Design of The PORTA EUROPA Portal (PEP) Pilot Project

Marco Pirri

Maria Chiara Pettenati

**Electronics and Telecommunications Department
University of Florence (Italy)**

**Library
European University Institute (EUI)
San Domenico di Fiesole
Florence (Italy)**

The aim of the project

- **To conceive a federation system to handle heterogeneous data sources**

- **In the framework of the PEP project**

- **Goal: to implement a PEP pilot project on a restricted domain - > history**

The problem of heterogeneous data sources

- Different format of data (Text, Audio, Video…)

- Different types in the same format (for example in text format: Pdf, Doc, Txt…)

- Different format to describe data (metadata)

- Different protocols to retrieve data, build services

Porta Europa Portal pilot project

- **Porta Europa purpose** is to present the EUI (European University Institute) on the Web
  - as a leader in the "European debate"
  - as a natural gateway to high-quality research information in the areas of law, political and social sciences, economics and history

- **What is it?** a specific Portal integrated inside the EUI Web Site
  - offering opportunities to link the currently dispersed information sources

- **The Pilot Project**: The Library will create the first prototype based on <u>three history projects</u>.

Voices on Europe - Resource 1/3

- It's an oral Archive of Interviews to Important European personalities - http://wwwarc.iue.it/webpub/

- Details: Access Database with 5 representative tables, Input made by Staff

- Today: interviews papers available in jpeg format (only on CD-ROM) and cassette available only at request to Archive

➔ Future: interview papers available in pdf format (online) and audio available (online with authorised access)

VL History Project -Resource 2/3

- **WWW VL History Project is part of Virtual Library the oldest catalog of the web, started by Tim Berners-Lee -** http://vlib.iue.it/

- **Run by a loose confederation of volunteers, who compile pages of key links for particular areas in which they are expert**

- **Today: 11 main Projects and 6 sub-projects are involved, all pages are in HTML and database is not implemented**

- **The Future: Create a database, transfer old data, build new and homogeneous output, administration interface.**

Biblio, Innopac - Resource 3/3

- **Biblio is the main catalogue of the EUI Library (more than 260.000 bibliographic records)**
  http://www.iue.it/LIB/Catalogue/

- **Based on Innopac Automation System**

- **Today: Staff administrate catalogue, but still some limits on extracting and elaborating data from catalogue**

- **Future: Innopac module, OAI compliant**

Summary of the three resources

- **Voices on Europe** - Oral Archive of interviews, access database

- **WWW Virtual Library History Project** - Website with a collection of links selected by experts in history fields organized by categories

- **Biblio, Innopac** - Main catalogue of Library of European University Institute - private database

Analysis of the three resources

Characterisitc of the archive	Voices on Europe	Virtual Library	Biblio Library catalog
Data Objects	Audio, tran.	HTML pages	Records
Collection of metadata structures	Archive organized in Access DB	Archive is structured in web pages	Proprietary DB, USMARC Format
Collection of services	Access: SQL queries, DB staff management, no log on or statistic	Access: through Web, managed by project admin. No log on or stat.	Information management through INNOPAC Lib. Automation System
Domain focus	European history		
Community of users	Everybody for information Search, Restricted access for full documents consultation, Admin.		

The adopted approach

- Metadata -> Dublin core

  - **International Standard based on XML**
  - **Extensibile**

- Protocol -> Open Archives Initiative

  - **Data and Service Provider**
  - **Harvesting (Retrieve data from Repositories)**

  - **More details in the project home page**

Meta Resource Card

PWP

DC element	Voices On Europe	VL History Project	Biblio Innopac
Title	Interviewee's name	Title	Title
Creator	Name of Interviewer	Author	Author
Subject	Level 1,2,3 (eurovoc)	Type 3	Subject
…	…	…	…
Rights	User Profile	Free	User Profile

User Profile

| Function	User
General Administration (Information management)	Administrator and Project leaders
Information search	Public
Full information access	Internal Users (iue members, professors, studends, etc.)
Restricted information access (restriction is due to property right on some resources contained in the archives)	External users, groups of users
Personalised services	Registered users

Architecture

PWP

- Layered Structure

- Layer provide services to the upper layer

- OAI federation Database

- Independent data source
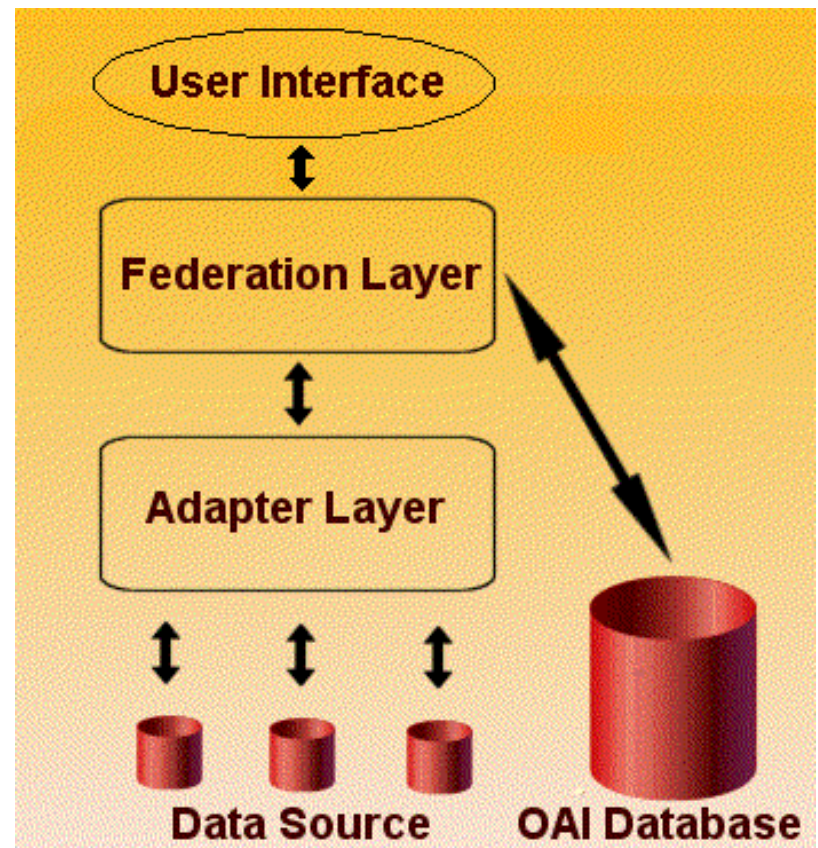
Data Source Layer

- **Voices on Europe - Oral Archive of interviews, access database**

- **WWW Virtual Library History Project - Website with a collection of links selected by experts in history fields organized by categories**

- **Biblio, Innopac - Main catalogue of Library of European University Institute - private database**
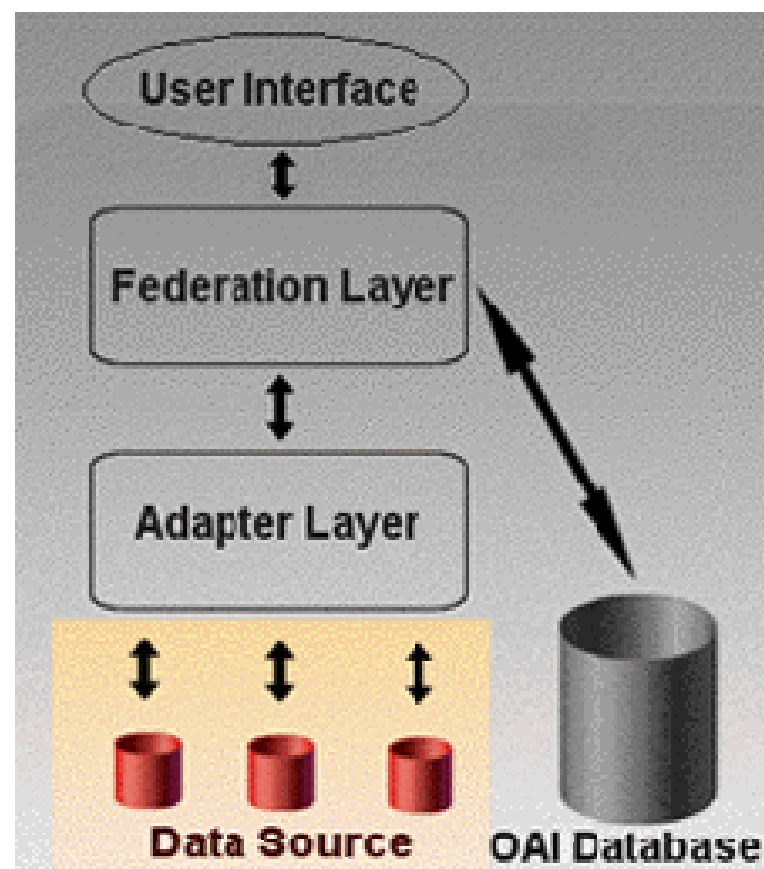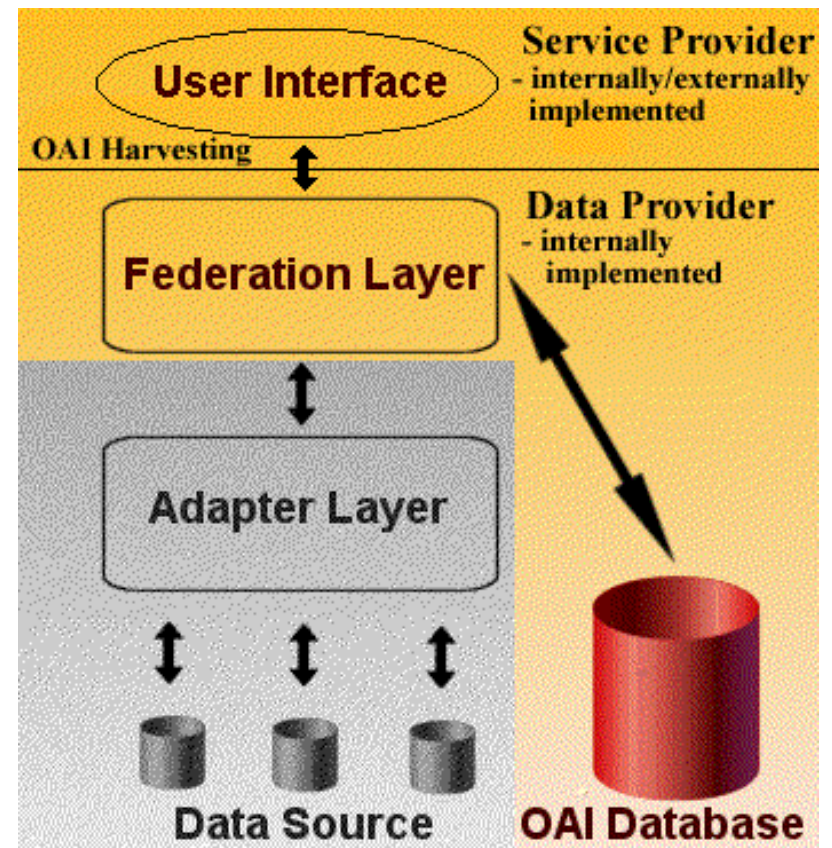
Adapter Layer

- **Extracting tools, SQL queries**

- **Extracting of data fields**
  - **automatic process**

- **Update Frequency**
  - **occasionally (every day/week/...)**
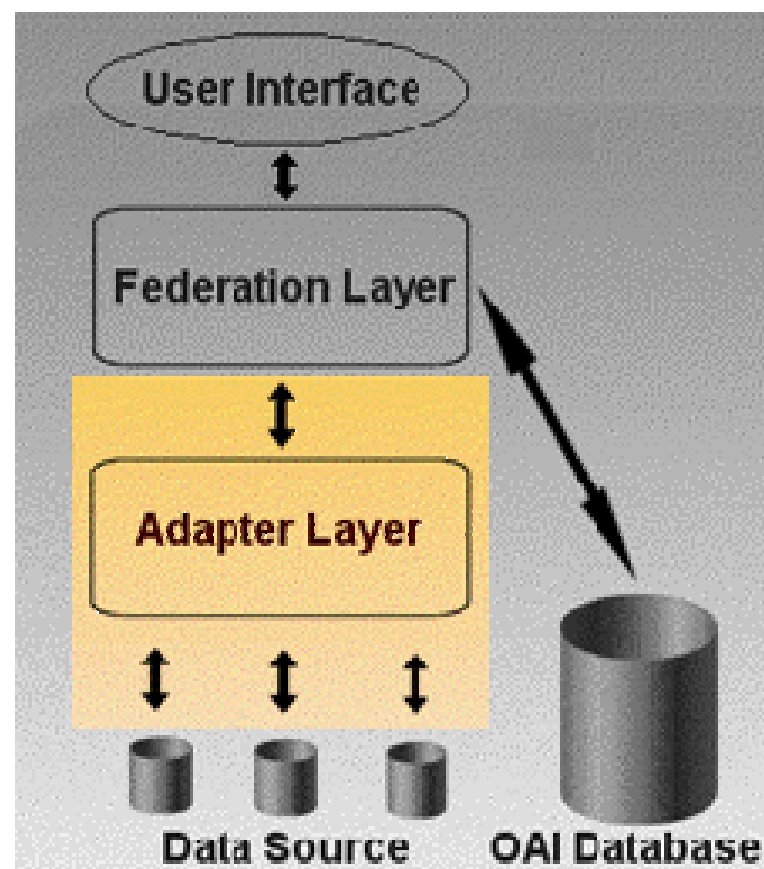
Federation Layer & User Interface

- **Meta Resource Card that map the fields extracted**

- **Management of OAI repository**

- **Implementation of OAI-PMH Protocol**

- **Communication with user interface (Service Provider)**

Eui, OAI Repository
XML Response (request: GetRecord)

PWP

```
<?xml version="1.0" encoding="UTF-8"?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/"    ← PROTOCOL VERSION (OAI-PMH 2.0)
...   ← PROTOCOL NAMESPACE DECLARATIONS, RESPONSE DATE,  METADATA PREFIX...
<GetRecord>    ← RESPONSE TO GETRECORD REQUEST
...    ← HEADER ( IDENTIFIER, DATE, SET)
 <metadata>   ← METADATA
   <oai_dc:dc .... ← RECORD NAMESPACE DECLARATIONS
     ...    http://www.openarchives.org/OAI/2.0/oai_dc.xsd">   ← SCHEMA LOCATION
   <dc:title>titolo1</dc:title>         ← DUBLIN CORE METADATA
   <dc:creator>autore</dc:creator>
   <dc:subject>argomento</dc:subject>
   <dc:description>descrizione</dc:description>
   <dc:publisher>editore</dc:publisher>
   <dc:date>data</dc:date>
   <dc:type>tipo</dc:type>
   <dc:format>formato</dc:format>
   <dc:identifier>identificativo</dc:identifier>
   <dc:source>source</dc:source>
   <dc:language>lingua</dc:language>
   <dc:relation>relation</dc:relation>
   <dc:coverage>coverage</dc:coverage>
   <dc:rights>diritti</dc:rights>
   </oai_dc:dc>
 </metadata>
...    ← CLOSING TAGS
</GetRecord>
</OAI-PMH>
```

Conclusion

- **Current state**
  - **Analisys of Resources (data source layer)**
  - **Meta Resource card**
  - **PEP Architecture**
  - **Test of OAI Repository (federation layer)**
- **Work in progress**
  - **Extracting tools (adapter layer)**
  - **User Interface (OAI Service Provider)**

- **PEP Homepage:**
  http://www.iue.it/Personal/Staff/pirri

PEP pilot project Home Page
http://www.iue.it/Personal/Staff/pirri

"Voices on Europe" Mapping

Dublin Core Element	Voices on Europe
Title	Interviewee's surname
Creator	Name of Interviewer
Subject	Level 1,2,3 (eurovoc)
Description	Number of Pages
Publisher	Eui
Contributor	-
Date	Date of recording
Type	Audio/Testo
Format	Pdf
Identifier	Url
Source	-
Language	Language
Relation	Additional Material
Coverage	-
Rights	User Profile

"Virtual Library History Project" Mapping

Dublin Core Element	Virtual Library
Title	Title
Creator	Author
Subject	Type 3
Description	Abstract
Publisher	Type 1
Contributor	-
Date	Date of input
Type	Text
Format	Html
Identifier	Url
Source	-
Language	English
Relation	-
Coverage	-
Rights	free

"Biblio - Innopac" Mapping

Dublin Core Element	Biblio
Title	Title
Creator	Author
Subject	Subject
Description	Note
Publisher	Imprint
Contributor	-
Date	Cat Date
Type	Text
Format	Pdf
Identifier	Isbn
Source	-
Language	Lang
Relation	-
Coverage	-
Rights	User Profile

Mapping Summary – Meta Resource Card

Dublin Core Element	Voices on Europe	Virtual Library	Biblio
Title	Interviewee's surname	Title	Title
Creator	Name of Interviewer	Author	Author
Subject	Level 1,2,3 (eurovoc)	Type 3	Subject
Description	Full text Interview	Abstract	Note
Publisher	Eui	Type 1	Imprint
Contributor	-	-	-
Date	Date of recording	Date of insertion	Cat Date
Type	Video/Audio/Testo	Text (Html)	Text
Format	Pdf	Html	Pdf
Identifier	Url	Url	Isbn
Source	-	-	-
Language	Language	English	Lang
Relation	Additional Material	-	-
Coverage	-	-	-
Rights	User Profile	free	User Profile

"Voices on Europe" - Table Interviews

Reference code	Collection title	Interviewee's surname	Interviewee's forename	Title
Date of birtd	Place of birtd	Nationality	Sex	Biographical notes
Type of recorder	Total number of tape	Type of format	Type of tape	Speed
Clearance Cassettes	Copyright agreement	Individual consultation	Public broadcast	Partial or entire publication
Pages	Topics-en	Full name	imgfo	appo1
Date(s) of recording (from)	Date(s) of recording (to)	Location of interview	Name of interviewer	Language
Original or copy	Transcription	Number of pages	Additional material	Clearance Transcription
Partial or entire reproduction	Closed until	Names of persons	Comments	Topics
appo2	appo3	appo4	Pages-en	tmp
dd1	mm1	yy1	National program	

Fields Description

- **Reference Code:** Code that identify the interview
- **Interviewee's surname:** The surname of the person interviewed
- **Interviewee's forename:** The forename of the person interviewed
- **Date(s) of recording (from):** The date of startin of interview

- **Name of interviewer:** The name of the person that made the interview
- **Language:** The language of the interviewed person
- **Number of pages:** The number of pages where the interviewed person is mentioned
- **Additional material**: The additional material with the interview

"Voices on Europe" – Table Evoc and description

Eurovoc	Language	level1	level2
level3	level4	level5	todos

- Level 1, 2, 3 – Thesaurus Eurovoc. This European multilingual listing of expressions has been developed for one-to-one indexing of documentary information from the European institutions

Dublin Core Fields Mapping

Felds Extracted from Tables

1) Reference Code → Identifier
2) Interviewee's surname and name → *Title*
3) Date(s) of recording (from) → *Date*
4) Name of interviewer → *Creator*
5) Language → *Language*
6) Number of pages → *Description*
7) Additional Material → *Relation*
8) level 1,2,3 (eurovoc)→ *Subject*

Fixed Fields

9) EUI→ *Publisher*
10) Audio/Text→ *Type*
11) Pdf→ *Format*
12) User Profile→ *Rights*

Fields Not Used

13) *Contributor*
14) *Source*
15) *Coverage*

"Voices on Europe" Mapping

Dublin Core Element	Voices on Europe
Title	Interviewee's surname
Creator	Name of Interviewer
Subject	Level 1,2,3 (eurovoc)
Description	Number of Pages
Publisher	Eui
Contributor	-
Date	Date of recording
Type	Audio/Testo
Format	Pdf
Identifier	Url
Source	-
Language	Language
Relation	Additional Material
Coverage	-
Rights	User Profile

"Voices on Europe" - Table Interviews

Dublin Core Element	Normalization
Title	No
Creator	No
Subject	No
Description	No
Publisher	No
Contributor	Not Used
Date	Yes – UTC Format ISO8601 – OAI
Type	Yes – Audio / Text
Format	Yes – Pdf / mp3
Identifier	No
Source	Not Used
Language	No
Relation	No
Coverage	Not Used
Rights	Yes – User Profile / free