CYCLADES

An Open Collaborative Virtual Archive Environment

www.ercim.org/cyclades

Umberto Straccia

I.E.I - C.N.R.

Pisa (ITALY)

Outline

- Info sheet
- Objectives
- Functionality
- Architecture
- Milestones

An Open Collaborative Virtual Archive Environment

CYCLADES

Info sheet

- Starting date: February,1st 2001
- Duration: 30 Months
- Effort: 214.8 p/m
- Budget
  - Total cost: 2,150,000 E
  - EU funding: 1,200,000 E
- Actual State: Month 15 (50%)

Participants

- IEI-CNR (Italy) - Scientific coordinator
  - Umberto Straccia
- UNIVERSITY OF DORTMUND (Germany)
  - Norbert Fuhr
- FORTH (Greece)
  - Dimitris Plexousakis
- FRAUNHOFER FIT (Germany)
  - Tom Gross
- ERCIM (France) - Administrative coordinator
  - Remi Ronchaud

Objectives

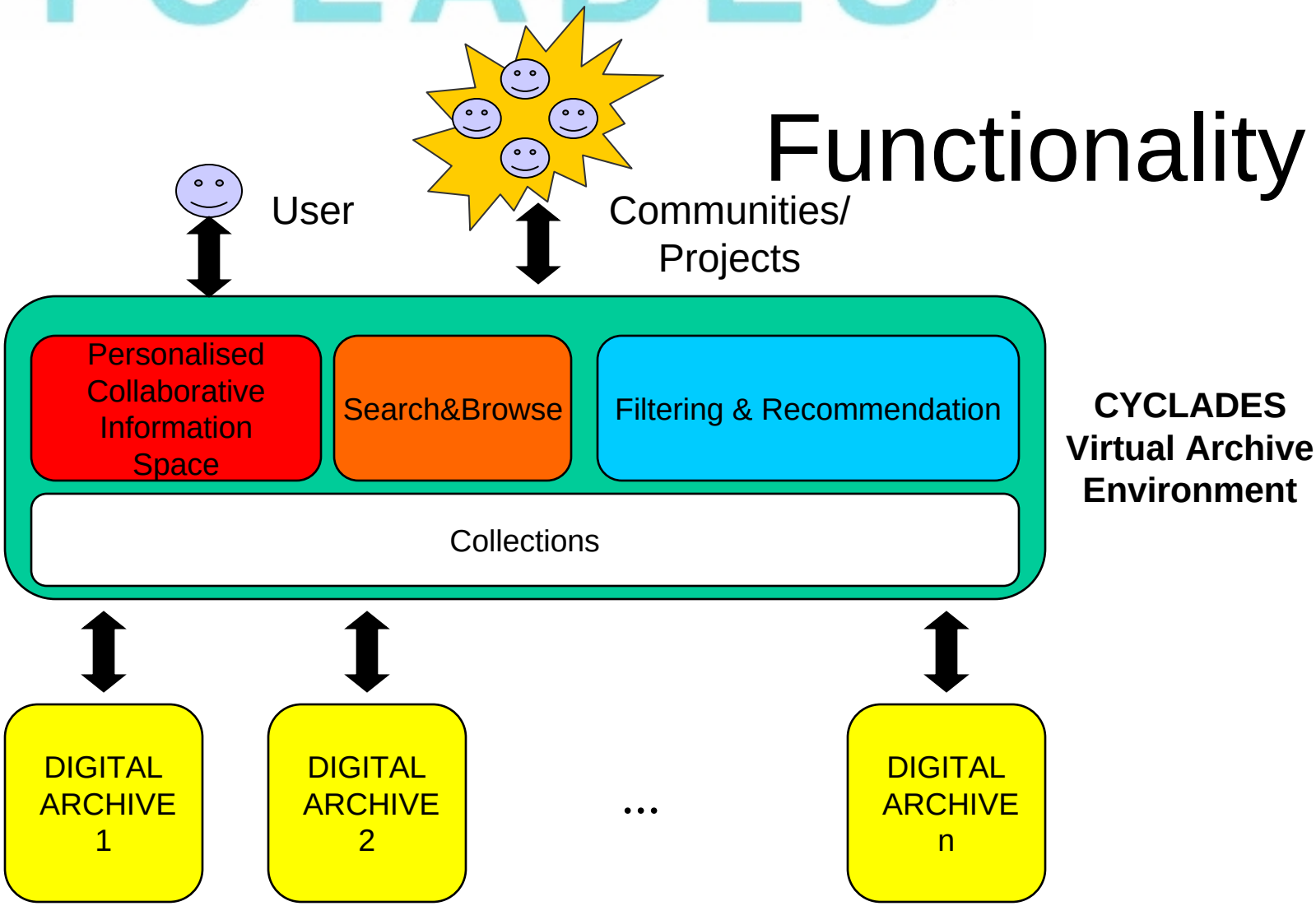
- Development of a system which provides *an open collaborative virtual archive environment*
- This environment will support
  - **Users**
  - **Communities** (and their members)
  - **Projects** (and their members)

  with functionality for advanced
  - ***Search*** in large, heterogeneous, multidisciplinary digital archives
  - ***Personalised Information Space Organisation***
  - ***Collaboration***
  - ***Filtering*** and ***Recommendation***

An Open Collaborative Virtual Archive Environment

CYCLADES

Functionality



Large, heterogeneous, multidisciplinary digital archives

The Digital Archives

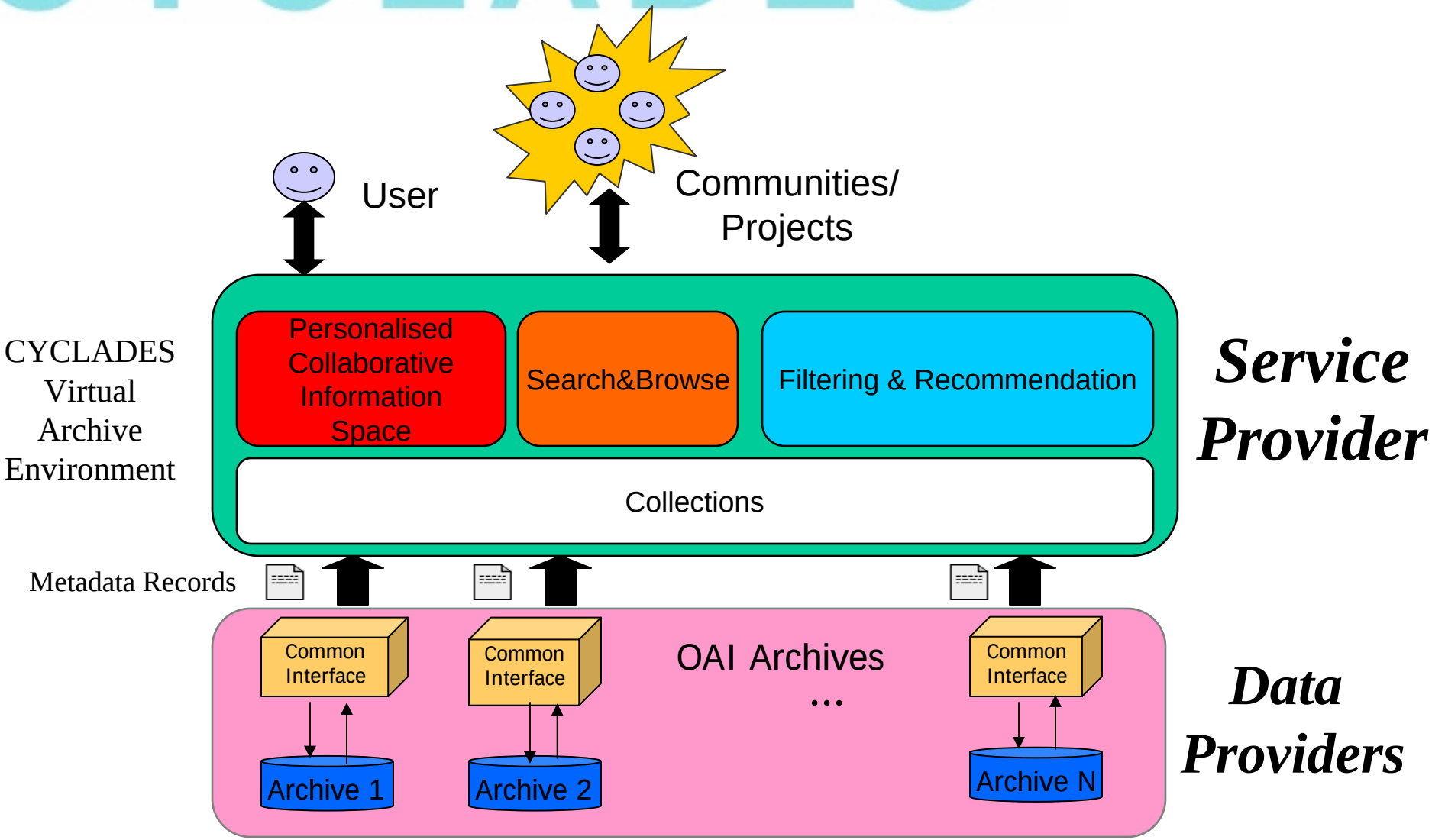
- CYCLADES will provide access to the Digital Archives compliant with the Open Archives Initiative (OAI)
  - CYCLADES will Harvest and index records from these archives

FOR MORE INFO...

http://www.openarchives.org

An Open Collaborative Virtual Archive Environment

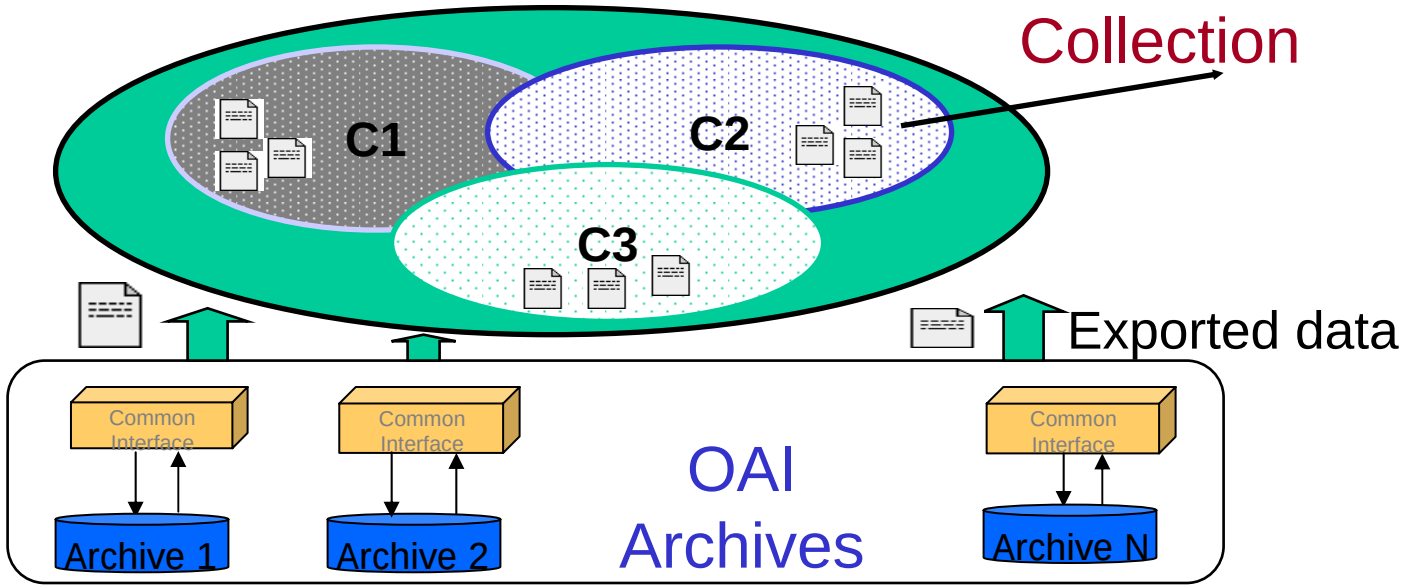
CYCLADES



Large, heterogeneous, multidisciplinary digital archives

Collections

- ***Collection***: a set of metadata records collected together, according to a record selection criteria, from the archives
- Collections are NOT materialized
- Collections represent the information space, while the OAI archives remain hidden to the users and communities
- ***Users* and *Communities* may define their own collections, e.g. by means of meet, join and refinement of existing ones**

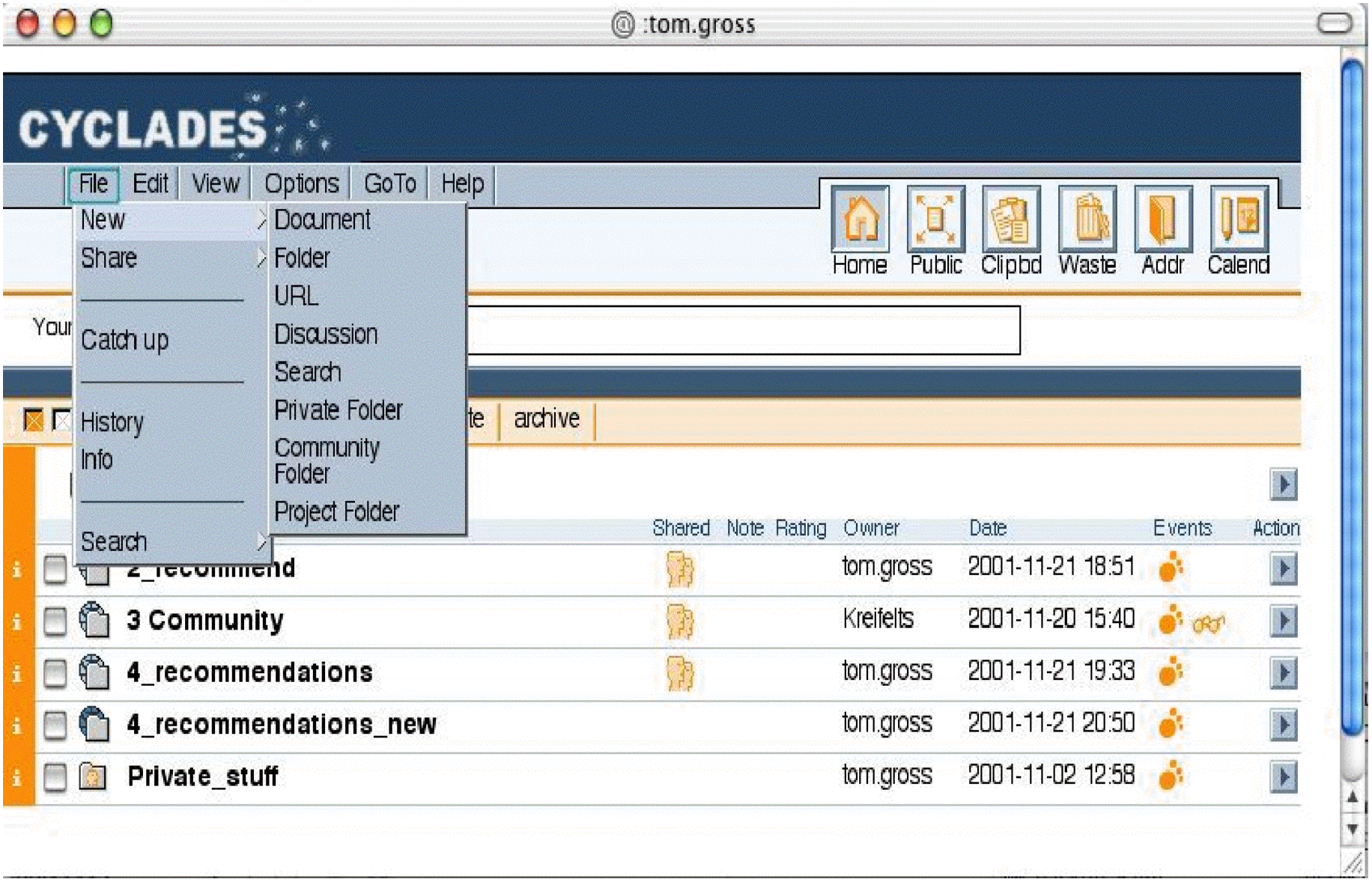


Personalised, Collaborative Information Space

- Users and communities may organize their own information and working space in terms of *folders*
- ***Folders***:
  - May be organized hierarchically
  - May contain
    - Metadata records (retrieved from the OAI archives)
    - Uploaded user documents
    - System recommendations (of: users, records, communities, projects)
    - Hyperlinks
    - Annotations
    - User ratings
    - Discussion forums
  - May be *shared* among community/project members
    - Members of the same community/project may share the same folder content

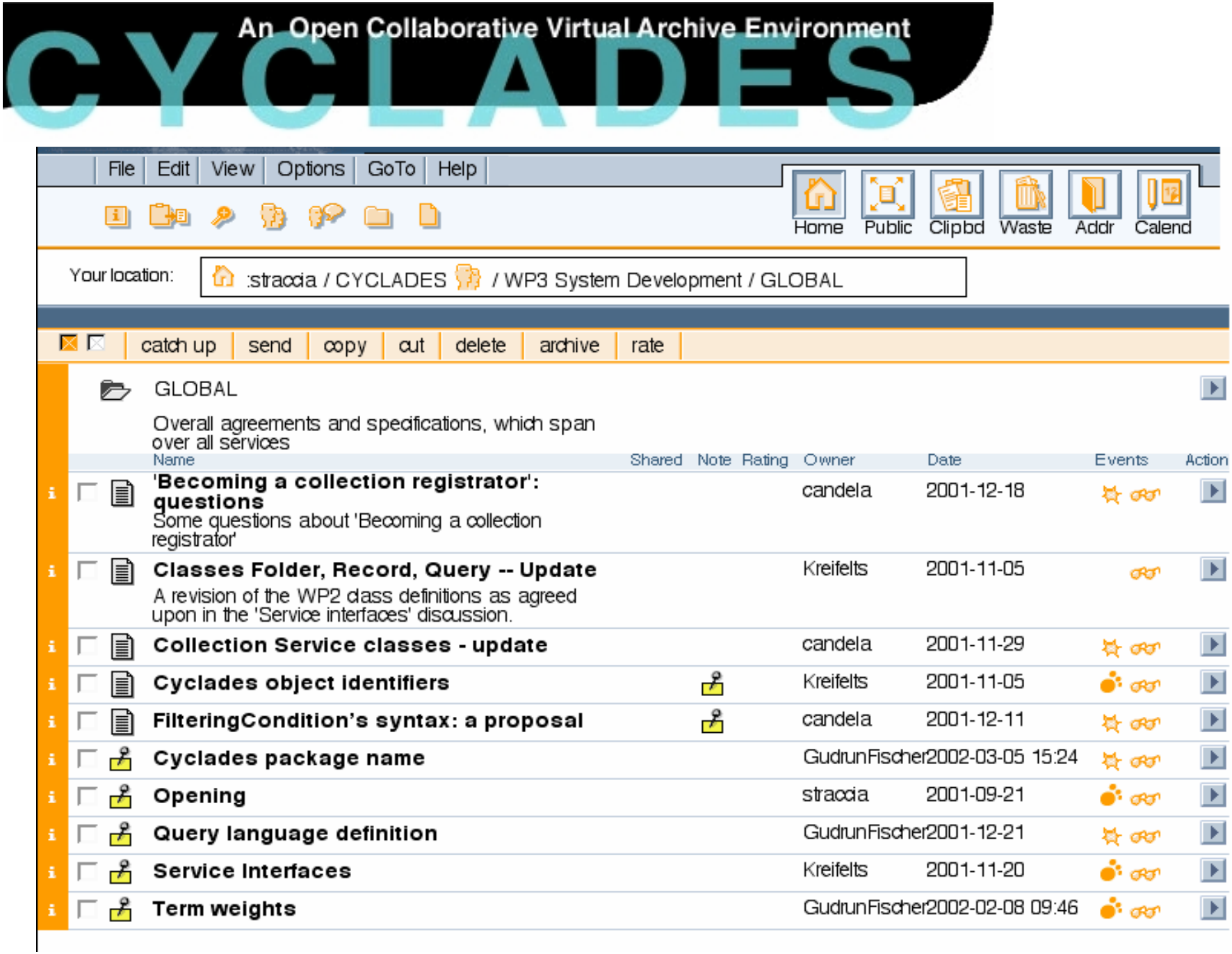
An Open Collaborative Virtual Archive Environment

CYCLADES



- ***Operations on folders***:
  - Create, Copy, Move folder
  - Edit folder properties

- ***Operations on items in folders***:
  - Documents:
    - Read, Delete, Move, Copy, Versioning
    - Upload, Download documents
    - Annotate documents
    - Rate documents
  - Discussion forum
    - Read, Delete, Reply
  - Edit item properties
    - Access preferences, …


An Open Collaborative Virtual Archive Environment
CYCLADES
File
Edit
View
Options
GoTo
Help
Home
Public
Clipbd
Waste
Addr
Calend
Your location:
:straccia / CYCLADES
/ WP3 System Development / GLOBAL
catch up
send
copy
cut
delete
archive
rate
GLOBAL
Overall agreements and specifications, which span over all services
Name
Shared
Note
Rating
Owner
Date
Events
Action
'Becoming a collection registrator': questions
Some questions about 'Becoming a collection registrator'
candela
2001-12-18
Classes Folder, Record, Query -- Update
A revision of the WP2 class definitions as agreed upon in the 'Service interfaces' discussion.
Kreifelts
2001-11-05
Collection Service classes - update
candela
2001-11-29
Cyclades object identifiers
Kreifelts
2001-11-05
FilteringCondition's syntax: a proposal
candela
2001-12-11
Cyclades package name
GudrunFischer
2002-03-05 15:24
Opening
straccia
2001-09-21
Query language definition
GudrunFischer
2001-12-21
Service Interfaces
Kreifelts
2001-11-20
Term weights
GudrunFischer
2002-02-08 09:46

Search & Browse

- Users may *search & browse* the collections
- *Search*: typical Search using the collection's search fields, e.g. author, title, description: multiple schema browsing allowed
- Query formulation through browsing

Search

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

Browse

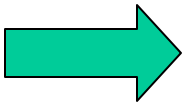
Schema | Attribute

Author	Castelli
Description	Straccia
Title	Thanos
...	...

- 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

Papers on Fuzzy Logic

Filtering & Recommendation

- Service learns the user information needs (*folder profile*) automatically from the user's folder content
- Uses of folder profiles
  - To be used by users to filter out irrelevant information during a search session: ***filtering of metadata records***
  - To be used by the system to automatically notify users about new data that the system judges relevant for them: ***recommendation of records, users, collections, communities***

Metadata records filtering

- *Personalised ad-hoc searching:*
  - Users formulates a query
  - User selects a folder
  - E.g. QUERY:=

    Collection = "*Logic*",
    Author = "*Straccia*",
    Description = "*Logic Programming*",
    Topic Folder = *"Papers on Fuzzy Logic"*

    "Find metadata records R in the collection **logic**, such that R refers to a document D written by **Straccia**, whose abstract is about **logic programming**, where R is relevant to my folder **Papers on Fuzzy Logic**"

- *Personalised on-demand searching:*
  - User selects a folder
  - E.g. QUERY:=

    Topic Folder = *"Papers on Fuzzy Logic"*

    "Find **new** metadata records relevant to my folder **Papers on Fuzzy Logic**"

Recommendations

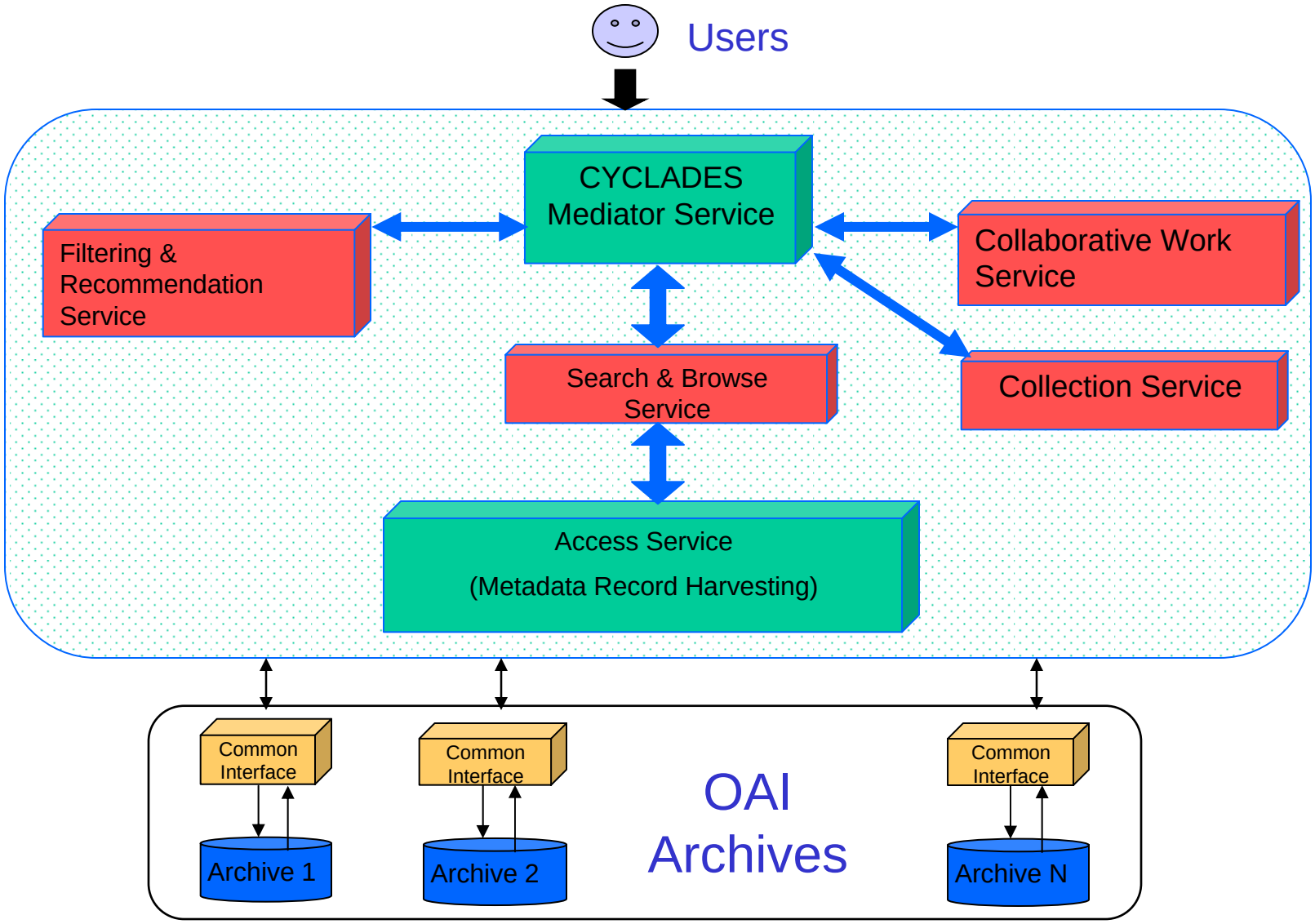
- All types of recommendations are relative to a user folder, i.e. user topic of interest
- *Metadata record recommendation:*
  - The system **recommends records**, based on folder profiles and user ratings (content-based collaborative filtering)
- *Collection recommendation :*
  - The system **recommends collections**, based on folder profiles, user ratings and collections' properties
- *User recommendation :*
  - The system **recommends users**, based on folder profiles, user ratings and users' properties
- *Community recommendation :*
  - The system **recommends communities**, based on folder profiles, user ratings and communities' properties

Architecture

- Modular
  - Basic functionality is allocated on servers
- Distributed
  - Servers may be over the Web
- Open
  - new services can easily be added
- Simple communication protocol
  - XML-RPC (SOAP will be evaluated)

An Open Collaborative Virtual Archive Environment

CYCLADES



Milestones

- User Requirements Collection and Analysis (April 2001)
- Functional and Architectural Specification (August 2001)
- Detailed Service Specification (February 2002)
- Service Prototypes (August 2002)
- System Integration and Testing (May 2003)
- Running Service (June 2003)

Summary

- CYCLADES will be a system providing advanced
  - Access to (at least) OAI compliant archives
  - Personalised collaborative working space
  - Filtering and recommendation functionality

Functionality:concepts

- Basic concepts in CYCLADES
  - Archive
  - Document
  - Metadata record
  - Metadata schema
  - Collection
  - User
  - Community
  - Project

Functionality:roles

- User roles
  - CYCLADES administrator
  - Archive administrator
  - Collection administrator
  - Single user
  - Community member
  - Community administrator
  - Project member
  - Project administrator

Functionality:actions

- **User actions**
  - Register, login
  - Search and browse records
  - Manage folders and records
  - Manage collections
  - Manage archives
- **System actions**
  - Harvest and index records from archives
  - Filter records
  - Generate recommendations

END