

I D _ T E C H N O L O G Y

OAI Data Provider

SOFTxBIB

6 Marzo 2008

Costantino Landino
Business Unit Manager
c.landino@idtech.it



L'OAI Data Provider è un software che permette in pochi passaggi di rendere compatibile un database con il protocollo OAI-PMH .

Questo software e' stato ideato e realizzato sfruttando le esperienze compiute in un progetto di ricerca calandole in un contesto operativo di integrazione reale.

Il software è distribuito in modalità gratuita dall'Osservatorio Tecnologico del Ministero per i Beni e le Attività Culturali (www.otebac.it), grazie ad un accordo con la Direzione Regionale dei Beni Culturali della Lombardia (www.lombardia.beniculturali.it).

Lo sviluppo e' partito da un software open source esistente (UIUC OAI Metadata Harvesting Project) e si è realizzato un modulo di estrazione e gestione di metadati che fosse utilizzabile senza necessità di programmazione.

I profili di metadati ad oggi utilizzabili sono Dublin Core (come da specifiche OAI) e PICO AP (Portale Italiano delle Cultura).

La gestione e la configurazione dell'applicazione avviene attraverso file XML e XSL con i quali è possibile modificare i set di estrazione OAI, il mapping degli schemi xml Dublin Core e le decodifiche dei vari campi.

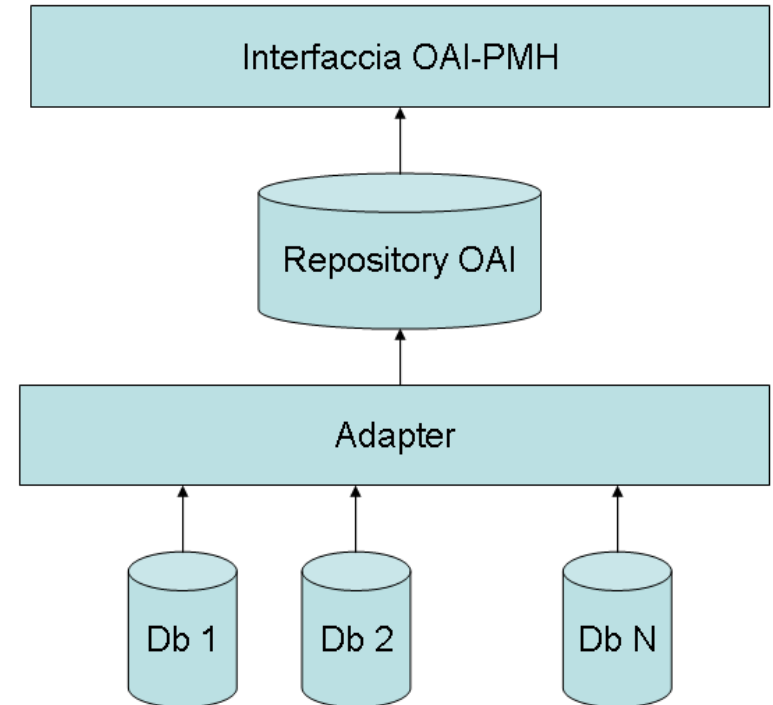


Il sistema è organizzato in due componenti: l'Adapter e l'interfaccia OAI-PMH. E' previsto un repository OAI per contenere tutti i metadati destinati ad essere pubblicati.

L'**Interfaccia OAI-PMH** è fornita dall'UIUC OAI Metadata Harvesting Project.

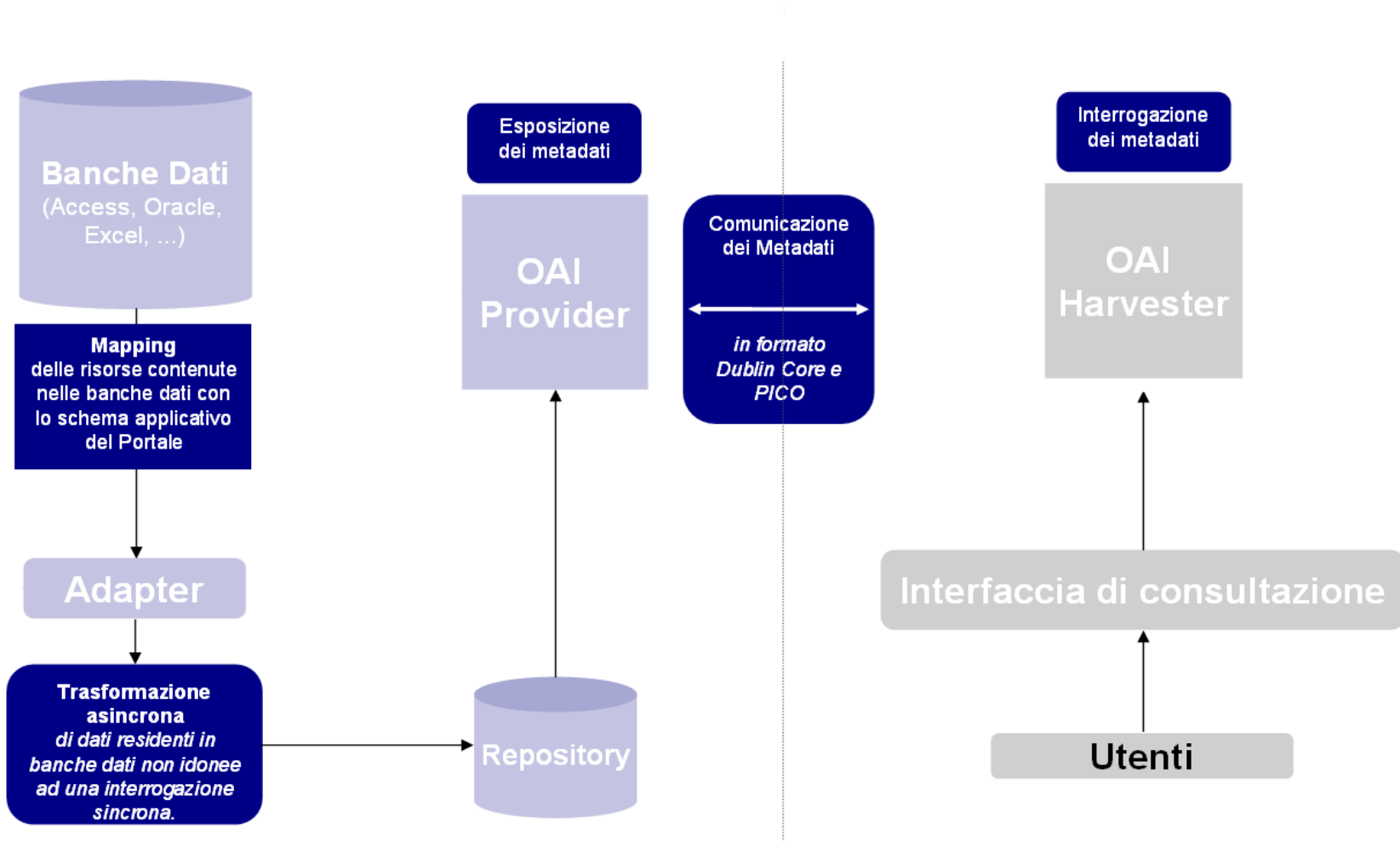
L'**Adapter** è una componente che consente l'aggiornamento del Repository OAI, effettuando le opportune query sui database sorgente, estraendo i risultati della ricerche e rielaborandoli. L'adapter si connette a ciascun datasource e al repository OAI via JDBC o attraverso il JDBC-ODBC bridge.

Il **Repository OAI** ha il compito di memorizzare le informazioni estratte dai singoli database. I dati sono organizzati in una struttura utile per la pubblicazione in formato Dublin Core e PICO.



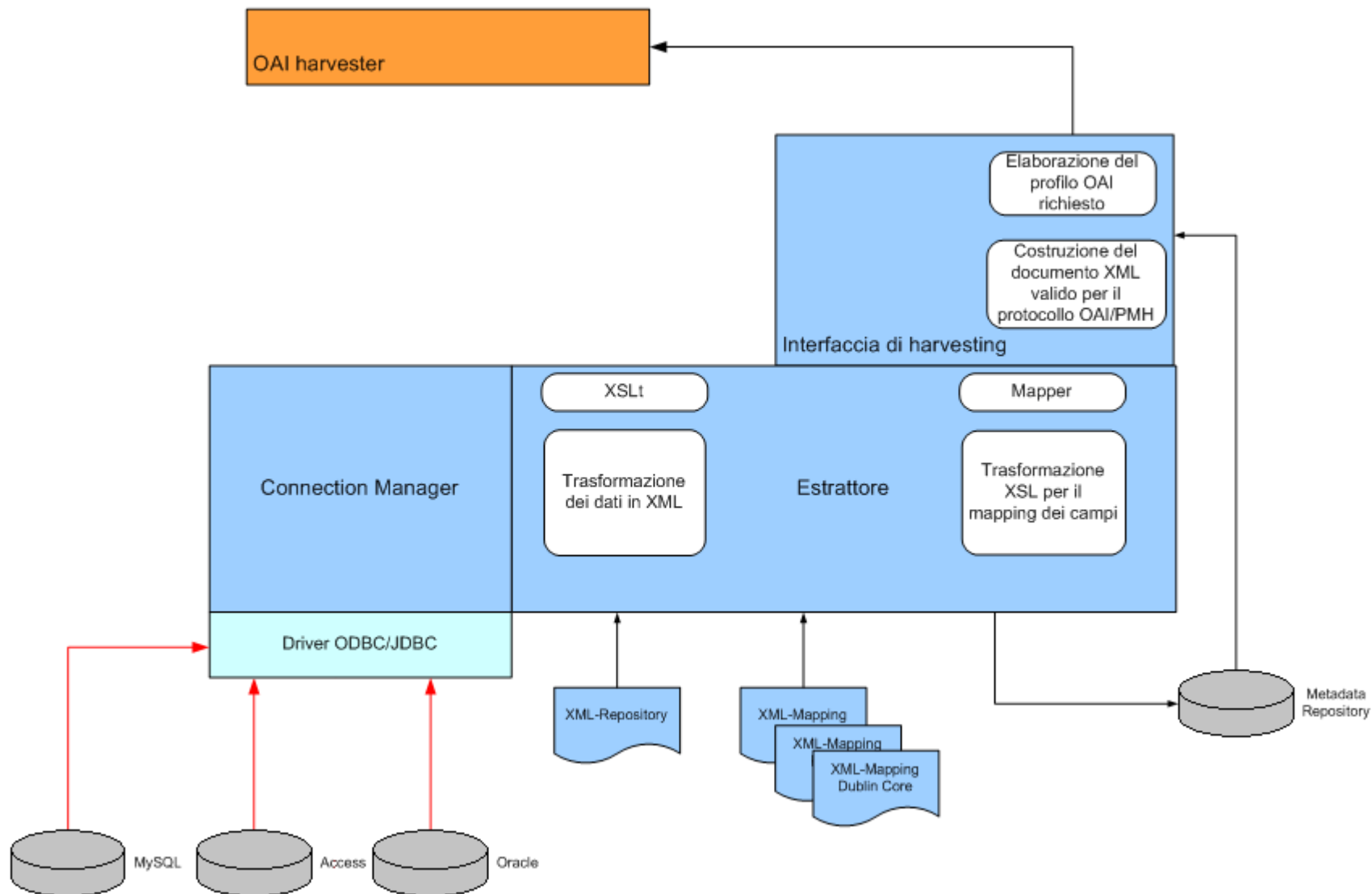


L'architettura e flussi di informazioni





Dettaglio dell'architettura



La procedura preliminare per l'utilizzo dell'adapter consiste nel configurare tutte le regole per l'estrazione e l'elaborazione dei dati per ciascuna applicazione sorgente.

I passaggi sono :

1 - Identificazione del database sorgente (datasource)

2 - Configurazione del mapping

- **<connection>**

specifica i parametri di connessione al database sorgente

- **<queries>**

specifica le query SQL da effettuare recuperare i dati che verranno memorizzati nell'OAI repository. Nell'ambito di ciascuna query si effettua la dichiarazione di variabili al fine di effettuare elaborazioni sui dati estratti.

- **<fields>**

contiene il mapping vero e proprio tra i dati recuperati dal datasource e gli elementi OAI da dichiarare.

3 - Generazione dei metadati

4 - Pubblicazione dei metadati



Identificazione del database sorgente

Le tipologie di datasource testati sono **Oracle, SQL Server, MySQL, Access, Excel, Livelink Collection Server (Basis).**

Esempio di configurazione

```
<connection>
  <tipo-datasource>
    MySql
  </tipo-datasource>
  <driver>
    org.gjt.mm.mysql.Driver
  </driver>
  <database>
    jdbc:mysql://<ip>/<db>
  </database>
  <username>
    dbuid
  </username>
  <password>
    dbpwd
  </password>
</connection>
```



Configurazione sezione queries del mapping 1/3

La sezione queries serve a definire le query SQL da effettuare sul datasource per recuperare i dati che saranno utilizzati per costruire i metadati Dublin Core o PICO. Nell'ambito di ciascuna query si effettua la dichiarazione di variabili al fine di effettuare elaborazioni fondamentali sui dati estratti.

La sezione può contenere più definizioni di queries.

All'interno delle dichiarazioni si possono definire variabili locali o globali. Il valore di ciascuna variabile viene ricalcolato ad ogni riga del risultato della query di riferimento. Il valore assegnato alla variabile può essere una concatenazione di testo, variabili già dichiarate in precedenza e campi presenti nella query SQL.

Nella sezione **queries** ed in quella **fields** è possibile definire alcune operazioni come:

Decod: indica che il valore calcolato deve essere decodificato secondo il mapping specificato da un file xml esterno;

id: consente di specificare l'ordinamento nelle dichiarazioni delle variabili e/o query;

nullify-if: indica che se l'espressione nella variabile assume uno dei valori specificati nell'attributo, allora alla variabile viene assegnato valore nullo.

Substring: indica la sottostringa che va estratta dal campo specificato.

Concat: indica che i valori vanno concatenati con il separatore indicato

Firstvalue: indica che il campo va valorizzato con il primo valore non vuoto.



Configurazione sezione queries del mapping 2/3

Esempi di queries su di un database relazionale su più tabelle.

```
<queries>
<query id="1" key="Maincounter">
<sql-clause> SELECT MainCounter, ogtn, ogtc, dmo, eri, dst, repa, reps, repf,
repp, ogtt, evie, idb, htm, pvcc, pvcl, ubvd, ubvn, ubvk, crdr, crdx, ORDER BY ID
</sql-clause>
<var-declaration>-----</var-declaration>
</query>
<query id="2" key="SubCounter" foreign_key="MainReference" parent="1">
<sql-clause>
SELECT MainReference, SubCounter, decd FROM Sub_dec ORDER BY
MainReference, SubCounter
</sql-clause>
<var-declaration>-----</var-declaration>
</query>
<query id="3" key="SubCounter" foreign_key="MainReference" parent="1">
<sql-clause>
SELECT MainReference, SubCounter, algi FROM Sub_alg ORDER BY
MainReference, SubCounter
</sql-clause>
-----
```



Configurazione sezione queries del mapping 3/3

```
<query id="1" key="idunivoco">
<sql-clause> SELECT IDUnivoco, autore, titolo, titolo_alternativo, note_tipografiche,
descrizione_fisica, curatore,tipografo,regione,località,data_edizione FROM DB
where Form='SkBib' ORDER BY idunivoco </sql-clause>
<var-declaration>
<var id="01" name="idunivoco">@{idunivoco}</var>
<var id="02" name="identifier">db3/@{idunivoco}</var>
<var id="03" name="autore">@{autore}</var>
<var id="04" name="titolo">@{titolo}</var>
<var id="05" name="titolo_alternativo">@{titolo_alternativo}</var>
<var id="06" name="note_tipografiche">@{note_tipografiche}. -</var>
<var id="07" name="descrizione_fisica">@{descrizione_fisica}</var>
<var id="08" name="descrizione">@{descrizione_fisica}</var>
<var id="09" name="curatore">@{curatore}</var>
<var id="10" name="tipografo">@{tipografo}</var>
<var id="11" name="regione"> @{regione} </var>
<var id="12" name="localita">@{localita}</var>
<var id="13" name="istat" decod="decod-istatcomuni-utf8.xml"
nomatch="">@{localita}</var>
<var id="14" name="url_scheda">http://XXX.XXX.XXX/OpenDocument?
@{idunivoco}</var>
<var id="15" name="url_img">http://XXX.XXX.XXX/@{idunivoco}=jpg</var>
```



Configurazione sezione field del mapping

La sezione field contiene il mapping vero e proprio tra i dati recuperati dal datasource e gli elementi OAI Dublin Core da dichiarare.

```
<fields>
<field dest="identifier" mandatory="yes">      #{identifier}      </field>
<field dest="dc_identifier">                  #{idunivoco}      </field>
<field dest="dc_language_ISO639_2">          ITA              </field>
<field dest="dc_type_dcmitype">              Physical Object </field>
<field dest="dc_subject_thesaurus">
  <values mode="concat" separator="|">
    <value id="1">                            Collezioni digitali </value>
    <value id="2">                            Manoscritti       </value>
  </values>
</field>
<field dest="dc_title"> #{titolo} </field>
<field dest="dcterms_alternative"> #{titolo_alternativo} </field>
<field dest="dc_description"> #{descrizione} </field>
<field dest="dcterms_spatial"> #{regione} </field>
<field dest="dcterms_spatial_istat"> #{istat} </field>
<field dest="dc_creator"> #{autore} </field>
<field dest="dc_rightsholder"> Proprietà Stato </field>
<field dest="dc_relation_uri"> #{url_scheda} </field>
</fields>
```



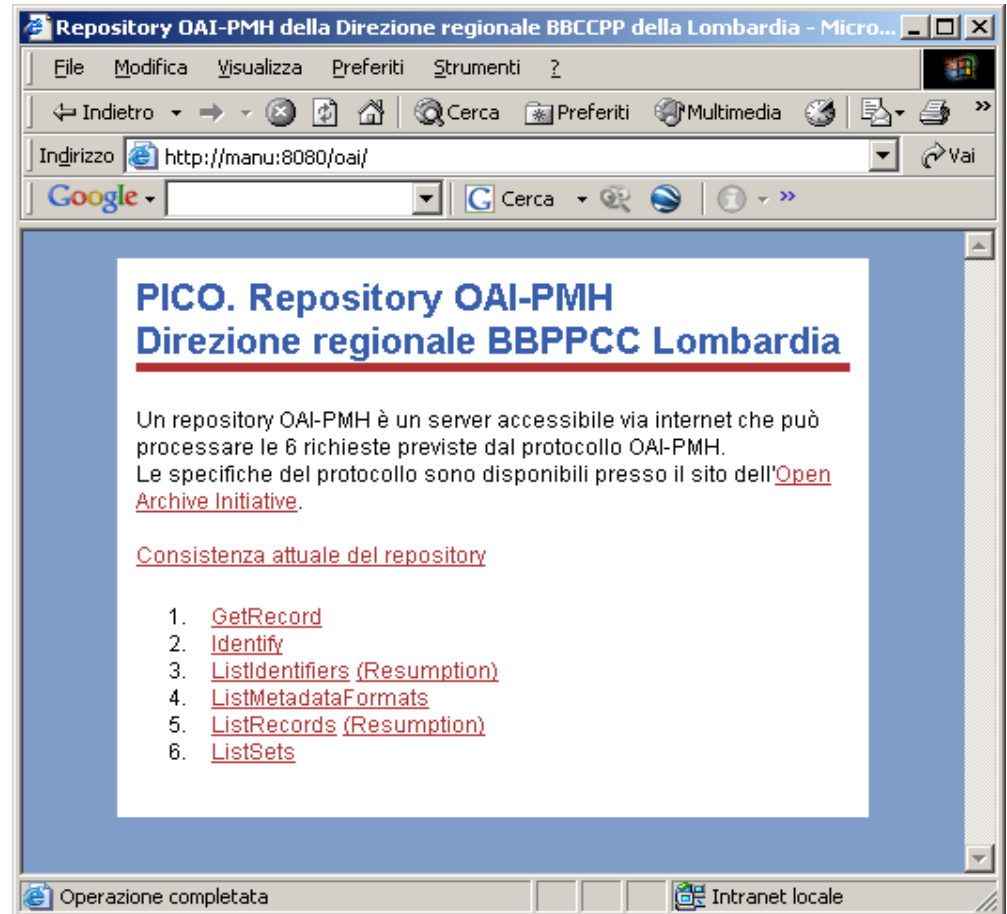
Generazione e pubblicazione dei metadati (1/2)

La procedura di generazione dei metadati utilizza tutte le specifiche dei passaggi precedenti e popola l'OAI Repository.

A questo punto i metadati sono disponibili per l'Harvesting attraverso un sistema OAI-PMH compliant

Per semplificare le operazioni di controllo dei risultati, si è sviluppata una semplice interfaccia di consultazione.

L'interfaccia permette di eseguire le richieste OAI (GetRecord, Identify, ListIdentifiers, ListMetadataFormats, ListRecords, ListSets) e di avere la consistenza attuale del repository suddivisa per tipologia di Set e di Formati.





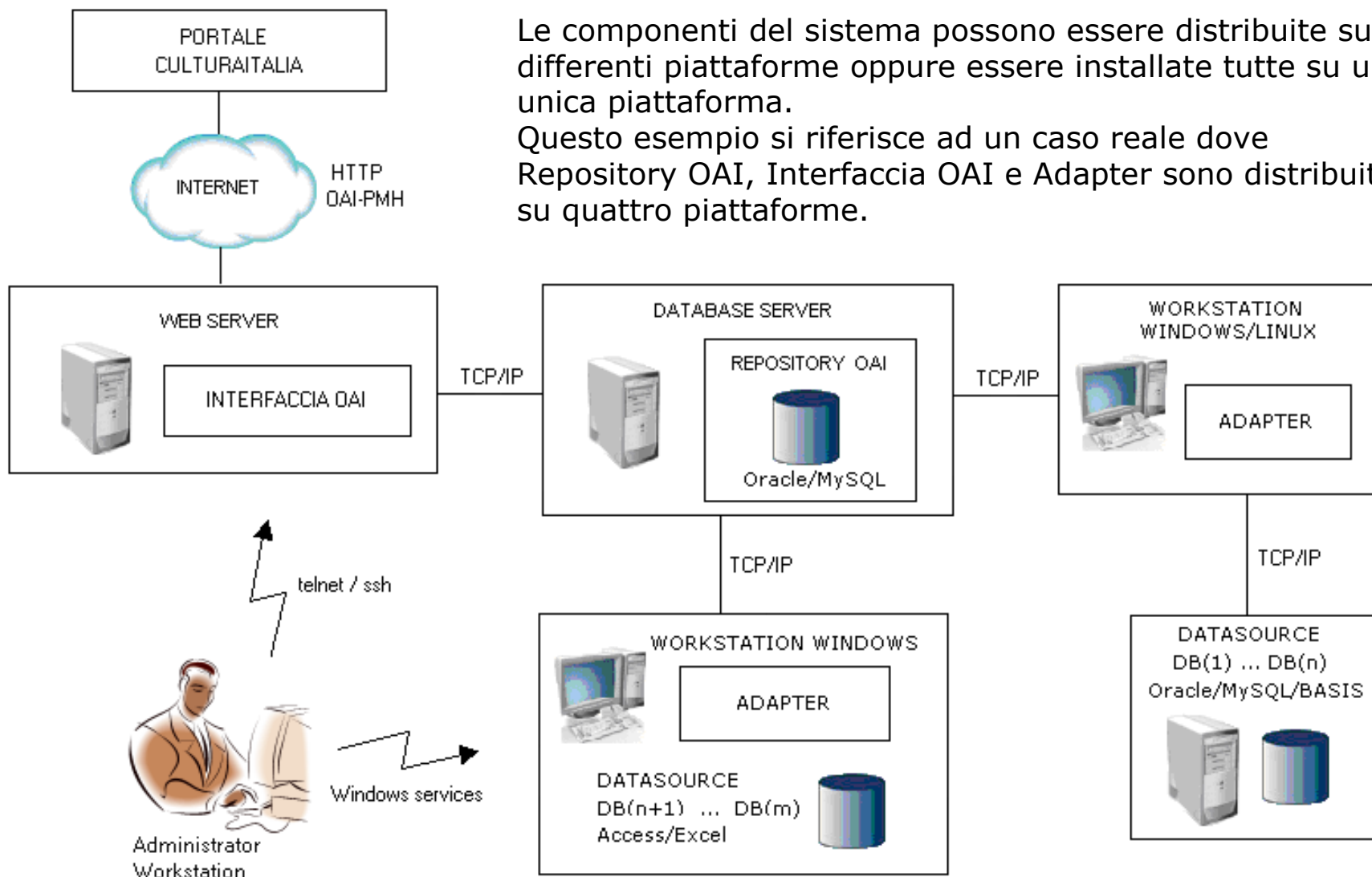
Generazione e pubblicazione dei metadati (2/2)

Esempio di richiesta:ListRecords

```
<?xml version="1.0" encoding="UTF-8" ?>
- <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2008-02-12T14:08:17Z</responseDate>
  <request verb="ListRecords" metadataPrefix="oai_dc">http://giove.lombardia.beniculturali.it:8080/oai/interfaccia.jsp</request>
- <ListRecords>
- <record>
+ <header>
- <metadata>
- <oai_dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title>IMMOBILE DENOM. FABBRICATO IN VIA BORGOVICO N. 182</dc:title>
  <dc:subject>NAME=Palazzi e ville;ID=20014007000000000</dc:subject>
  <dc:description>Bene architettonico di interesse storico artistico soggetto a tutela</dc:description>
  <dc:type>Physical Object</dc:type>
  <dc:identifier>DB01_0130750003</dc:identifier>
  <dc:relation>http://giove.lombardia.beniculturali.it:8080/oai/immagini/benia/DB01_0130750003a.jpg</dc:relation>
  <dc:coverage>NAME=Como;YEAR=2001;CODE=0301307510003</dc:coverage>
  <dc:coverage>COMO, 22100, VIA BORGOVICO 182</dc:coverage>
  <dc:coverage>Dati catastali: C.T. Fig. 4 - mapp. 68, 62, 67.</dc:coverage>
  <dc:coverage>east=1505142; north=5073681; projection=UTM</dc:coverage>
</oai_dc>
</metadata>
</record>
- <record>
+ <header>
- <metadata>
- <oai_dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:title>IMMOBILE DENOM. FABBRICATO IN VIA PRIMO TATTI N. 12-16</dc:title>
  <dc:subject>NAME=Palazzi e ville;ID=20014007000000000</dc:subject>
  <dc:description>Bene architettonico di interesse storico artistico soggetto a tutela</dc:description>
  <dc:type>Physical Object</dc:type>
  <dc:identifier>DB01_0130750004</dc:identifier>
  <dc:relation>http://giove.lombardia.beniculturali.it:8080/oai/immagini/benia/DB01_0130750004a.jpg</dc:relation>
  <dc:coverage>NAME=Como;YEAR=2001;CODE=0301307510003</dc:coverage>
  <dc:coverage>COMO, 22100, VIA PRIMO TATTI 12-16</dc:coverage>
  <dc:coverage>Dati catastali: Fig. 4 - mapp. 400 (sub. 1, 2, 3, 4, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23).</dc:coverage>
  <dc:coverage>east=1506378; north=5072992; projection=UTM</dc:coverage>
</oai_dc>
</metadata>
</record>
+ <record>
```



Scenario di integrazione

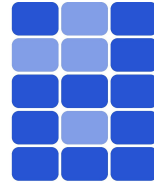


Le componenti del sistema possono essere distribuite su differenti piattaforme oppure essere installate tutte su un'unica piattaforma.

Questo esempio si riferisce ad un caso reale dove Repository OAI, Interfaccia OAI e Adapter sono distribuiti su quattro piattaforme.



info@idtech.it



I D _ T E C H N O L O G Y

Milano

Corso di Porta Romana, 6 - 20122 Milano
t +39 02 80 51 131
f +39 02 80 51 163

Roma

Via G. Bettolo, 3 - 00195 Roma
t + 39 06 37 29 945
f + 39 06 37 51 83 28

Cosenza

Via Sabotino, 55 - 87100 Cosenza
Tel +39 0984 21 84 2

www.idtech.it