ENABLING SEMANTIC SEARCH IN STRUCTURED P2P NETWORKS VIA DISTRIBUTED DATABASES AND WEB SERVICES

Maria Teresa Andrade
FEUP / INESC Porto
mandrade@fe.up.pt ; maria.andrade@inescporto.pt
http://www.fe.up.pt/~mandrade/ ; http://www.inescporto.pt

Presentation walkthrough

- Introduction:
  - current situation and challenges in p2p networks
- Context of the work
  - MOSAICA
    - goals, system architecture, functionality
- The proposed approach
  - concept
  - tools used, modular functionality, layered architecture
- P2P Content Management System
  - use cases, uploading and searching content, deployed testbed
- Conclusions

Introduction

- Current situation and challenges in p2p networks

Current situation in p2p

- unstructured architectures
  - the network is flooded with query messages to locate streams and files (all nodes receive the query)
- structured architectures
  - based on Distributed Hash Tables (DHT) that maintain an index to the available resources, facilitating location
  - query messages are forwarded to only some nodes, which are more likely to have the location of the desired resource

Challenges in p2p for multimedia

- generally, users when searching for multimedia resources
  - do not know the exact name of the file they are looking for
  - often want to receive a useful set of results that are clearly related to the subject they are interested in
- while unstructured designs allow this kind of proximity behavior
  - at the cost of larger search times and increased network traffic
- common structured, DHT-based, designs do not
  - search is performed using a hash key and only exact matches are returned

Challenges in p2p - semantic queries

- Current research initiatives include
  - Locality Sensitive Hashing (LSH) techniques (nearest-neighbor search problem in a Euclidean space)
  - hierarchical multiple indexes, allowing query-to-query mapping and thus the implementation of a recursively query process
  - the CUBIT project at Cornell University using an overlay where peers are organized in concentric rings
- though promising good results, require complex management and/or processing of intermediate results

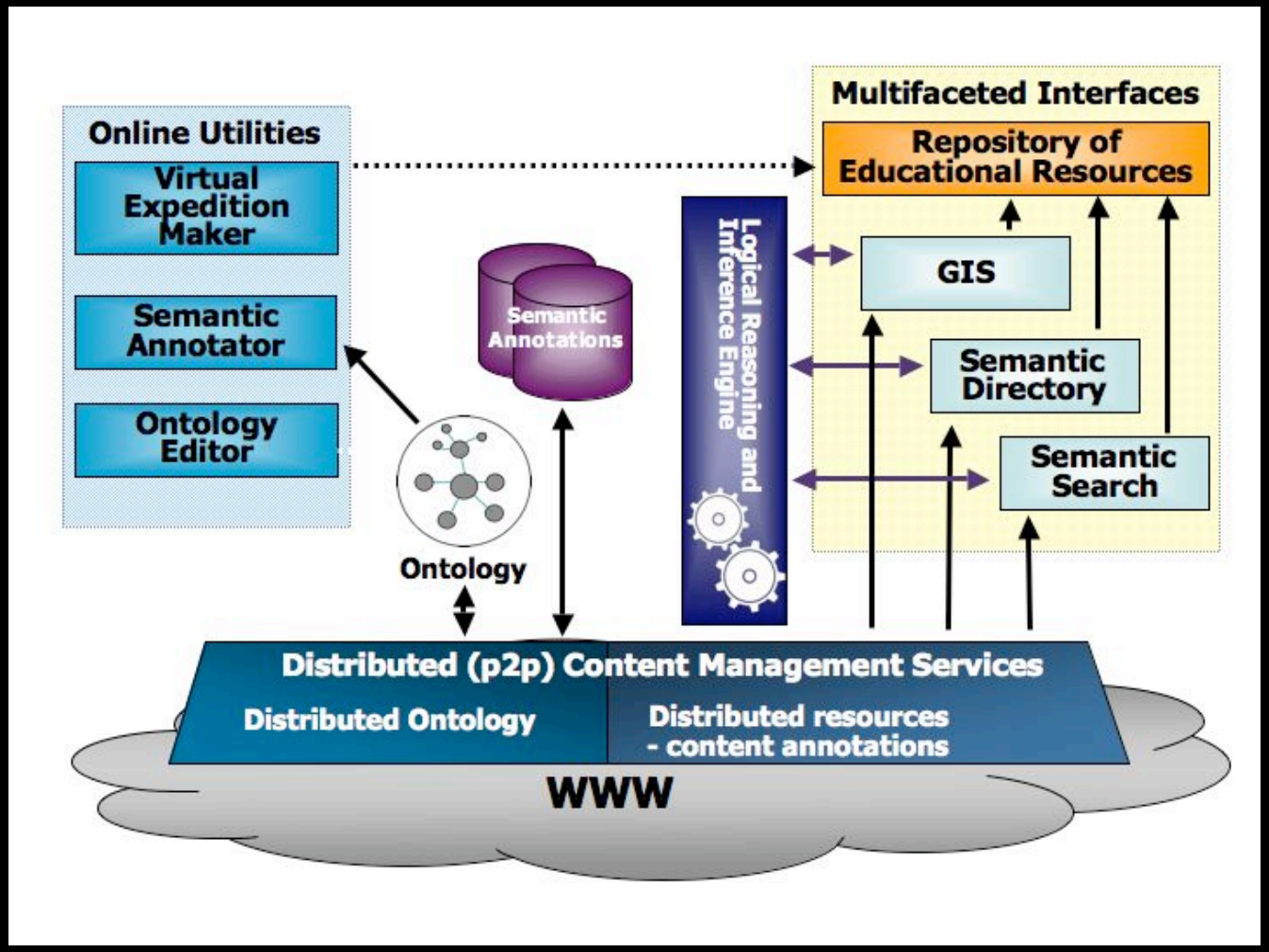
Context of the work

- MOSAICA
  - goals, system architecture, functionality

MOSAICA goals

- Leveraging Web2.0-based technologies to enhance critical thinking and open-mindedness
- tool-box of Web based technologies for the preservation and sharing of cultural heritage resources (photos, documents, video, sound, etc.)
  - sharing of cultural resources owned by private people or organizations such as museums.
  - motivate users to create innovative experiences, multimedia virtual expeditions, telling their personal stories, and sharing them with remote peers, via a rich interactive environment
- reach as many people as possible
  - low cost, high availability, simplicity

MOSAICA architecture

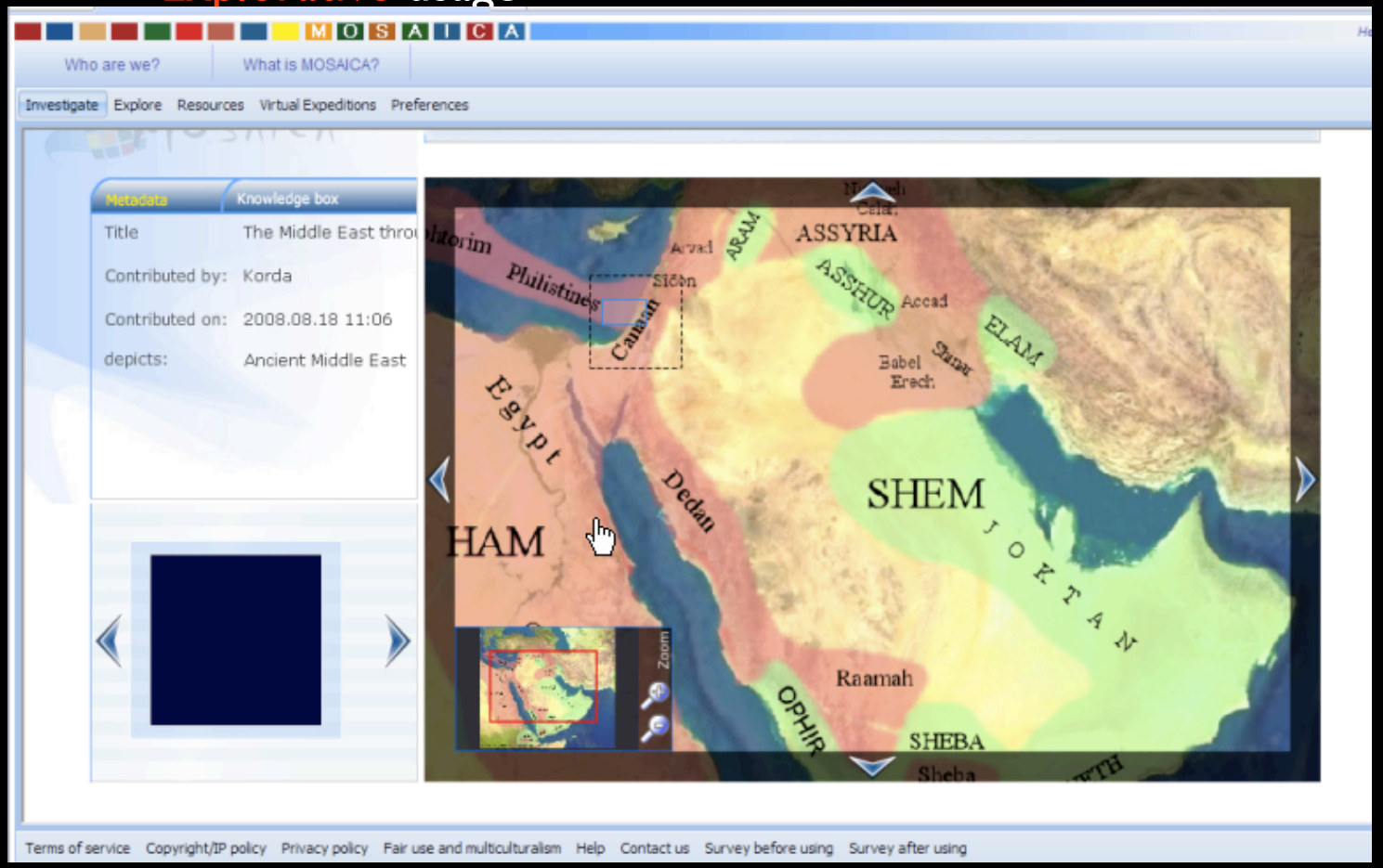


MOSAICA functionality

- Explorative usage
  - visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  - or by exploring MOSAICA semantic directory, or by submitting a query.
- Collaborative usage
  - annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  - submit and exposed to the public annotated resources
  - design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository
- Guided usage
  - select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository

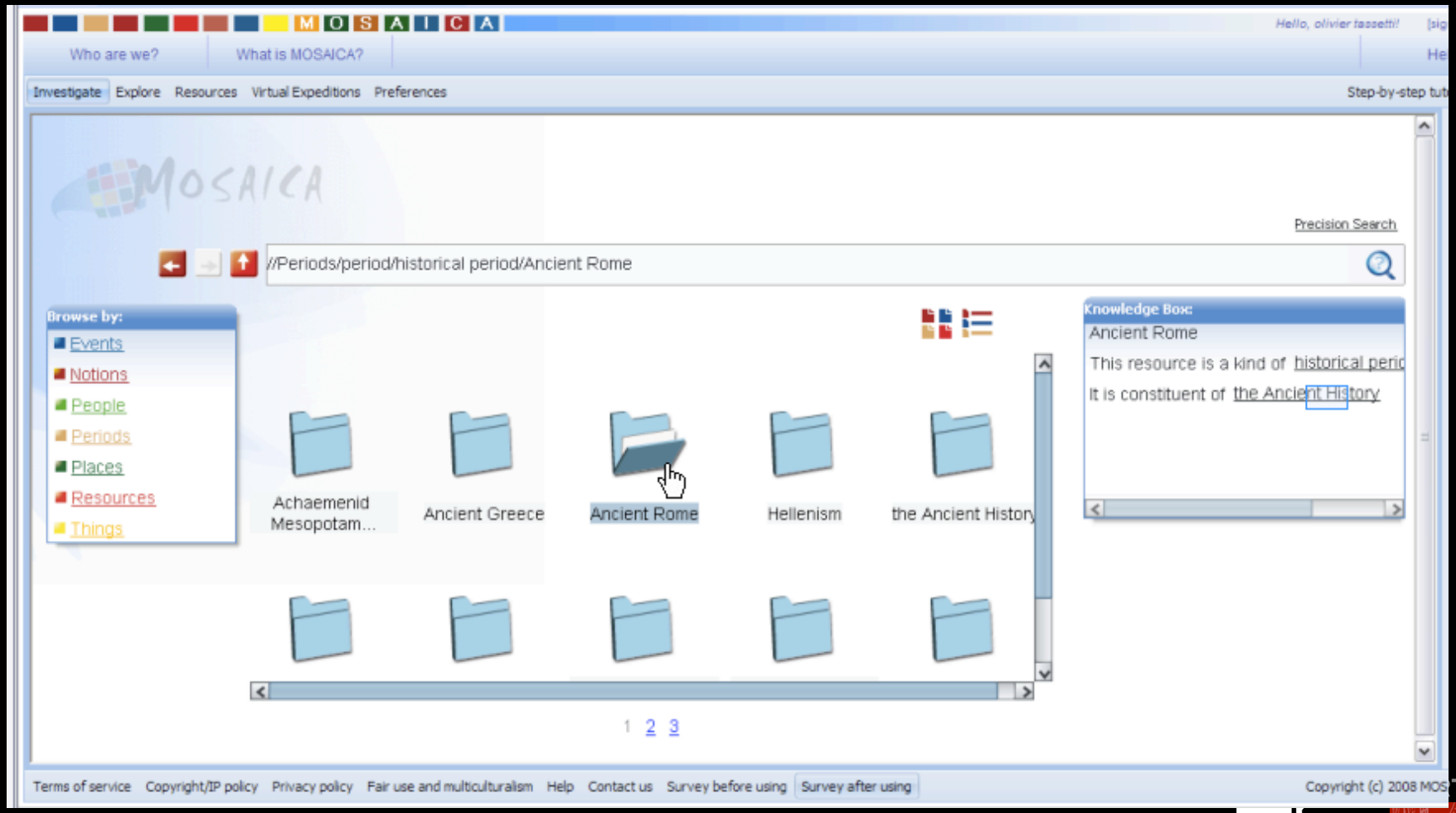
MOSAICA functionality

- Explorative usage



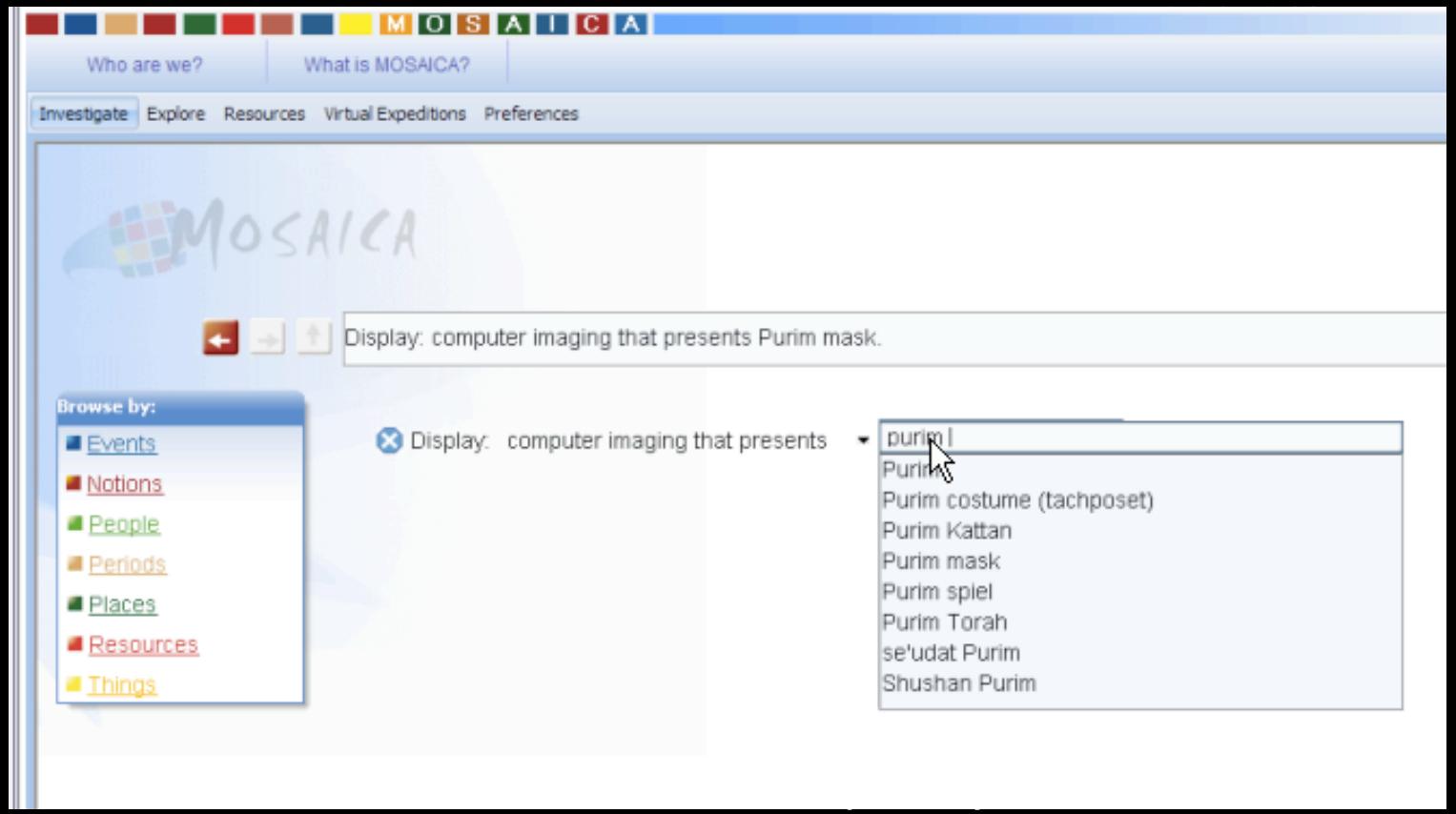
MOSAICA functionality

- Explorative usage

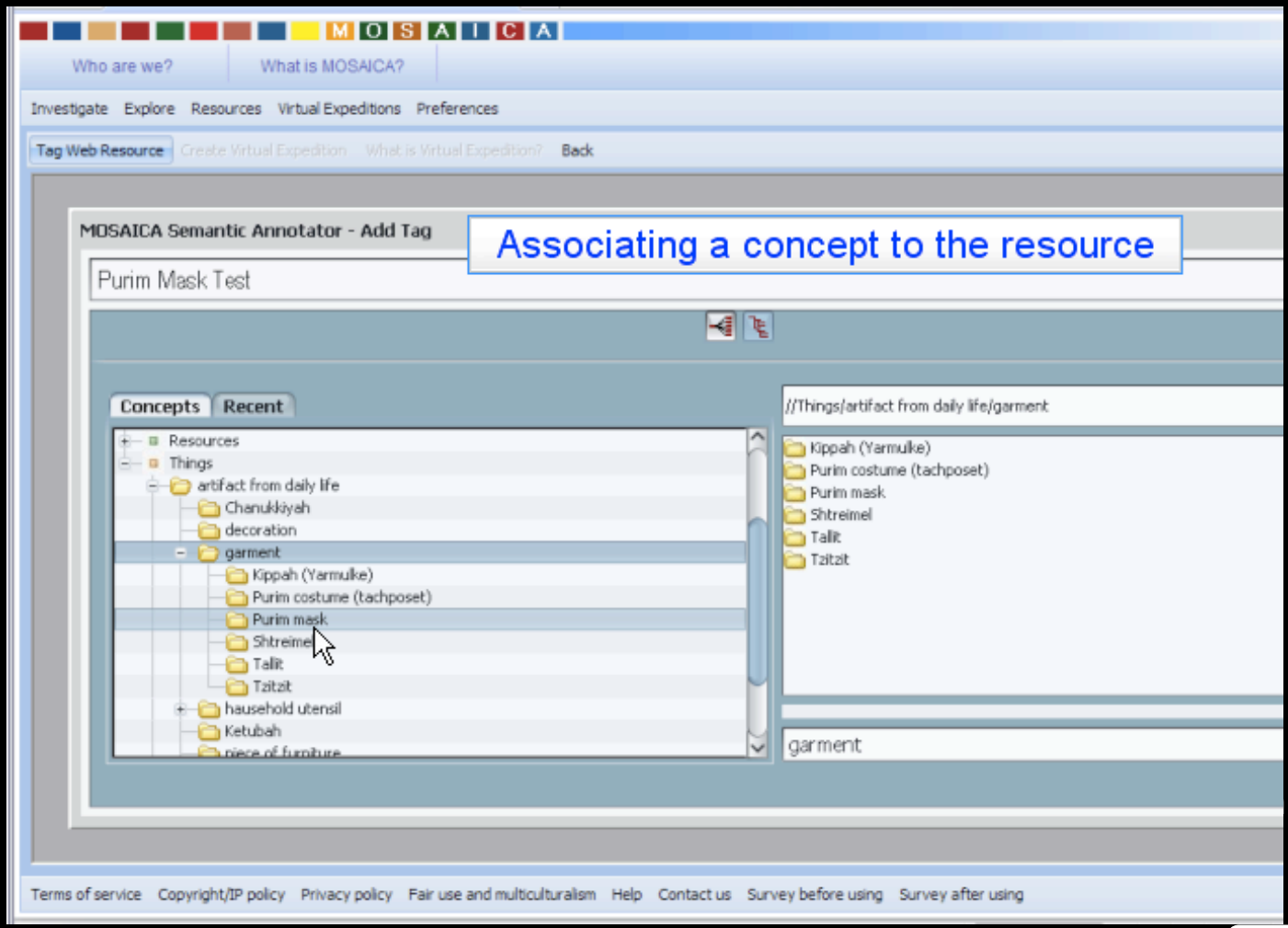


MOSAICA functionality

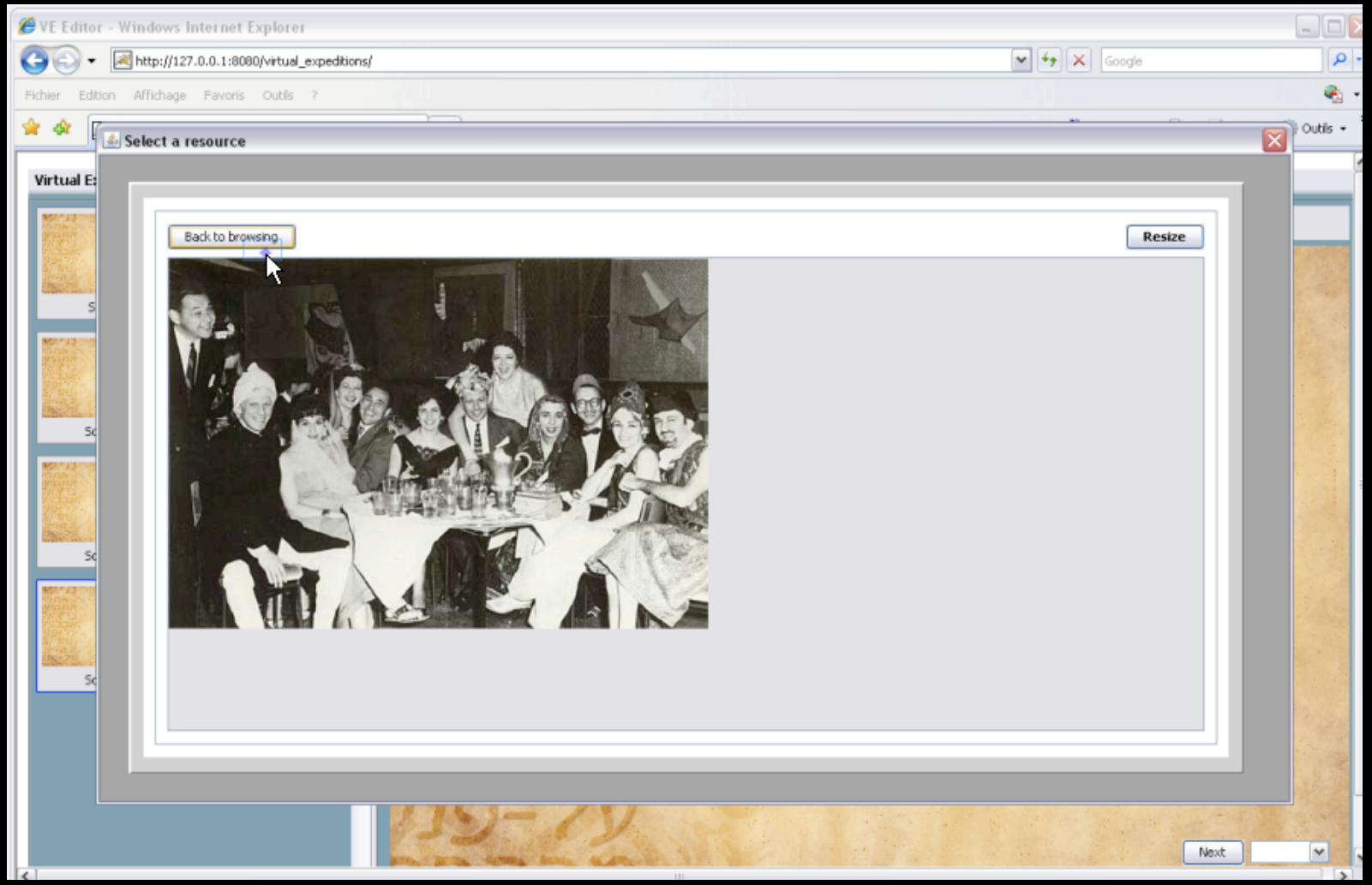
- Explorative usage



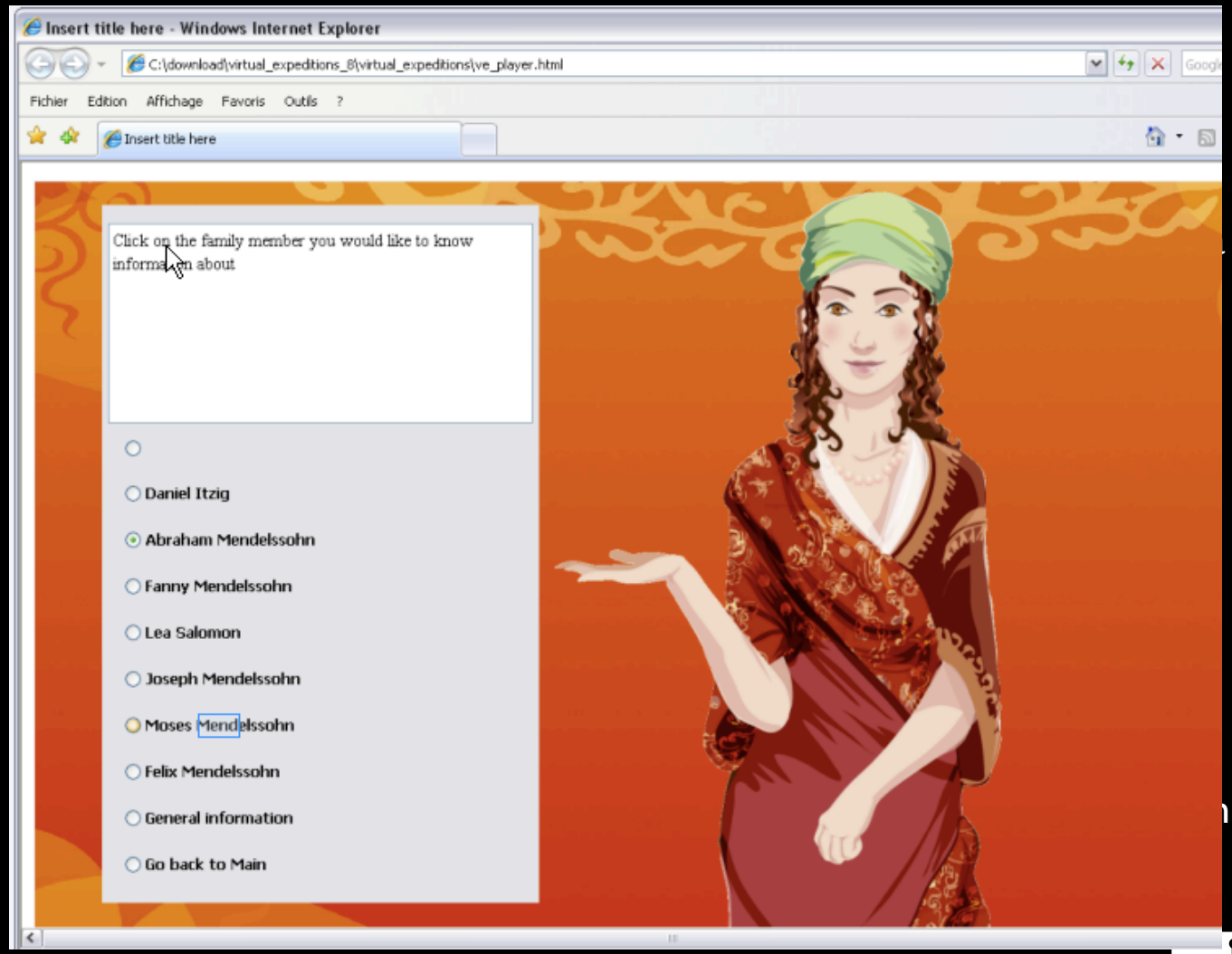
MOSAICA functionality



MOSAICA functionality



MOSAICA functionality



The proposed approach for the distribution and management of resources

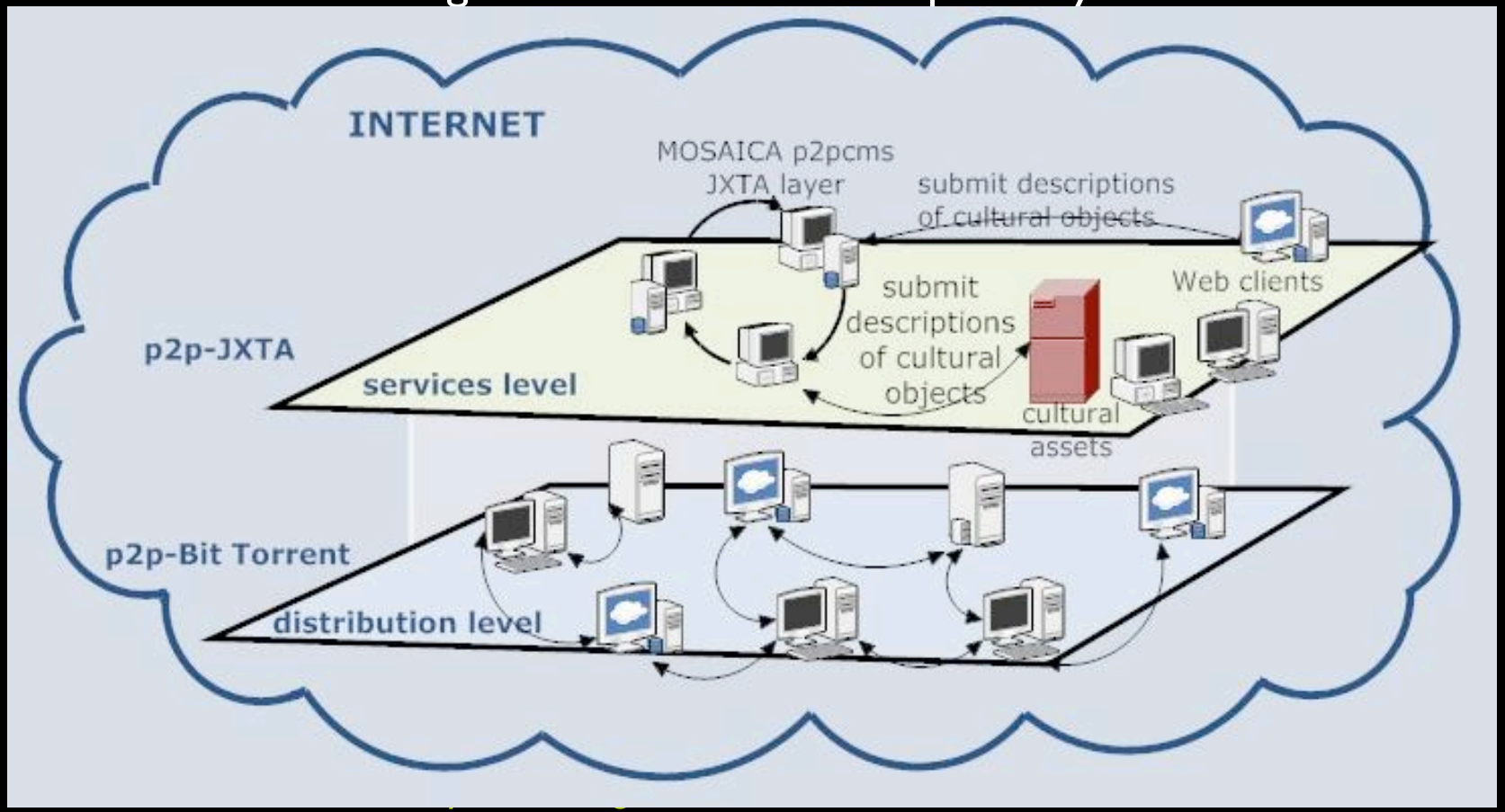
- concept, tools used
- modular functionality, layered architecture

Concept

- The challenge for the MOSAICA repository and content management system
  - to allow efficient semantic search
  - low-cost, usability, high-availability, simplicity
- The solution
  - p2p design adopting a two layer approach
    - upper layer handling semantic and keyword based queries
      - exposing distributed services through a Web Services interface
        - notably the Ontology service
      - built using the JXTA framework and distributed databases
    - lower layer handling media resources distribution and download
      - based on DHT p2p architecture, notably using Bit Torrent protocols

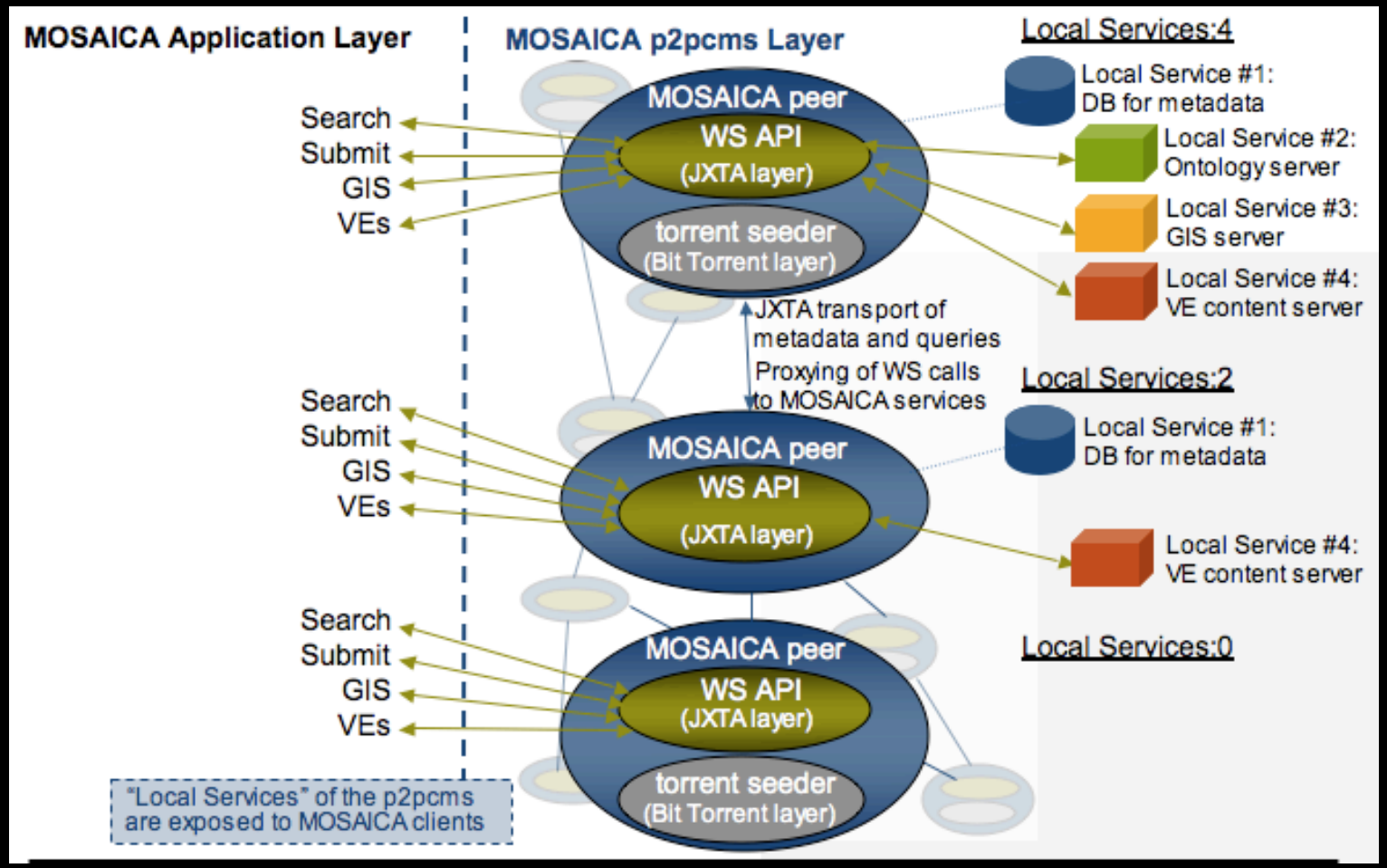
Concept

- The challenge for the MOSAICA repository and content



- based on DHT p2p architecture, notably using Bit Torrent protocols

p2p-cms modular functionality

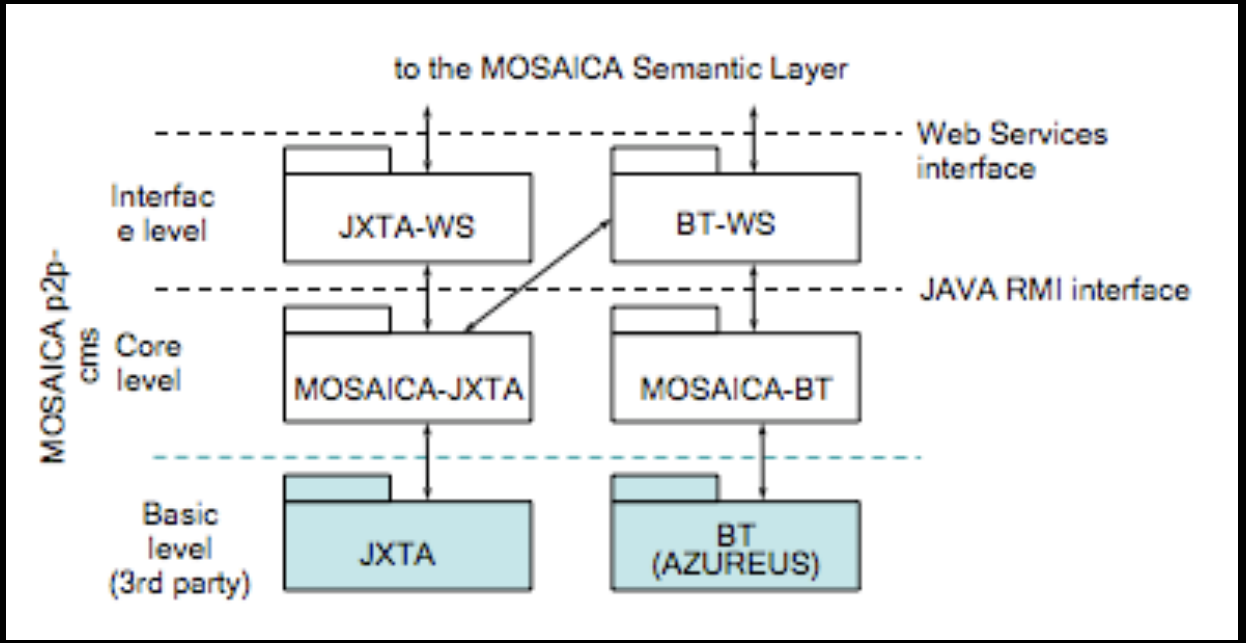


p2p-cms functionality

Method	Description
login(LUser, password)	Login method. It returns a session hash if user credentials are correct.
registerContent(LSession, contentDat)	Content data XML is sent. ContentID is returned
requestContent(LSession,contentID, boolean wait)	XML describing content (title, torrent binary, etc.) is returned
searchContent(LSession,XMLquery, boolean wait)	Query contents using a XML query. XML list of found contents, for this user groups, is returned
removeContent(LSession,content ID)	Remove index to content
addMetadata(Lsession, content ID, indextype, property, value)	Submits new semantic descriptions to be propagated to the distributed databases
removeMetadata (Lsession, content ID, indextype, property, value)	Eliminates a metadata record
changePassword (Lsession, user,oldpassword, newpassword)	Updates password for user

p2p-cms architecture

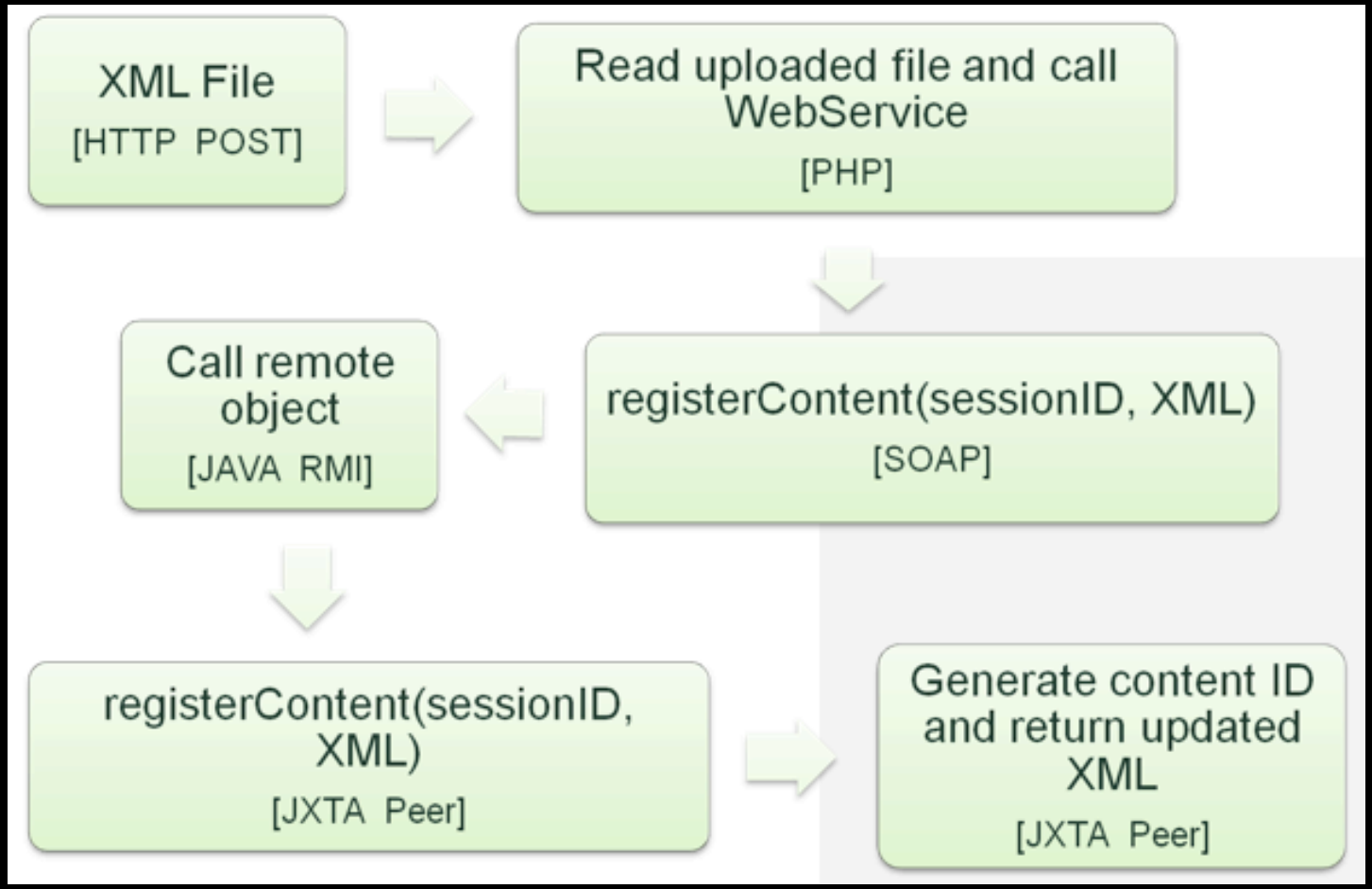
- modular and layered architecture



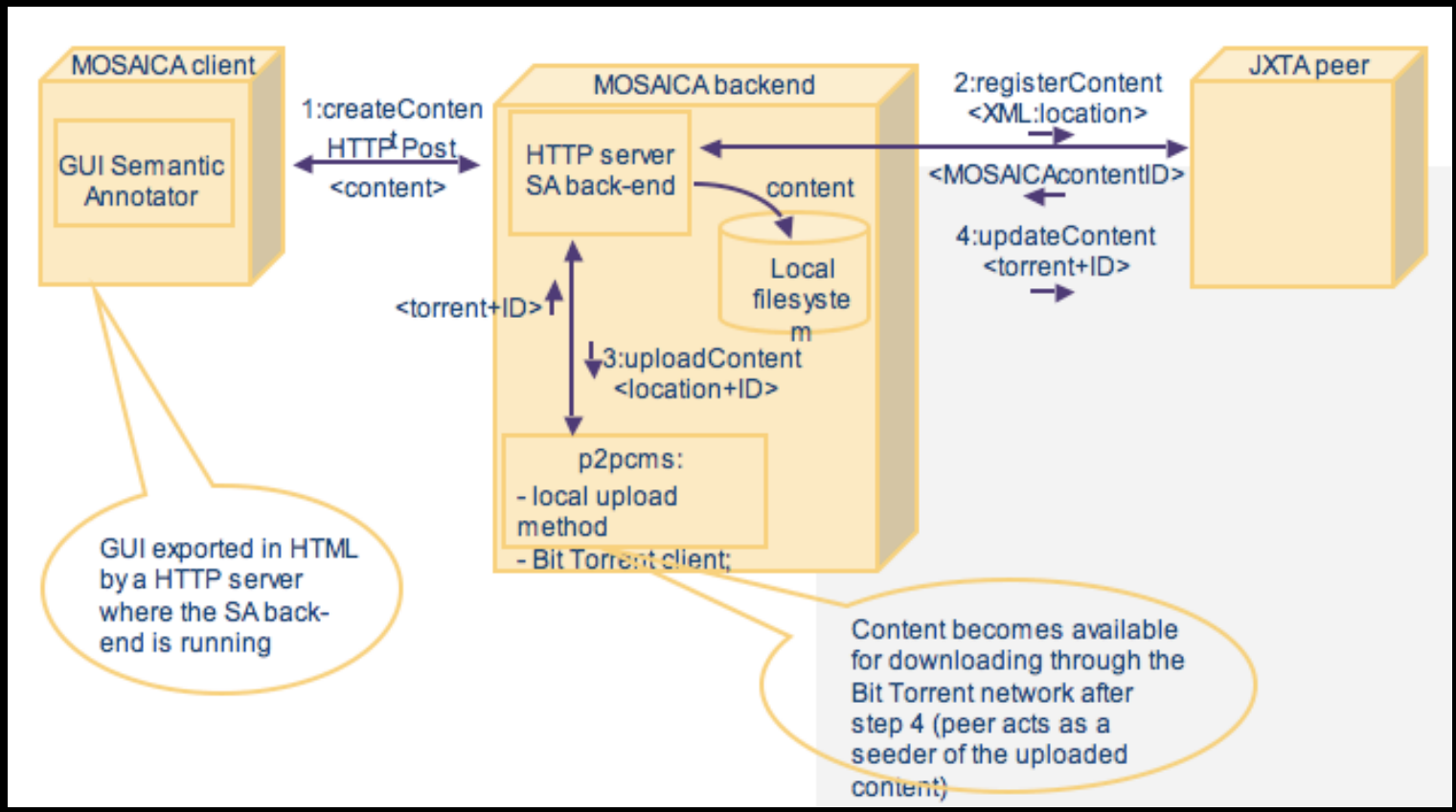
P2P Content Management System

- use cases supported
  - uploading and searching content
- testbed deployed

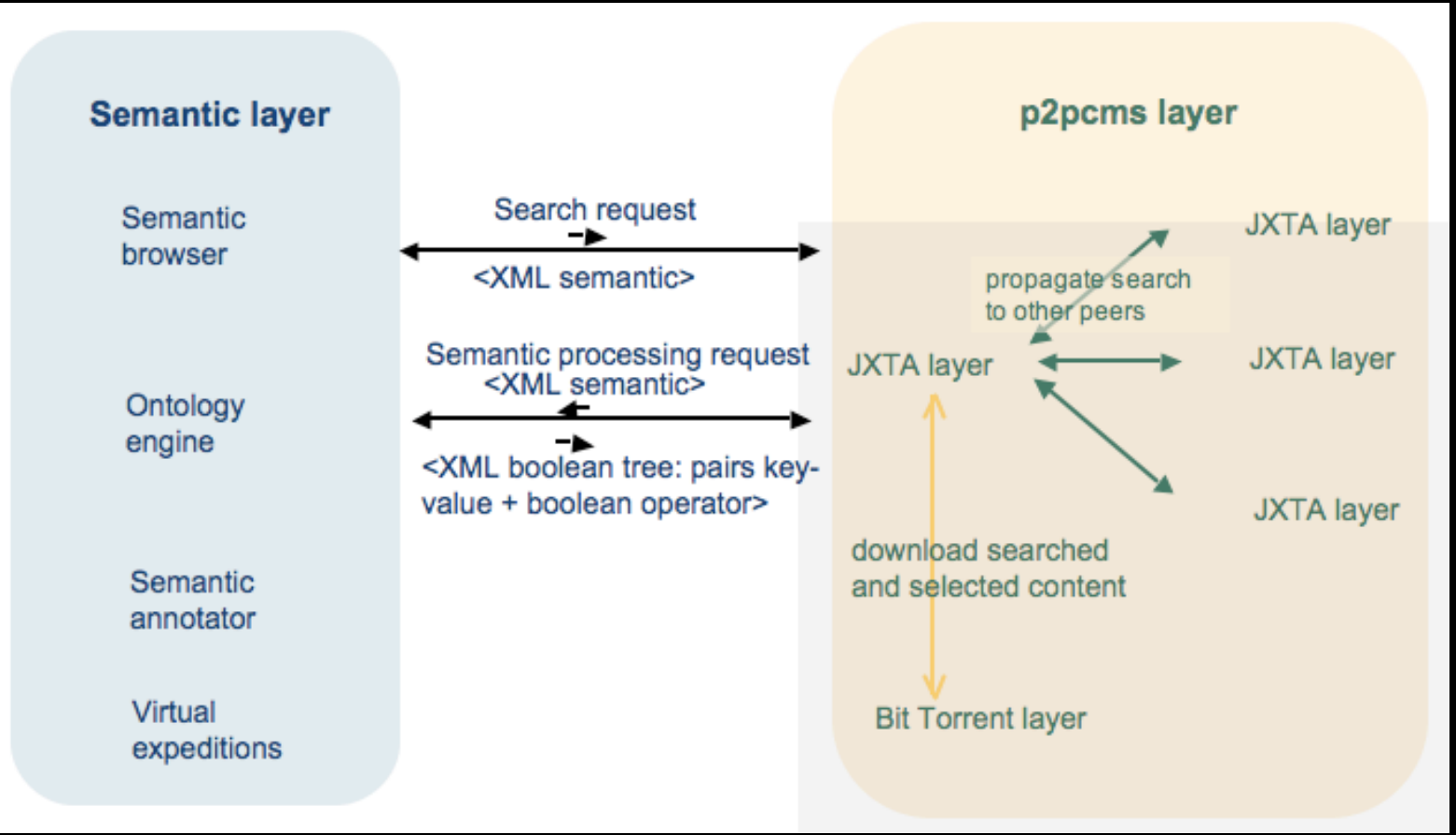
Some important use cases



Some important use cases



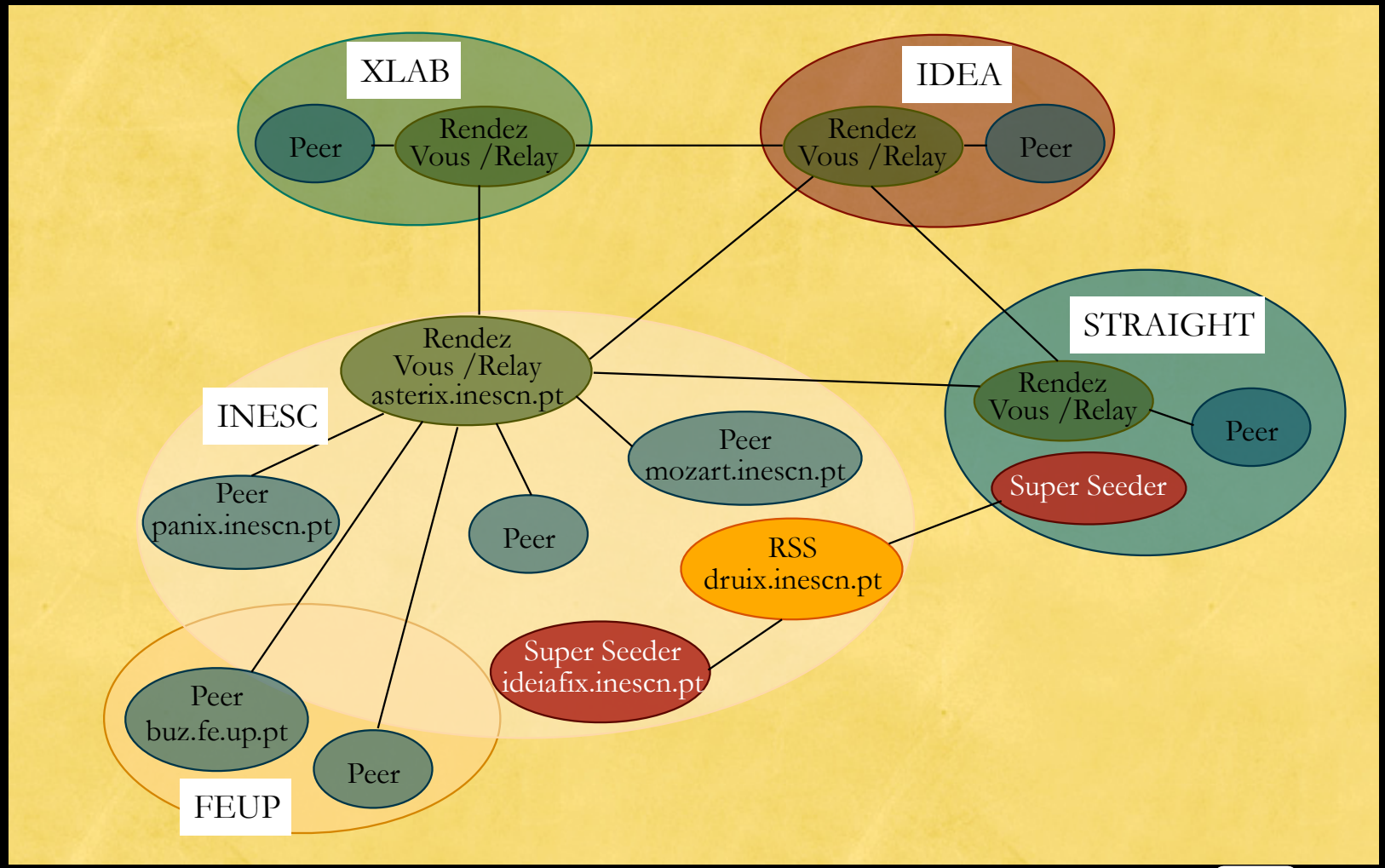
Some important use cases



testbed deployed

- Needed for operational testing, integration and data gathering for simulation and evaluation
- Includes JXTA and Bittorrent overlays
- A Rendezvous/Relay peer, normal peer, super seeder and RSS server can be installed in only one machine (independent applications)
- Needed software : P2P-cms application, Azureus, JXTA Shell, Web server (for RSS feeds), Mysql and AXIS 2 (with RSS and peer webservices)

testbed deployed



Conclusions

- The two-layered approach
  - by which metadata is manipulated in one layer and content in the other layer
- enables to obtain an optimal solution
  - on one side, it enables to take advantage of the great efficiency of DHT (Distributed Hash Tables) for the distribution of and access to media resources;
  - on the other hand, through the implementation of distributed databases for managing metadata, it enables to overcome the limitation of DHT-based P2P networks for performing semantic-based searches
  - additionally it enables the use of simple search mechanisms
- however, semantic search is only possible through the use of an ontology server
  - XML based semantic queries are decomposed in a series of boolean expressions which are then used for indexing and searching
- Still, performance and scalability tests need to be performed

Thank you very much for your attention!