XML-Based Content Management Framework for Digital Museum

Joshua J.S. HONG,
National Chi Nan University
TAIWAN
jshong@csie.ncnu.edu.tw

Samuel CRUZ-LARA,
LORIA / INRIA Lorraine
Language & Dialog Project
FRANCE
Samuel.Cruz-Lara@loria.fr

國立暨南國際大學
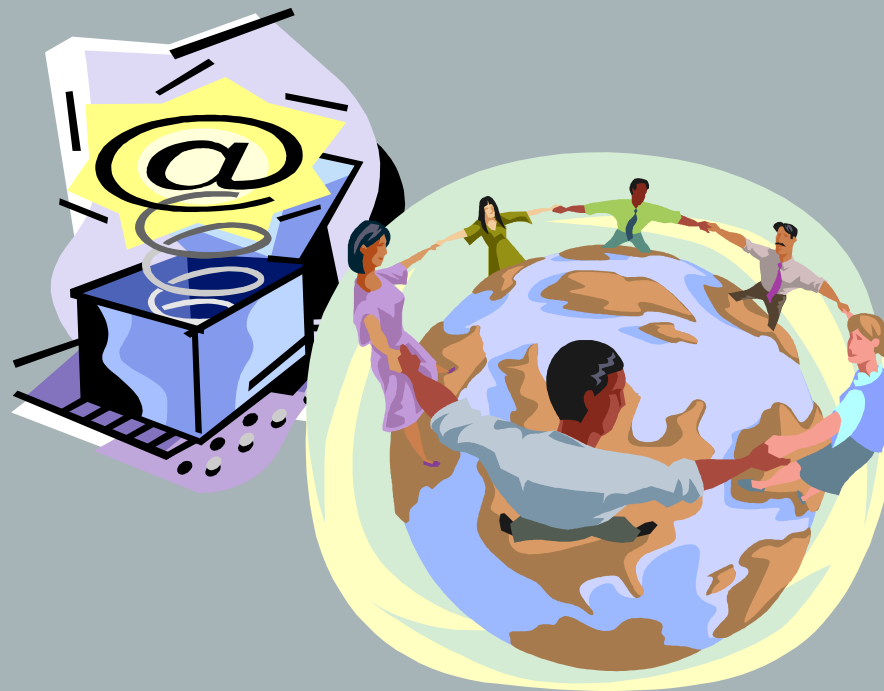National Chi Nan University

INRIA Lorraine

loria

Contents

Preface

- *This is not (yet !) a presentation from a concrete implementation of a project*
- *Rather, this is a presentation outlining the main objectives of a collaboration (NCNU - LORIA INRIA Lorraine) that is only at the very beginning*
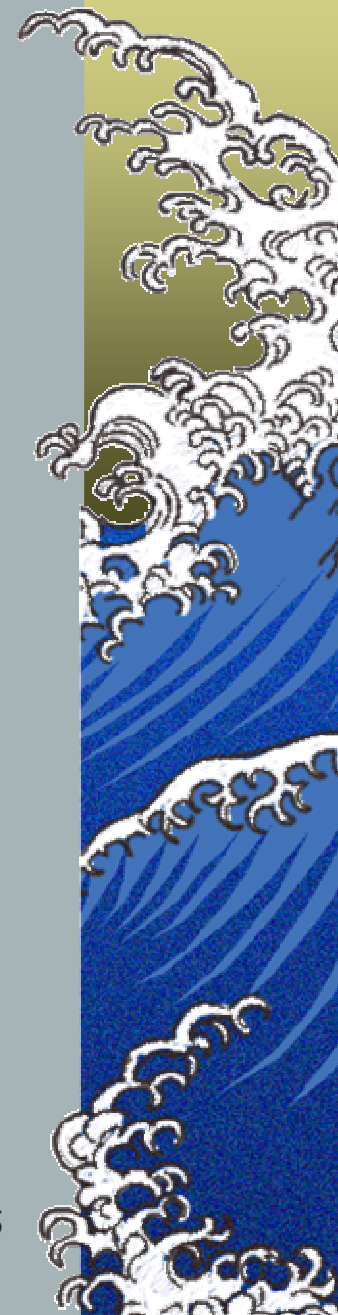- *National Science Council (Taiwan) & INRIA (France) Collaboration*

The "Digital Museum Project"

Introduction

- *A Digital Library (DL) focuses on conserving, cataloguing, accessing, and tracking the usage of digitized material*

Introduction

⋏ *On the other hand, a Digital Museum (DM), other than being a simple digital archive, indeed emphasizes more on providing users with highly educational and motivating exhibitions*

Introduction

- *To efficiently transform the organized media objects deposited in the digital archive into educational experiences, there is a need of a novel Content Management Framework (CMF)*
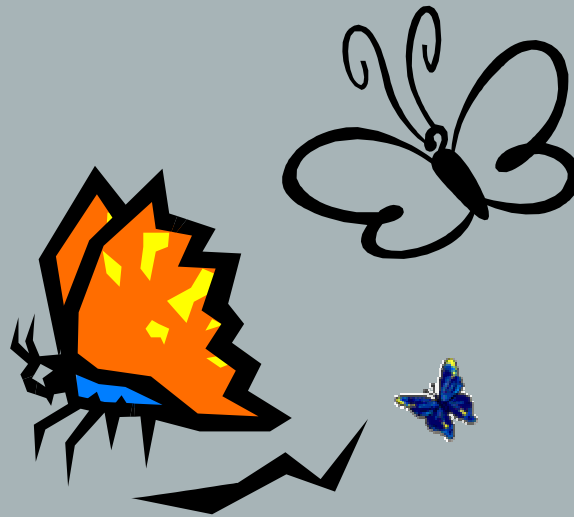
Introduction

⋏ *This novel CFM shall be used for organizing digital collections and for quickly selecting, integrating, and composing objects from the collection to produce exhibitions of different presentation styles*
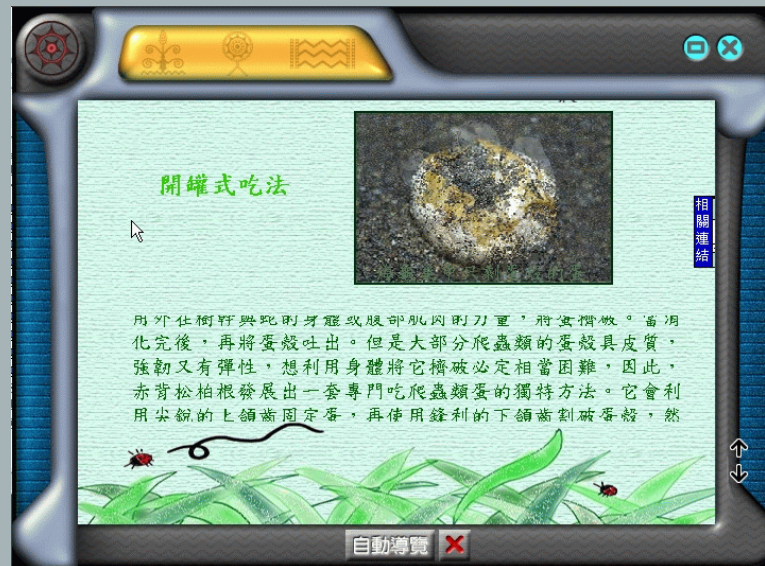
National Chi-Nan University

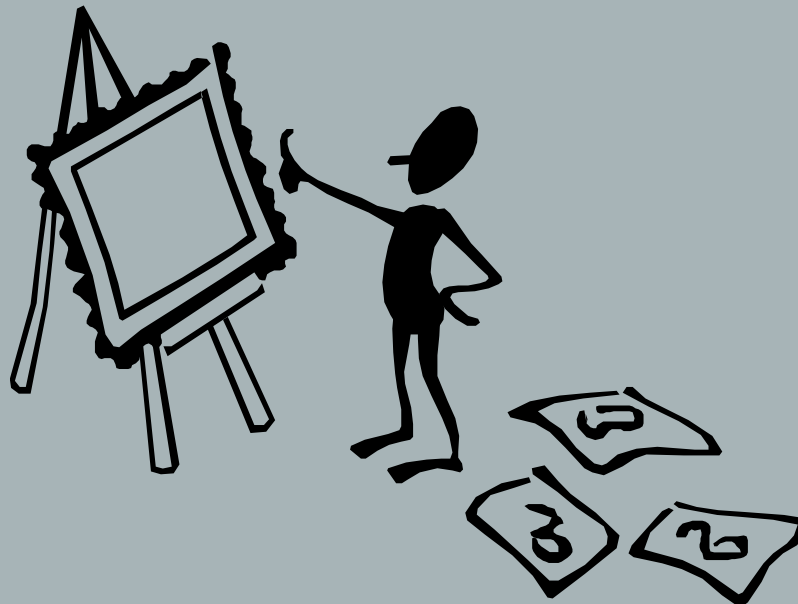- *A Digital Museum of Taiwanese Butterflies*

National Chi-Nan University

⋏ *The Lanyu Digital Museum on-line Exhibition*

National Chi-Nan University

- *XML-based hypermedia digital museum content management framework*

LORIA - INRIA Lorraine

⋏ *The SILFIDE Network : An Interactive Service for Using, Studying, Distributing and Sharing Natural Language Resources"*

Serveur
Interactif, pour la
Langue
Française, son
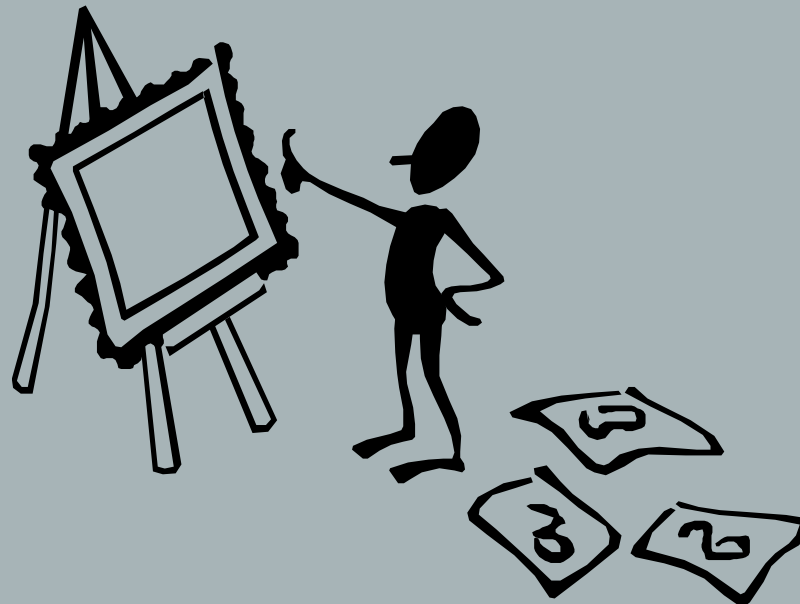**I**dentité, sa
Diffusion et son
Etude

LORIA - INRIA Lorraine

⚐ *ELAN : "European Language Activity Network" (Multi-Lingual Information Society)*

LORIA - INRIA Lorraine

➤ *XML-based documents management and networked environments*

MLIS-ELAN

- *MULTI-LINGUAL INFORMATION SOCIETY*
  - *European Language Activity Network*
  - *http://solaris3.ids-mannheim.de/elan/*
  - *http://www.loria.fr/projets/MLIS/ELAN/*

MLIS-ELAN

- *Companies, research teams and individuals involved in language engineering or in undertakings such as translating, dictionary making or philology require LARGE corpora, lexicons and similar electronic resources*

- *On the other hand such linguistic databases have already been created for most European languages*

MLIS-ELAN

⌖ *It can be observed that the need exists, the product exists and yet the two have until now rarely been brought together !*

⌖ *This paradox can largely be explained by the fact that the language resources in question often cannot be accessed EASILY and usually cannot be exploited using STANDARDIZED procedures*

MLIS-ELAN

- *MLIS-ELAN is a distributed language resources system offering access to existing resources to their potential users throughout Europe*

18

MLIS-ELAN

- *MLIS-ELAN should provide standardized resources for the following languages : Albanian, Belo Russian, Bulgarian, Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian, Lithuanian, Polish, Portuguese, Romanian, Russian, Serbian, Slovakian, Slovene, Spanish, Swedish and Uzbek*
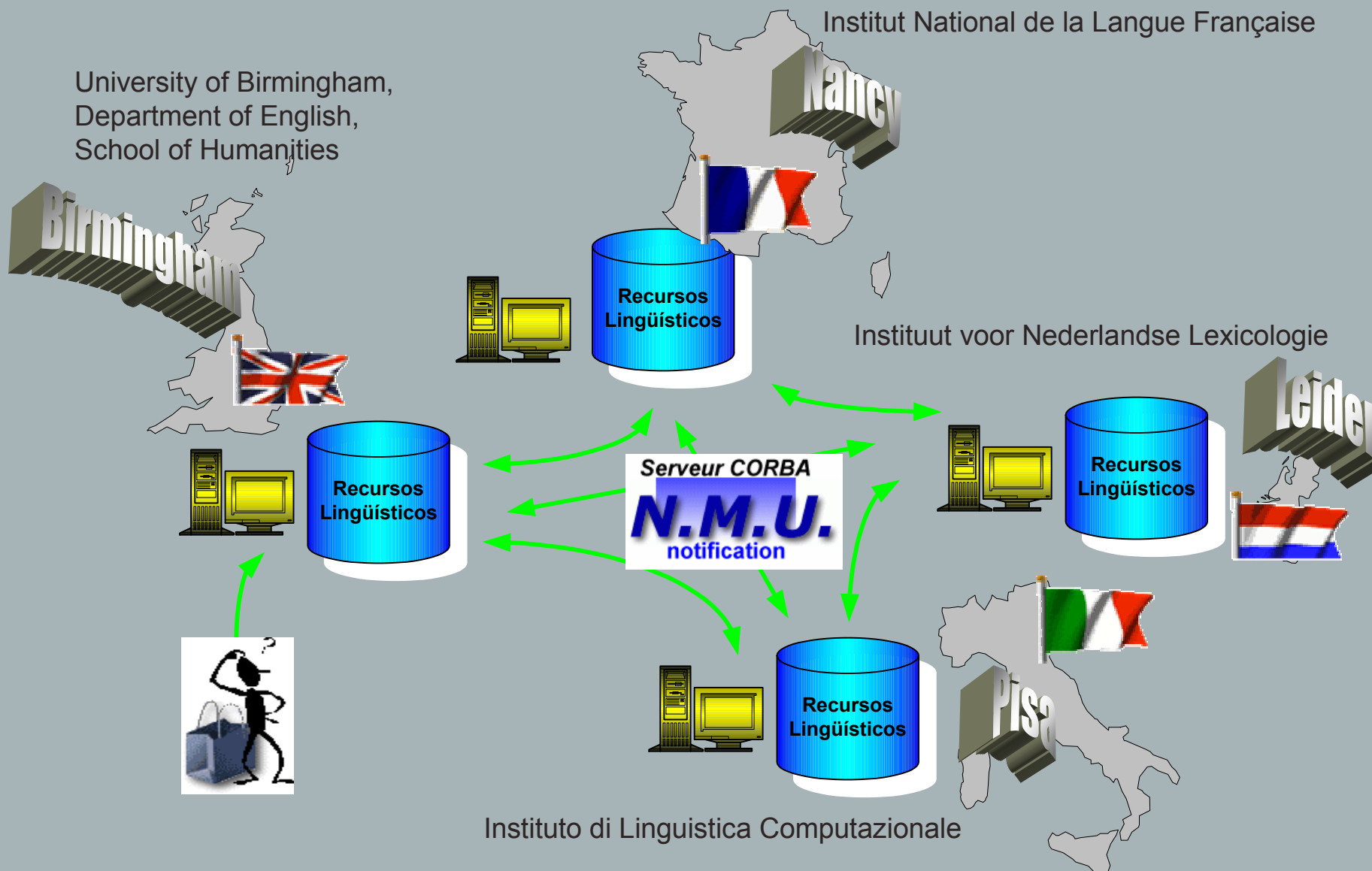
19

Main Goals

- *To build a distributed architecture for accessing and sharing natural language resources that are stocked in heterogeneous Data Bases*

- *To define a "Common Query Language" used by all servers in the network*
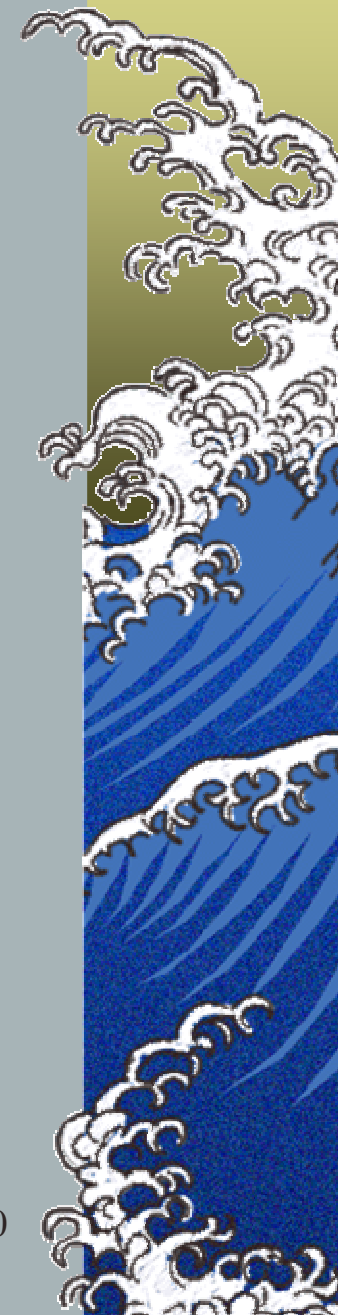
The First ELAN Network

100% XML

- *XML is obviously used for normalizing the representation of natural language resources*
- *XML is also used for normalizing the protocols we use to exchange information between all servers in the network*

ELAN Network Architecture
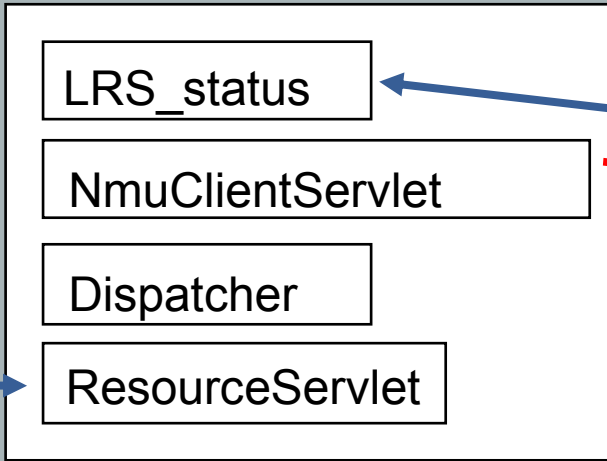
- *Three main actors :*
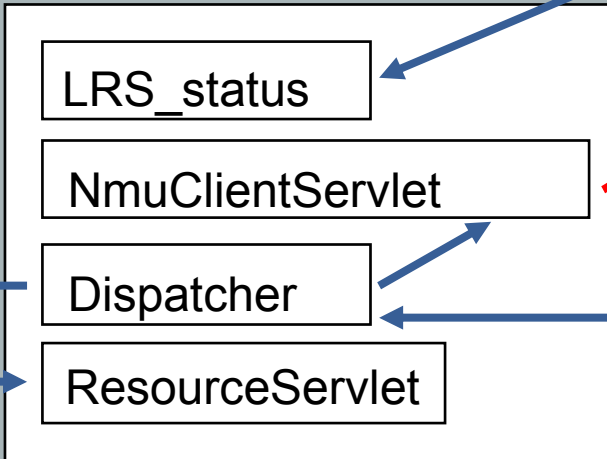 - *Network Management Unit*
  - *LR Servers*
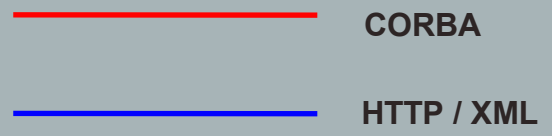 - *Java enabled Web browsers, client side*

23

Server 1

LRS_status

NmuClientServlet

Dispatcher

ResourceServlet

Server 2

LRS_status

NmuClientServlet

Dispatcher

ResourceServlet

Web Browser

Client Applet

NMU

Administration

CORBA

HTTP / XML
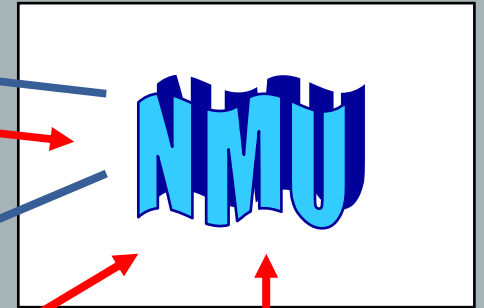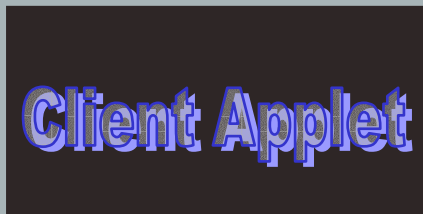
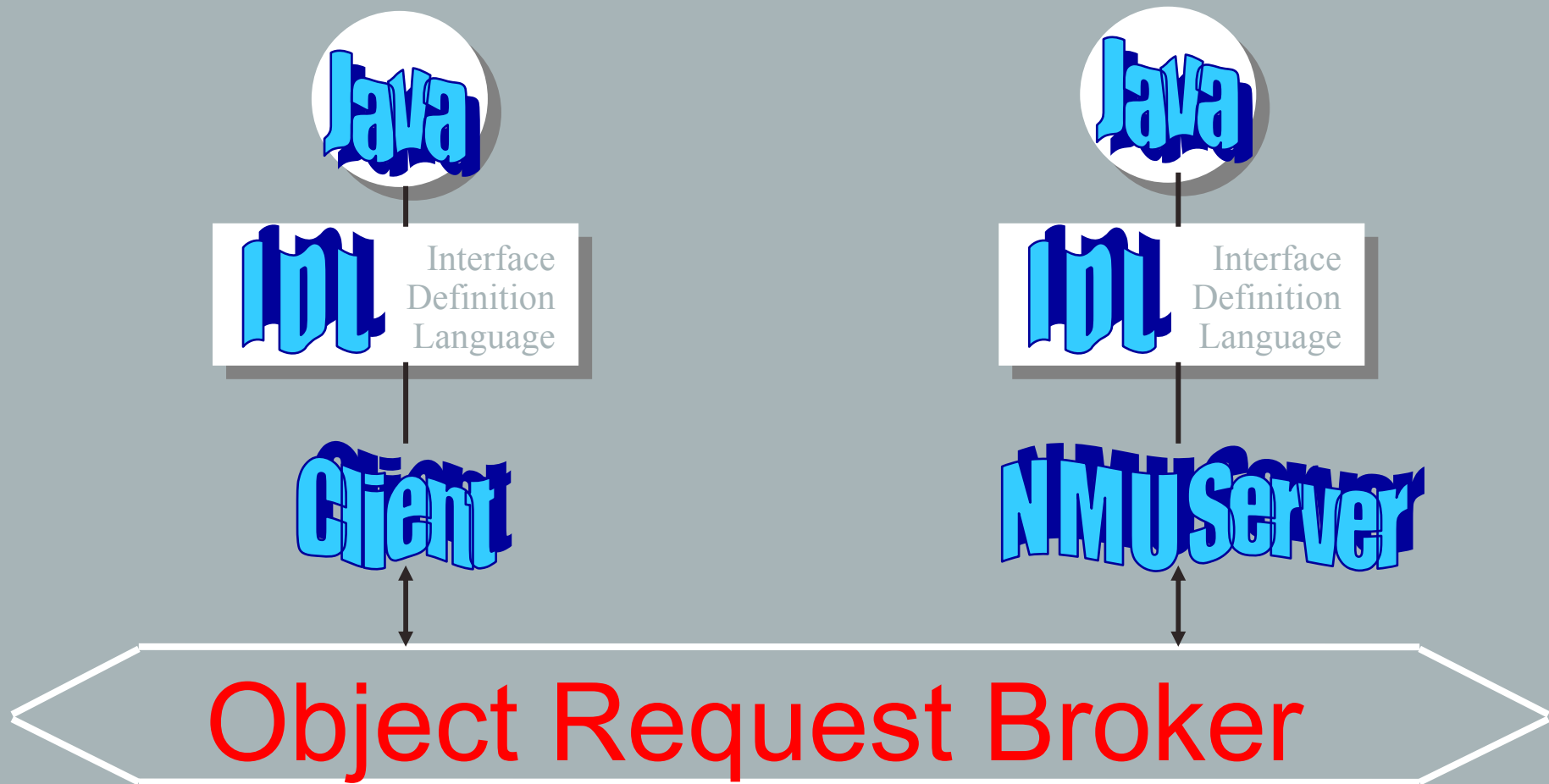# Network Management Unit

- *Keep information about all LR servers*
- *For each server :*
  - *its name*
  - *its address (Internet URL or IP address)*
  - *a profile of server contents*
- *Every LR server consult the NMU to :*
  - *get the list of registered servers on the network*
  - *get information about a specific server*
- *Contact regularly every server to check their status (offline or online)*

NMU Architecture

Based on a CORBA Clients-Server architecture

Java

IDL Interface Definition Language

Client

Java

IDL Interface Definition Language

NMUServer

Object Request Broker

NMU Administration

- *A single person or instance is responsible for adding, removing and updating servers on the network*

- *2 tools are provided :*
  - *Command line administration*
    - *fast to use*
    - *can be used from a distant computer (via telnet protocol)*
  - *Administration with a graphical user interface*
    - *more user-friendly*

Servlet Concept

- *It's a Java application, running on the server, and waiting for HTTP queries*
- *It works like a CGI (Common Gateway Interface) with the following advantages :*
  - *Hardware and OS independent*
  - *Take advantage of the power of the Java API*

28

LR server's architecture

- *A LR server is basically based on a java http server that implements the Servlets technology*
- *The purpose of these servlets include :*
  - *Implementing communication with the NMU*
  - *Dispatching of the queries to each selected server*
  - *Parsing of the queries and interaction with the linguistic resources database*
  - *Loading and saving of workspaces*
  - *General administration and users database management*

Communication with the NMU

- *This is implemented through a specific servlet which is the CORBA client from the NMU point of view*

- *This servlet purpose is to :*

  - *Contact the NMU to answer queries related to the other servers (list, name, address, status,...)*

 - *Maintain a local servers database in the case the NMU server is unavailable*

Communication with the NMU

⌑ *Sample queries sent to the NMU Client Servlet :*

⌑ *Getting the list of servers :*

⌑ *Http://myServer.a.com/NmuClientServlet*

⌑ *Getting the status of the myServer server :*

⌑ *Http://myServer.a.com/NmuClientServlet?server status=server_name*

Dispatching of linguistic queries

- *A servlet get the query from the client*
- *It's purpose is :*
  - *To dispatch the query to each selected server*
  - *To get the answer from each server*
  - *To format the answers in a MIME document which is sent back to the client*
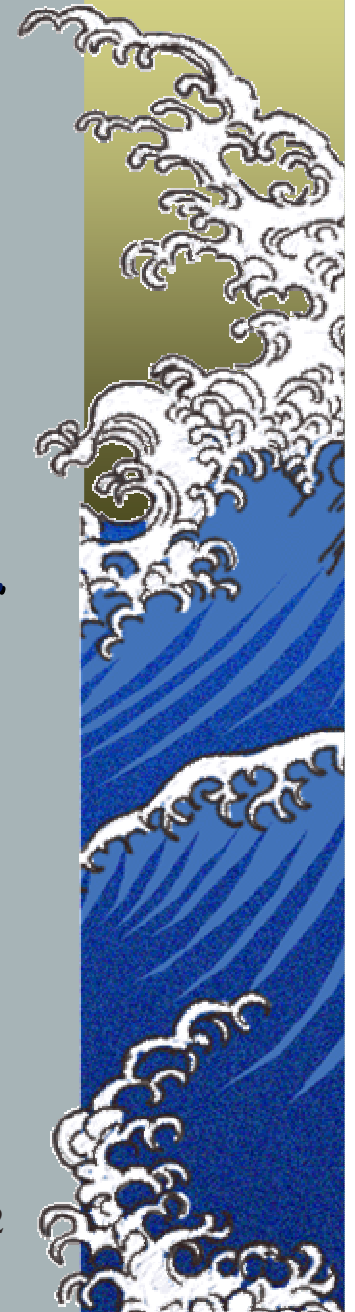
Workspaces

- *There's one workspace per client.*
- *It's an XML document that contains :*
  - *The list of selected servers*
  - *The user preferences (language…)*
  - *One or more saved meta-data corpus*
  - *An history of user actions*
- *A specific servlet is used to load and save the user's workspaces*

Communication between LR servers

- *Implemented through the HTTP Protocol*
  - *assure that every server is reachable from any point of the network*
  - *simplifies the installation of the server (no firewalls problems)*

34
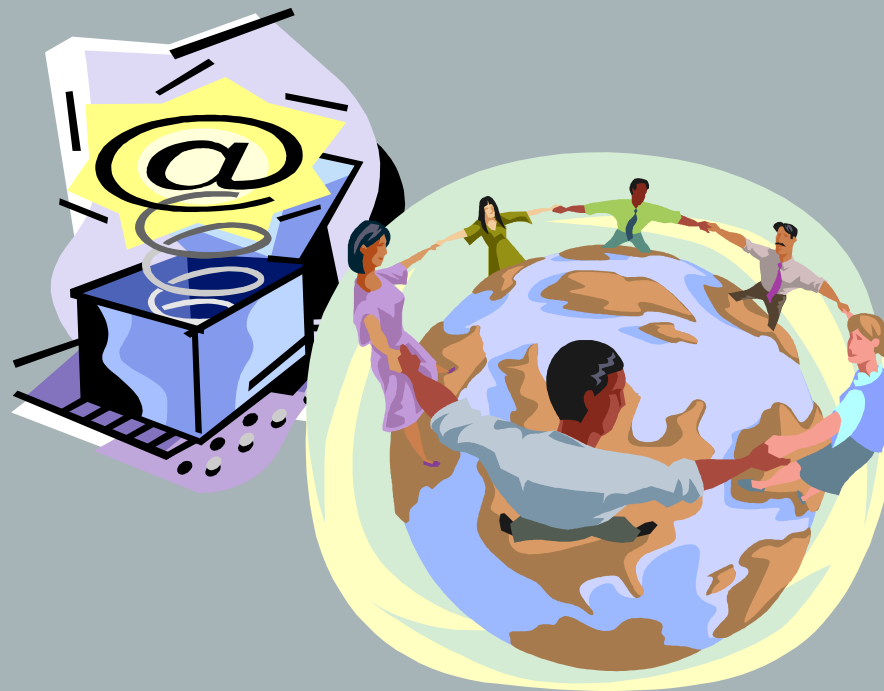
From the client side

- *A server is reachable through a simple Web browser, via the HTTP protocol*
- *After user registration, every interaction with the server take place in a multi-purpose applet :*
  - *selection of the servers*
  - *queries on meta-data and selection of linguistic resources*
  - *use of linguistic tools on the selected corpus*
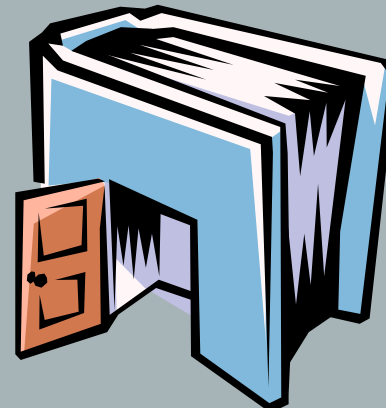- *The graphic interface is based on the Swing technology*
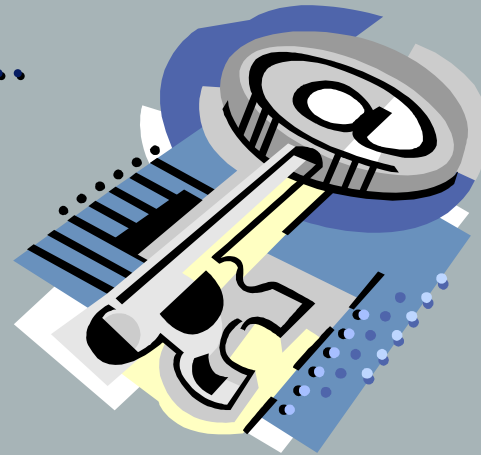
The "Digital Museum Project"

Main Objectives

- *Development of crucial techniques for the efficient producing, storage, and retrieval of XML-based documents*
  - *XLink, XPointer, XSLT, XSL FO, RDF,...*

Main Objectives

- *Development of a general XML-based distributed software architecture for accessing and sharing multimedia resources*
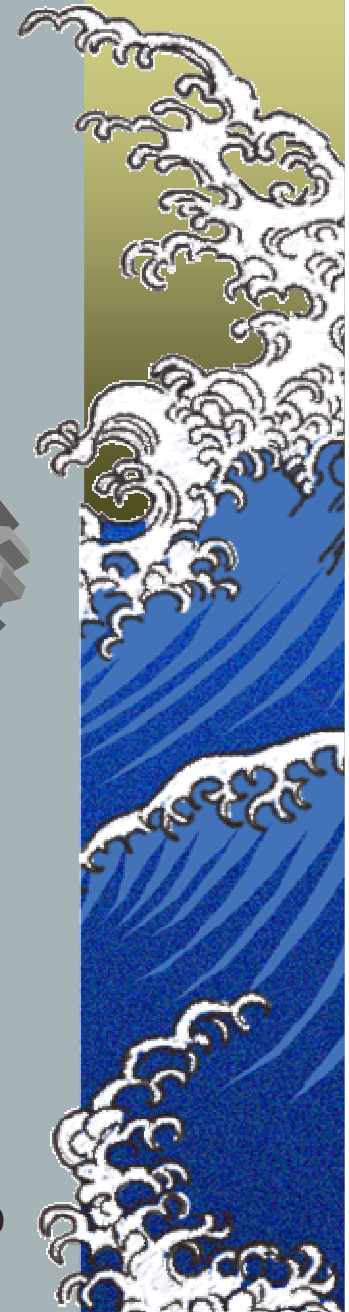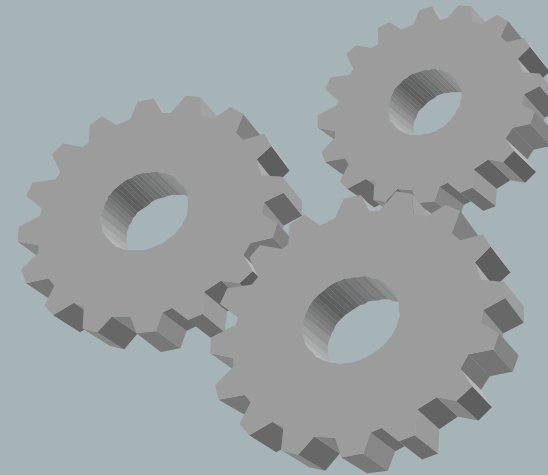 - *SOAP & Web Services, ...*

Main Objectives

- *Design of optimal synchronized hypermedia document models for digital museum exhibition*
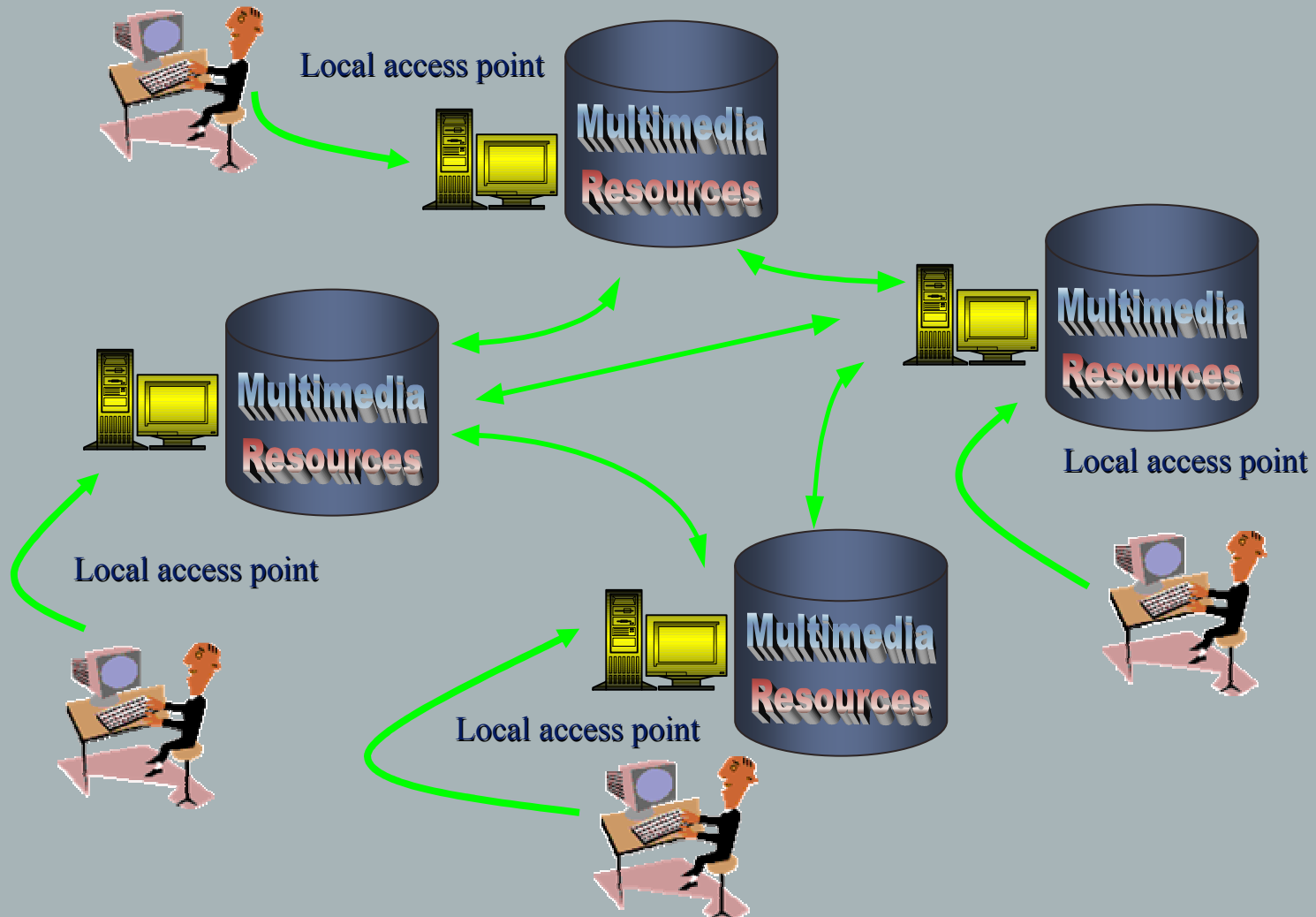 - *SMIL, ...*

System Architecture Issues

- *A multilevel architecture :*
  - *Administrator Level*
    - *Organizing LOCAL media*
  - *Exhibition Design Level*
    - *Organizing exhibitions*
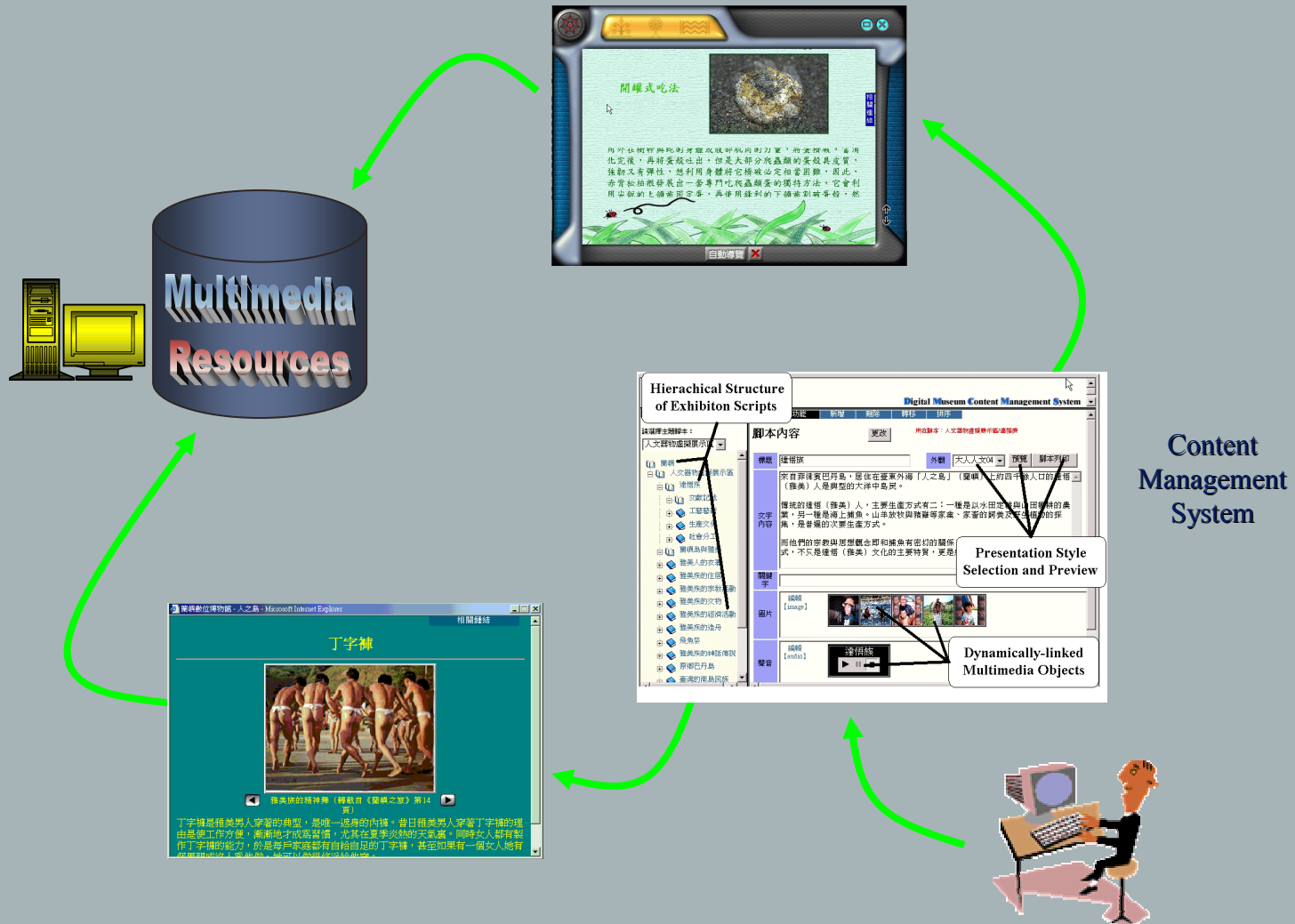 - *Final User Level*
    - *Accessing exhibitions*

System Architecture Issues

Local access point

Multimedia Resources

Local access point

Multimedia Resources

Local access point

Multimedia Resources

Local access point

Multimedia Resources

Administrator Level

Content Management System

Hierachical Structure of Exhibiton Scripts

Presentation Style Selection and Preview

Dynamically-linked Multimedia Objects

Administrator Level

- *Content Management System*
  - *Multimedia Resources Database Feeding*
  - *Multimedia Resources Integration*
  - *Digitization Process*
  - *Multimedia Resources Composition*

43

Exhibition Design Level

Local access point

Multimedia Resources

Multimedia Resources

Multimedia Resources

Multimedia Resources
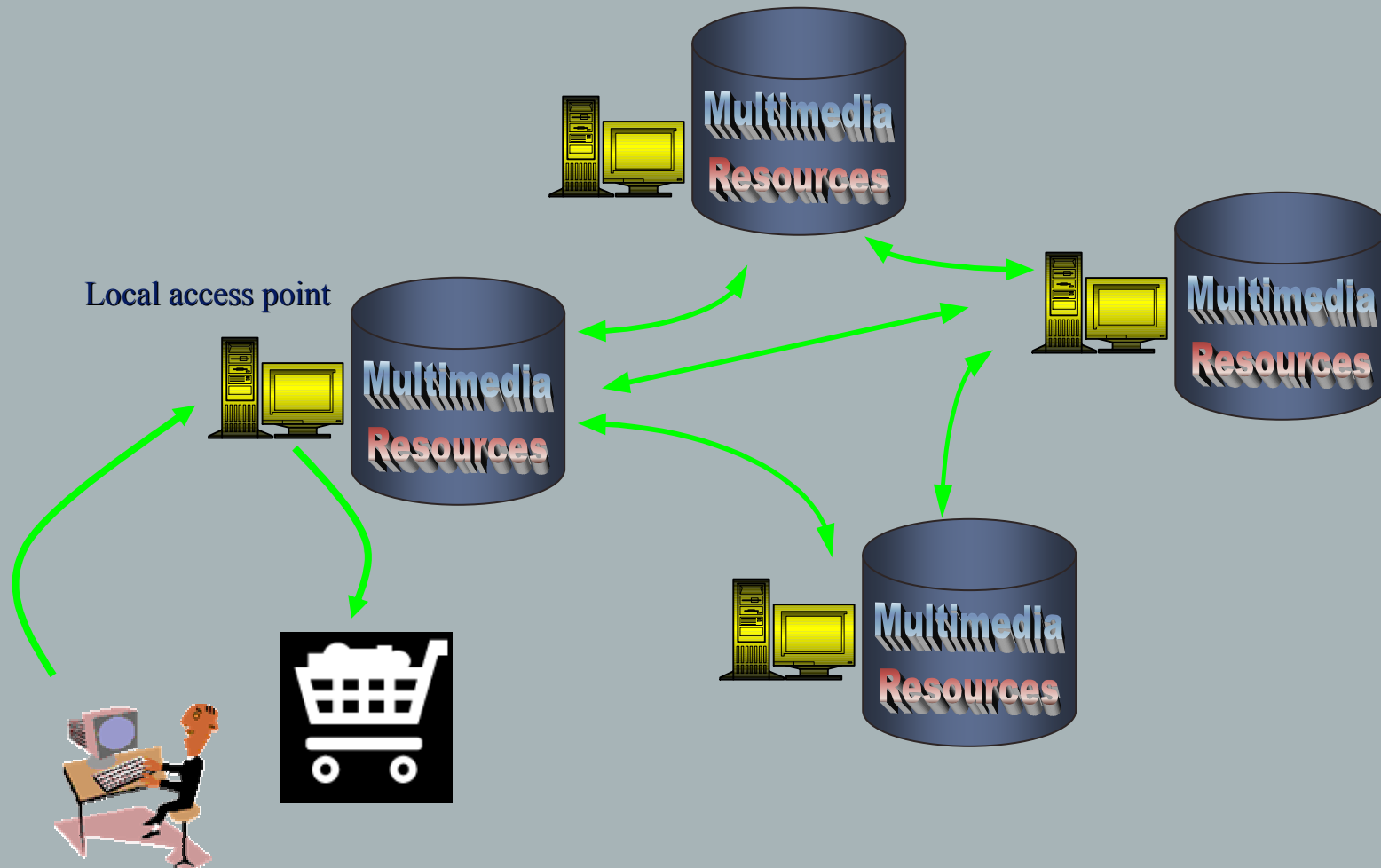
Exhibition Design Level

- *Building exhibitions by collecting multimedia resources allover the network*
  - *Selection of the working servers*
  - *Selection of multimedia resources ("Shopping Cart Paradigm")*
  - *Manipulation of the selected resources*
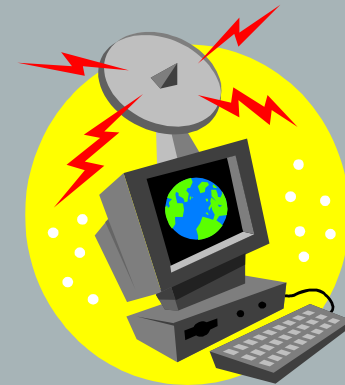- *"Local" access for all multimedia resources*

45

Final User Level

Final User Level

- *Web Browser-based access*
- *"Direct" access to exhibitions containing "distributed multimedia resources"*

47

Technical Approach

- *XML-based technologies*
  - *XML, XSLT, XSL FO, SMIL, RDF, ...*
- *Platform Independent Software Applications*
  - *Java*
- *Distributed Software Applications*
  - *SOAP*

48

XML
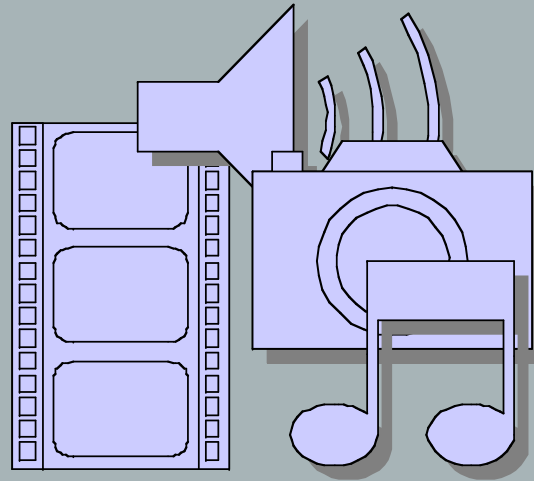
- *Resource Description Framework (RDF)*
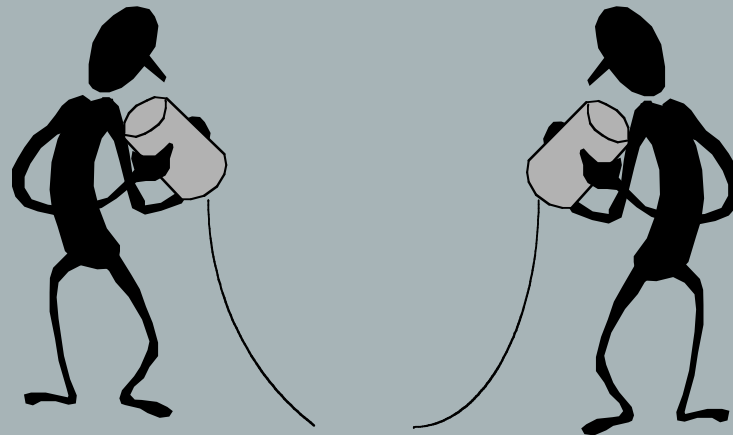 - *Meta Data*
  - *Ontologies*
  - *Semantics*

XML

- *XML will be used for normalizing the presentation of multimedia resources (SMIL, …)*

XML

⮝ *XML will also be used for normalizing the protocols we will use to exchange information between all servers in the network (SOAP, …)*
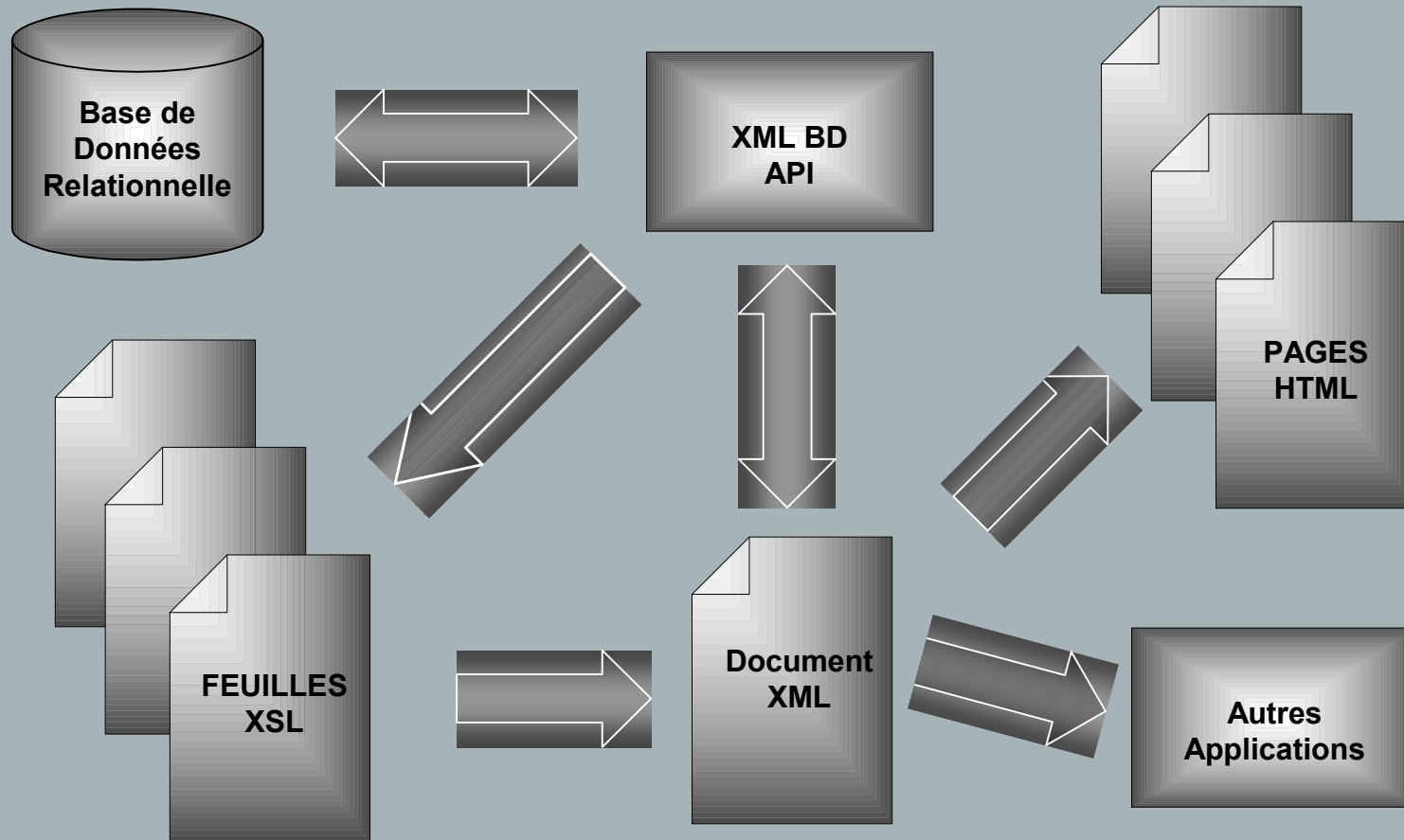
51

XML

- *XSLT & XSL FO*
  - *Transformations*
  - *Style Sheets*
  - *PDF documents*

XML

Base de
Données
Relationnelle

XML BD
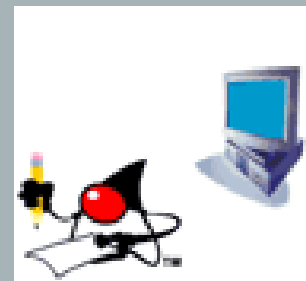API

PAGES
HTML

FEUILLES
XSL

Document
XML

Autres
Applications

Expected Results

- *Implementation of a Java-XML-based digital museum content management system functioning properly both taiwanese and french computer/network environment*

54

Synthesizing Remarks

- *Three major axes :*
  - *Design of Optimal Synchronized Hypermedia Documents for Digital Museum Exhibitions (SMIL, ...*
  - *Optimal Storage and Retrieval of XML-based Documents (XML, XLink, XPointer, XSLT, XSL FO, RDF, ...)*
  - *XML-based Distributed Software Architectures (SOAP & Web Services)*

Questions