ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

Information Society Technologies

Open D-Lib

Vittore Casarosa ISTI-CNR

MICHAEL-DELOS meeting

Roma, 11 Aprile 2006

OpenDLib

OpenDLib is a **Digital Library Management System** that allows to create and maintain distributed DL

- It is composed by a federation of services that can be activated on one or more servers
- It supports plug and play of new services
- Each service is customizable
- It supports dynamic and automatic reconfiguration of the services

User perspectives

- System Administrator
- Librarian
- End-user

System administrator (1)

The DL System Administrator creates and maintains the DL by

- preparing the hosting nodes with the appropriate application framework,
- selecting the appropriate services forming the DL,
- deciding how to distribute them and how many replicas (if needed)
- configuring the DL,
- maintaining the instance of the DL

System administrator (2)

Global configuration:

- Document Formats (DoMDL)
- Metadata formats (+mappings between formats)
- Manifestation formats (+adapters & behaviours)
- External Manifestation formats
- Back-up, Archiving, Compression, Encryption
- Query Language (operators, attributes)
- Access Rights

Library administrator

Library administrators access the DL using customizable user interfaces.

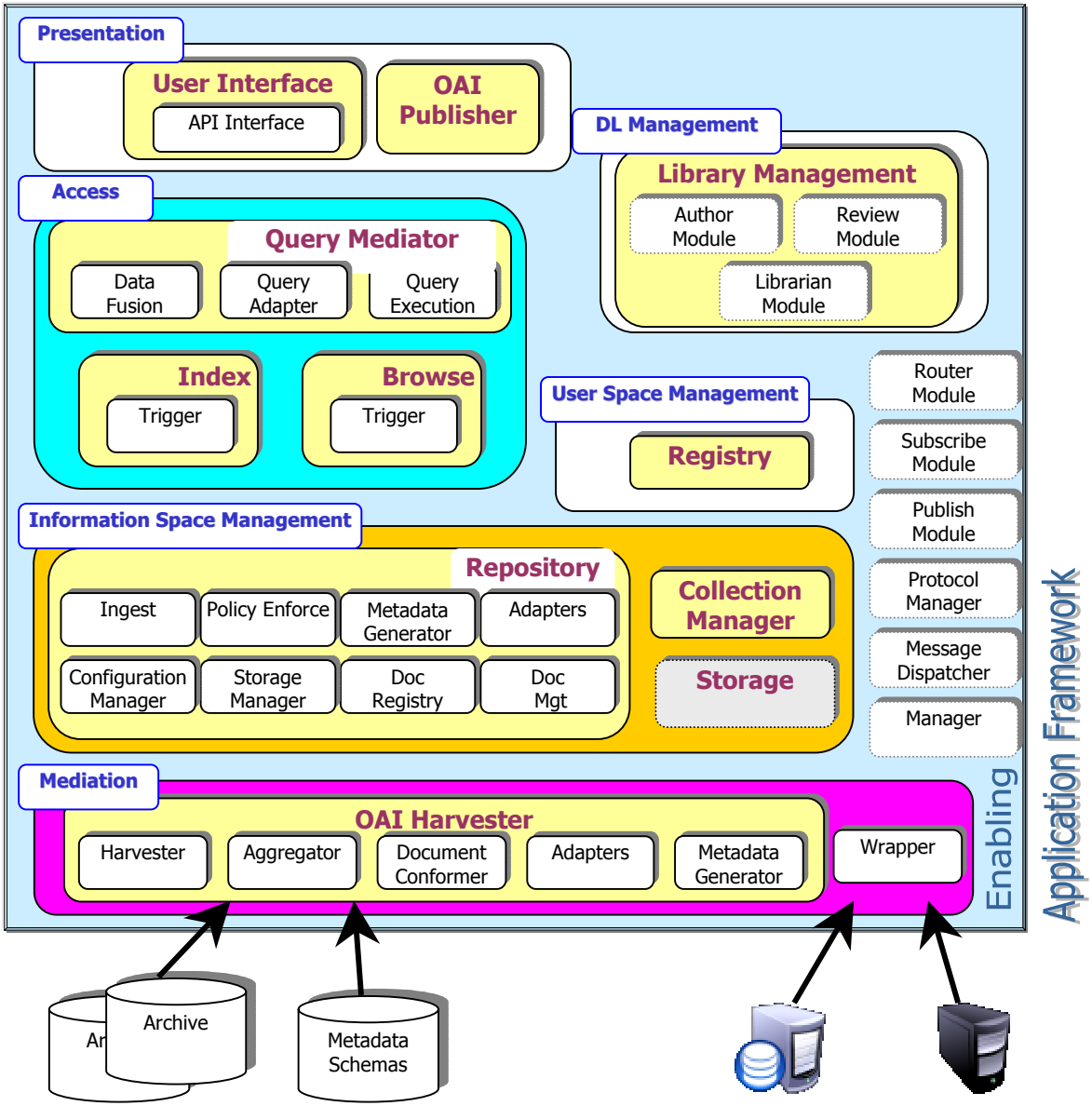
- manage the DL assigning rights
  - Identify the Repository administrators
    - Identity authors, reviewers, and editors
    - Create virtual views of the repository content
  - Assign roles to end-users
    - Group Manager
    - Collection Manager
    - Author
    - Reviewer
    - *DL specific*
  - Customize user interface
- monitor DL status and usage

End-user

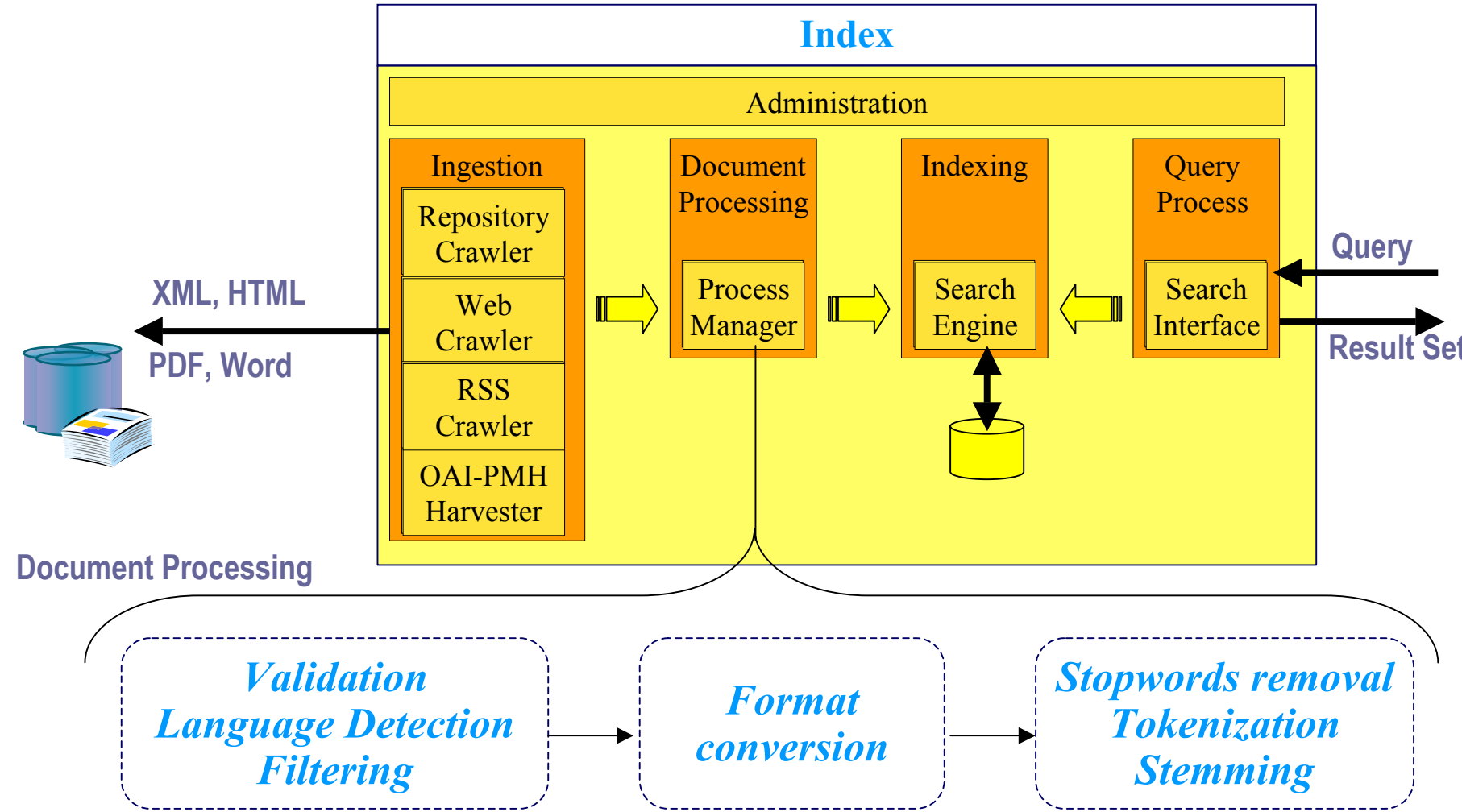
End-users access the DL using customized user interfaces.

- *Author* submits document for inclusion in the DL
- *Reviewer* reviews the submitted document (Library administrator final approval)
- *Generic user*
  - searches, browses, and accesses document;
  - customizes the information space (e.g. virtual collections);
  - customizes the working session;
  - customizes the result-set formats;
  - shares private documents within defined groups.

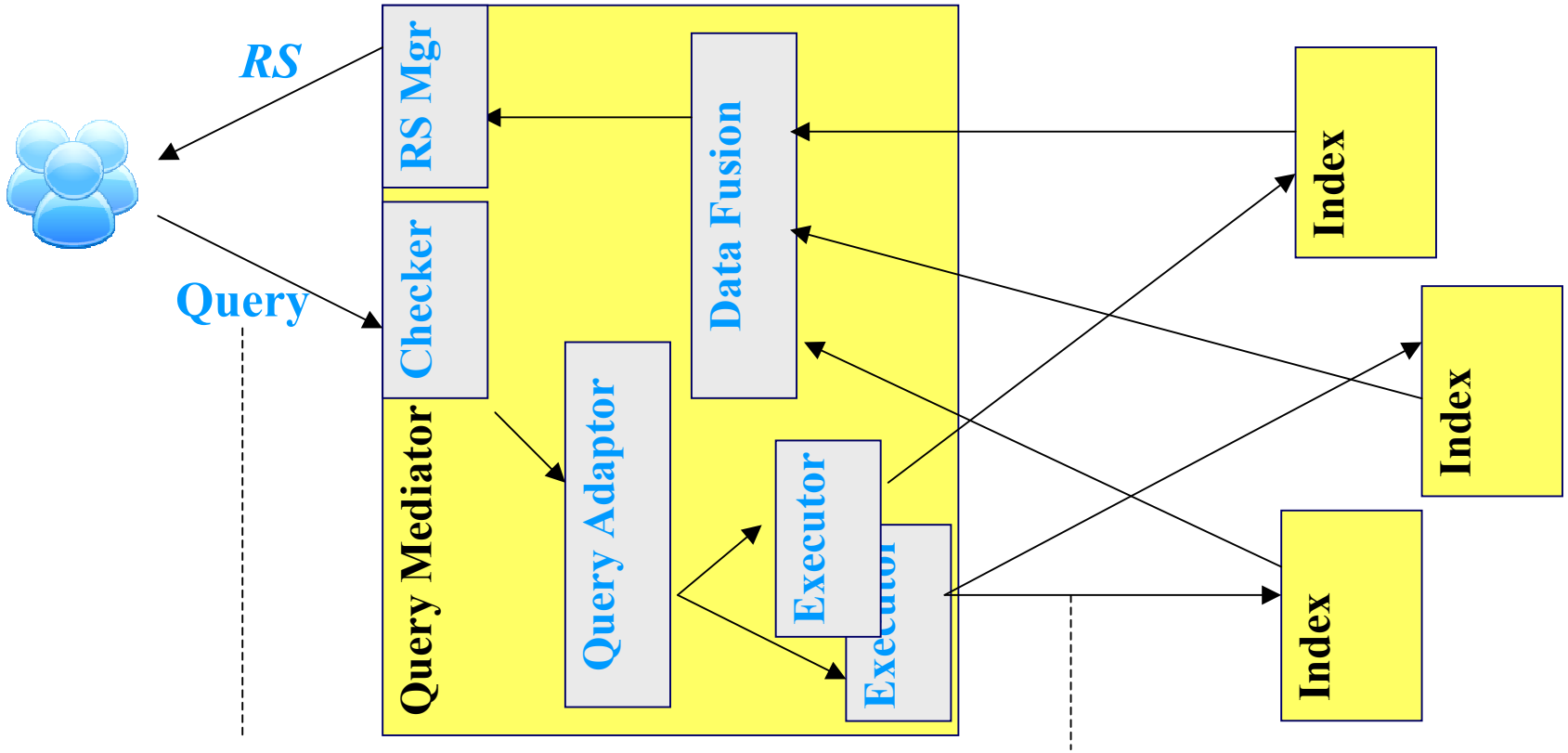
Architecture



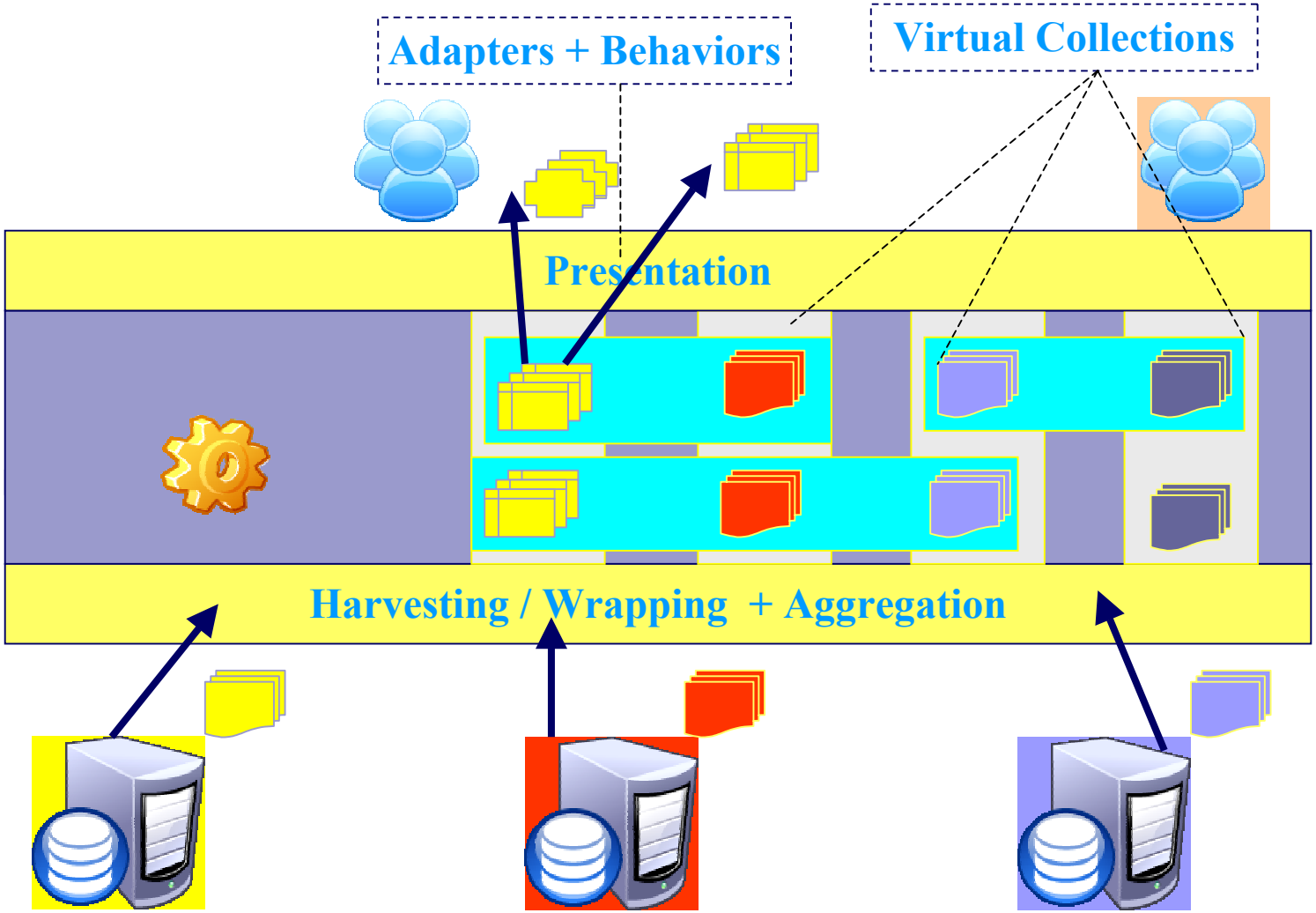
Index Management



Search Management



Content and Metadata Management (1)



Content and Metadata Management (2)

- Features
  - access to data over multiple Repositories
  - management of private working areas
  - definition of custom metadata formats
  - one document can have multiple metadata formats at the same time (all of them are indexed)
  - automatic format migration and manipulation
  - SDK to develop custom services
  - virtual collection management

Index and Search Management (1)

- User Condition
  - Simple query
    - Google like
      - *UC* := String
    - Attribute-value pair
      - *UC* := MD:attribute *Op* value
      - Op := = | contain | phrase | #n | …
  - Advanced query
    - Attribute-value pairs composition
      - UC := *Con Pair+*
      - Pair := MD:attribute *Op* value
      - Op := = | contain | phrase | #n | …
      - Con := and | or
    - Attribute-value pairs structured composition
      - UC := *Pair* | *Pair Con Pair*
      - Pair := MD:attribute *Op* value
      - Op := = | contain | phrase | #n | …
      - Con := and | boolean-and | or | and-not | filter-accept | filter-reject

“digital library”

“dc:creator = pagano”

“and (dc:creator = pagano dc:title contain DL)”

“dc:creator = pagano and (dc:title contain DL or (….boolean-and …))”

Index and Search Management (2)

- Features
  - distributed query process (preparation of the query plan, selection of the appropriate Indexes, invocations of Indexes, fusion of the various result-sets, end user presentation)
  - virtual collection materialization at query time
  - cross archive search/browse support
  - distributed indexing process multilingual indexes
  - boolean and probabilistic query operators support
  - relevance feedback support
  - result-set paging
  - session management