDL

ECHO

ACM DL

D4Science

DSPACE

DRIVER

OPENDLIB

FEDORA

DILIGENT

TEL

HOPE

EFG

# REFERENCE MODEL