



Open Archives Initiative e Metadata harvesting nel contesto delle Biblioteche Digitali

*Università di Parma
Corso Editoria Digitale*

30 Novembre 2006

*Dr. Valdo Pasqui
Università di Firenze
valdo.pasqui@unifi.it*

Contenuti

- **Parte I**
 - Contesto
 - Modello
 - Servizi e applicazioni
- **Parte II**
 - OAI metadati gestionali e gestione dei diritti
 - Uso del protocollo per veicolare risorse
- **Parte III**
 - Il Protocollo OAI-PMH in sintesi

Parte I

- Contesto
- Modello
- Applicazioni e Servizi

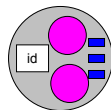
Contesto

- Inizio: Ottobre 1999, Santa Fe
- Trae origine nell'ambito della comunicazione accademica (scholarly) per:
 - favorire l'accesso agli archivi e-print
 - sviluppare e promuovere standard di interoperabilità allo scopo di facilitare la disseminazione efficiente dei contenuti
- Si è esteso oltre questi limiti perché è indipendente dai contenuti e dai modelli organizzativi ed economici circostanti

Contesto: Esempi “storici” di Open-Archives

- **ArXiv** del Los Alamos National Laboratory (LANL)
- **Networked Computer Science Technical Science Library (NCSTLR)** (architettura DIENST, CNR Digital Library)
- **CERN Document Server (CDS)**
- **Cognitive Sciences Eprint Archive (CogPrints)** University of Southampton (=> Eprints.org)

Contesto: terminologia



Digital Object: A data structure whose principal components are digital data and key-metadata.

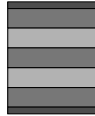
Digital data can be a *Datastream* or a *Digital Object*, i.e. a *Digital Object* may have one or more other *Digital Objects* as nested components.

Key-metadata must include an identifier for the *Digital Object*.



Datastream: An ordered sequence of bytes.

Contesto: terminologia



Repository: a networked system that provides services pertaining to a collection of *Digital Objects*.

"Currently, we do not really have a shared definition of a **digital repository**[.....]
The best thing we can say is that if it doesn't support the Open Archive Initiative Protocol for Metadata Harvesting it is probably not a very well-behaved digital repository, assuming it is a repository" (Clyff Linch, CNI)

Contesto: terminologia



Data Model:

An abstraction for *Digital Objects* such that each *Digital Object* can be seen as an instance of the class defined by a *Data Model*.

Example *Data Models* include the Pathways Core model, the MPEG-21 Digital Item Declaration model, METS, etc.

Definizioni tratte da: "Augmenting Interoperability across Scholarly Repositories", April 20-21, 2006, New York. URL: <http://msc.melon.org/Meetings/Interop>

Contesto: diversi servizi

Anna Maria **Tammaro**

International Conference on Scholarly Communication and **Academic Presses**

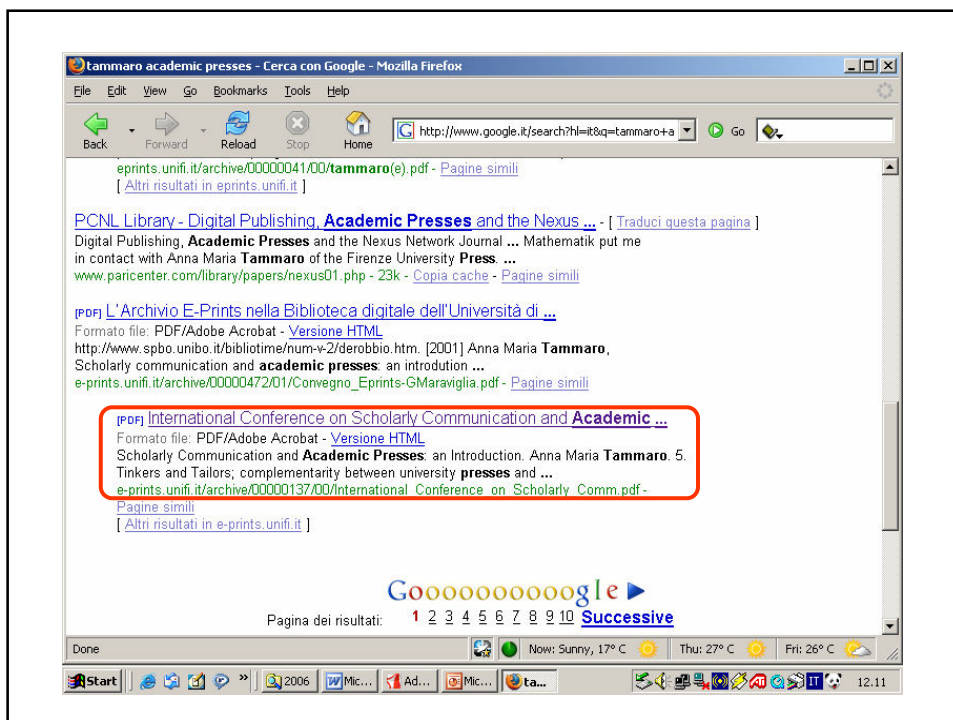
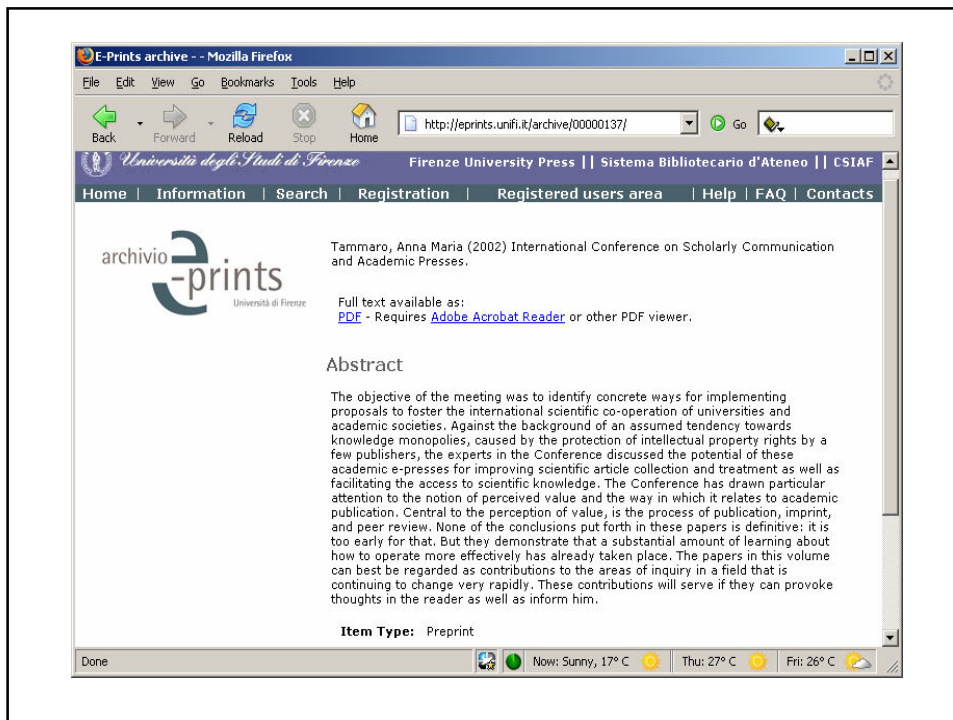
- **Archivio e-prints Università di Firenze**
<http://eprints.unifi.it/perl/search/simple>
- **Google**
<http://www.google.it/search?q=tammaro+academic+presses&hl=it&lr=&start=20&sa=N>
- **OAISTER** <http://oaister.umdl.umich.edu/o/oaister/>
- **PLEIADI** <http://www.openarchives.it/pleiadi/>
- **OPAC SBN** <http://www.internetculturale.it/>

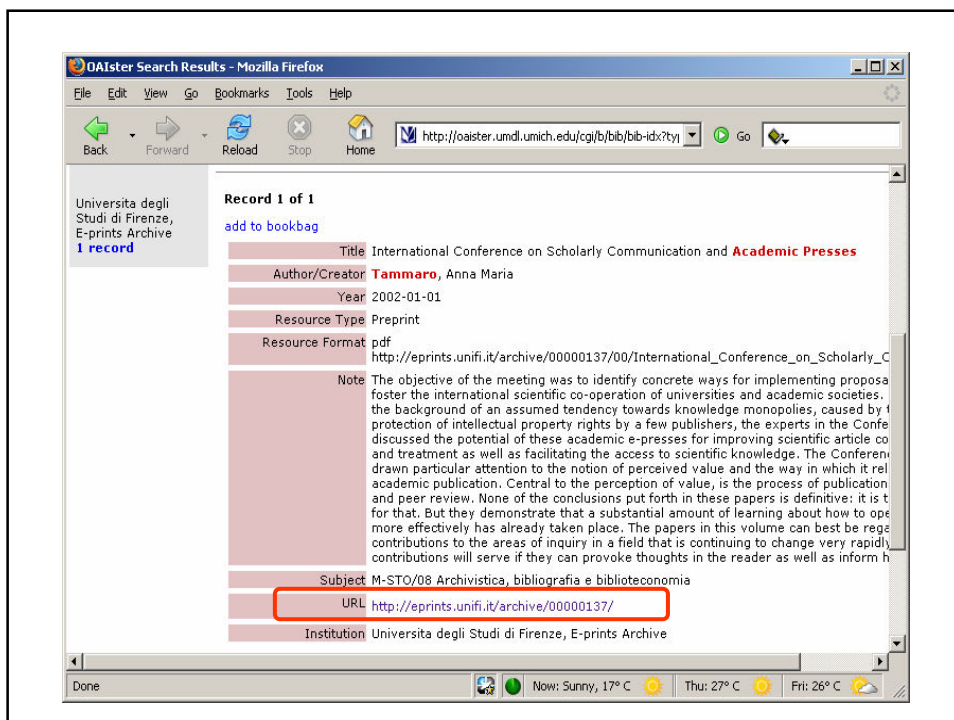
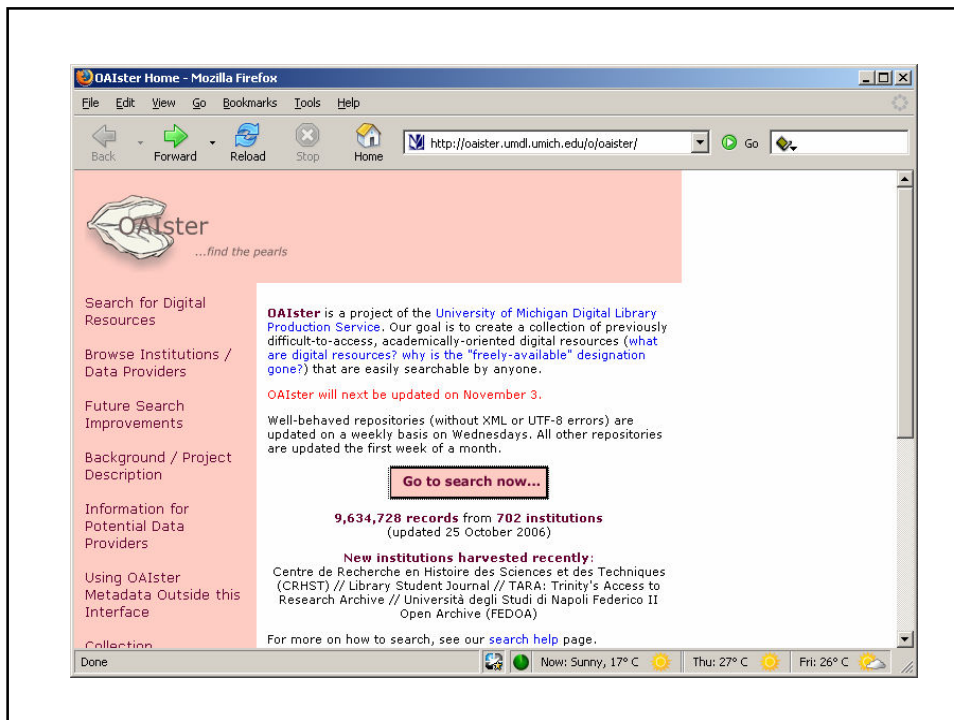
Parma 30/11/2006

Valdo Pasqui Università di Firenze

9

The screenshot shows a Mozilla Firefox browser window titled "E-Prints archive - Results for Simple Search". The address bar contains the URL "http://eprints.unifi.it/perl/search/simple?abstract?". The page header includes the University of Florence logo and navigation links: Home, Information, Search, Registration, Registered users area, Help, FAQ, and Contacts. The main content area displays the "archivio e-prints" logo and search results for the query "tammaro". The results show a single entry: "Tammaro, Anna Maria (2002) International Conference on Scholarly Communication and Academic Presses". The search interface includes "Refine search" and "New search" buttons.





PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali - Archivi Aperti ed A...

File Edit View Go Bookmarks Tools Help PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi I

PLEIADI

PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali

Login
 Nome utente:
 Password:
 Login utente
 Hai perso la password?

Registrati ora!
 Menu Principale
 Home
 Ultime notizie
 RSS: Notizie dal mondo OA
 Forum
 Ricerca negli archivi
 Risorse Web
 Progetto PLEIADI
 Chi siamo

Ricerca in OpenArchives
 tammaro academic presses
 Cerca
 Altre modalità di ricerca
 Record attualmente in archivio: 7729

Manifesto del progetto PLEIADI
 Il Progetto PLEIADI (Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali) scaturisce dalla collaborazione fra due importanti consorzi interuniversitari italiani, CASPUR e CILEA, nell'ambito del progetto AEPIC. [continua...]

Notizie recenti

- La Open Archives Initiative annuncia ORE - Object Reuse and Exchange (13/10/2006)
- Cosa può significare l'accesso aperto per gli studi umanistici (4/10/2006)
- SPARC and PKP annunciano una partnership per l'editoria Open Access (3/10/2006)
- OA15 a Ginevra - 18-20 aprile 2007 (28/9/2006)
- Fermo servizio: venerdì 22 settembre h.20 - sabato 23 settembre h.10 (CET) (21/9/2006)
- ELPUB 2007 - 1st CFP - Openness in Digital Publishing (4/9/2006)
- Uno scienziato etico che pubblica solo su Arxiv (2/9/2006)

Done Now: Sunny, 17° C Thu: 27° C Fri: 26° C

PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali - Ricerca negli archi...

File Edit View Go Bookmarks Tools Help PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi I

PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali

Login
 Nome utente:
 Password:
 Login utente
 Hai perso la password?

Registrati ora!
 Menu Principale
 Home
 Ultime notizie
 RSS: Notizie dal mondo OA
 Forum
 Ricerca negli archivi
 Risorse Web
 Progetto PLEIADI
 Chi siamo

Scegli la lingua
 Realizzato da

1 - 1 of 1 results

1 International Conference on Scholarly Communication and Academic Presses

Description

The objective of the meeting was to identify concrete ways for implementing proposals to foster the international scientific co-operation of universities and academic societies. Against the background of an assumed tendency towards knowledge monopolies, caused by the protection of intellectual property rights by a few publishers, the experts in the Conference discussed the potential of these academic e-presses for improving scientific article collection and treatment as well as facilitating the access to scientific knowledge.

The Conference has drawn particular attention to the notion of perceived value and the way in which it relates to academic publication. Central to the perception of value, is the process of publication, imprint, and peer review.

None of the conclusions put forth in these papers is definitive: it is too early for that. But they demonstrate that a substantial amount of learning about how to operate more effectively has already taken place. The papers in this volume can best be regarded as contributions to the areas of inquiry in a field that is continuing to change very rapidly. These contributions will serve if they can provoke thoughts in the reader as well as inform him.

Authors Tammaro, Anna Maria
 Subjects MST008
 Archive FI_uni
 Link <http://e-prints.unifi.it/archive/00000137/>

Done Now: Sunny, 17° C Thu: 27° C Fri: 26° C

Ricerca Bibliografica - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Back Forward Reload Stop Home

http://www.internetculturale.it/moduli/opac/opac. Go

Ministero per i Beni e le Attività Culturali • Dip. per i Beni Archivistici e Librari •
Dir. Gen. per i Beni Librari e gli Istituti Culturali • ICCU

INTERNET Culturale

Ricerca bibliografica

ENGLISH FRANÇAIS ESPAÑOL

Trovati : 2

Ordina per: [Salva ricerca](#) [Raffina ricerca](#) [Nuova ricerca](#)

Risultati : 1 : 2

[Visualizza selezionati](#) [Visualizza tutti](#) [Annulla selezione](#) [Stampa](#)

- [M] Testo a stampa - Scholarly communication and academic presses : proceedings of the International conference : 22 March 2001, Univer - Firenze - W0 1-2002
- [M] Archivio elettronico - Scholarly communication and academic presses : proceedings of the International conference : 22 March 2001, Univer

[Visualizza selezionati](#) [Visualizza tutti](#) [Annulla selezione](#) [Stampa](#)

© 2005 Copyright ICCU

Done

Now: Sunny, 17° C Thu: 27° C Fri: 26° C

Ricerca Bibliografica - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Back Forward Reload Stop Home

http://www.internetculturale.it/moduli/opac/opac. Go

Ministero per i Beni e le Attività Culturali • Dip. per i Beni Archivistici e Librari •
Dir. Gen. per i Beni Librari e gli Istituti Culturali • ICCU

INTERNET Culturale

Ricerca bibliografica

ENGLISH FRANÇAIS ESPAÑOL

Scheda: 1 [Aggiungi al segnalibri](#) [Esporta](#)

Livello bibliografico: Monografia
Tipo documento: Testo a stampa
Titolo: Scholarly communication and academic presses : proceedings of the International conference : 22 March 2001, University of Florence, Italy / edited by Anna Maria Tammaro

Publicazione: Firenze : Firenze university press, c2002

Descrizione fisica: 114 p., 24 cm

Note Generali: In testa al front.: Università degli studi di Firenze

Numeri: ISBN - 88-8453-049-0
Bibliografia Nazionale - 2003-5091

Nomi: Tammaro, Anna Maria
Università degli studi <Firenze>

Soggetti: Pubblicazioni elettroniche
Pubblicazioni elettroniche - Congressi - 2001
Pubblicazioni scientifiche - Diffusione

Classificazione: 070.5 - EDITORIA
070.579 - EDITORIA. TIPI SPECIALI DI PUBBLICAZIONE

Paese di pubblicazione: ITALIA
Lingua di pubblicazione: inglese
Codice del documento: IT\ICCU\CFI\0548139
Localizzazioni: Tutte | Prestito | OPAC locale | Anagrafe Biblioteche |

Scheda: 2 [Aggiungi al segnalibri](#) [Esporta](#)

Done

Now: Sunny, 17° C Thu: 27° C Fri: 26° C

Ricerca Bibliografica - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

Back Forward Reload Stop Home

http://www.internetculturale.it/moduli/opac/opac. Go

Ministero per i Beni e le Attività Culturali • Dip. per i Beni Archivistici e Librari • Dir. Gen. per i Beni Librari e gli Istituti Culturali • ICCU

INTERNET Culturale Ricerca bibliografica

ENGLISH FRANÇAIS ESPAÑOL

Lingua di pubblicazione: inglese
Codice del documento: IT\ICCU\CFI\0548139
Localizzazioni: Tutte | Prestito | OPAC locale | Anagrafe Biblioteche |

Scheda: 2 **Aggiungi al segnalibri** Esporta

Livello bibliografico: Monografia
Tipo documento: Archivio elettronico
Titolo: Scholarly communication and academic presses : proceedings of the International conference : 22 March 2001, University of Florence, Italy / edited by Anna Maria Tammaro
Descrizione fisica: Firenze : Firenze university press, 2002
Note Generali: Modalità di accesso: WWW. URL: <http://digital.casalini.it/fulltext/is.asp?isbn=8884530490>
Requisiti tecnici e modalità di accesso sul sito del distributore
Titolo del front.
Numeri: ISBN - 88-8453-049-0
Nomi: Tammaro, Anna Maria
Soggetti: Editoria
Classificazione: 070 - GIORNALISMO, EDITORIA, GIORNALI
Paese di pubblicazione: ITALIA
Lingua di pubblicazione: inglese
Codice del documento: IT\ICCU\UFI\0426662

Visualizza in formato: Formato etichette Vai

Invia E-Mail Stampa Lista dei Risultati

Done

Now: Sunny, 17° C Thu: 27° C Fri: 26° C

OAI <http://www.openarchives.org/>

- **Modello Funzionale**
 - ◆ Data Provider
 - ◆ Service Provider
- **Modello Tecnico**
 - ◆ harvesting di metadati
- **Modello dei dati**
 - ◆ resource-item-metadata-record

Modello Funzionale: Data Provider e Repository

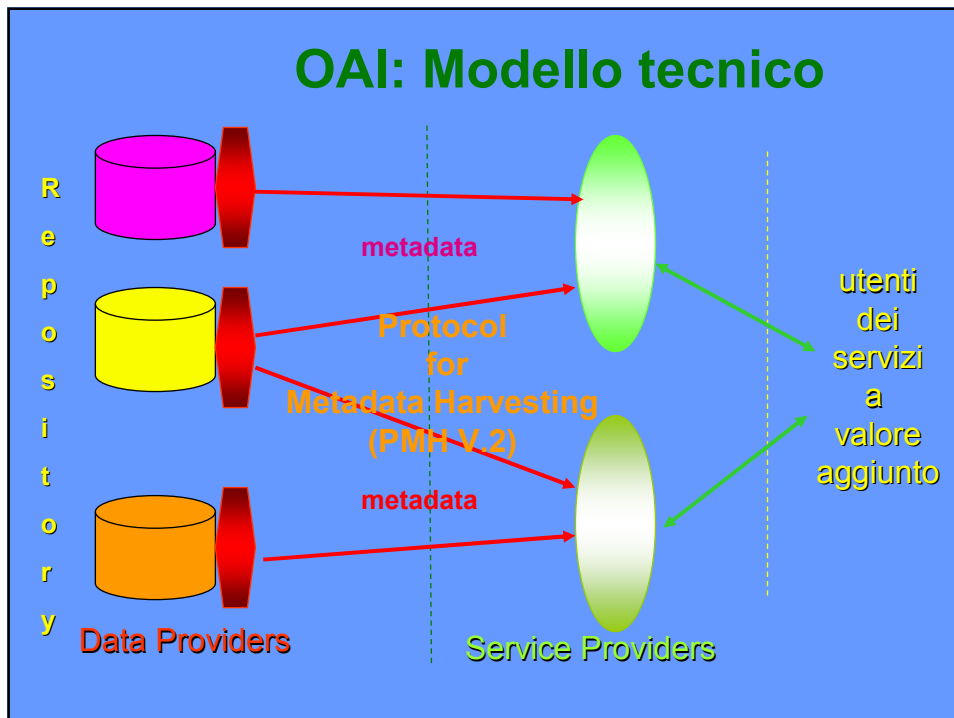
- Un **Data Provider** :
 - ◆ **gestisce uno più repository (archivi)/ collezioni di oggetti digitali**
 - ◆ è responsabile del mantenimento di questi oggetti e della **generazione dei metadati** che li caratterizzano
 - ◆ **supporta il protocollo OAI** per consentire l'accesso ai metadati sul suo contenuto
 - ◆ va inteso in senso esteso: **non solo e-print ma il più ampio spettro delle collezioni di oggetti digitali**

Modello Funzionale: Service Provider

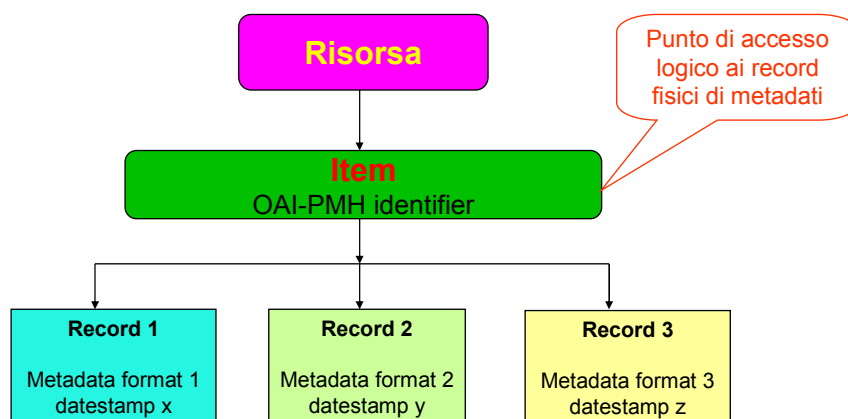
- I **Service Provider** :
 - ◆ **gestiscono i servizi a valore aggiunto** per la aggregazione ed indicizzazione dei metadati (ricerca, scoperta, localizzazione degli oggetti digitali)
 - ◆ **forniscono interfacce utente che si avvalgono della tecnologia dei portali e dei middleware (OAI-PMH, OpenURL, WebServices, Z39.50, ISO ILL, NCIP)**
 - ◆ **interrogano i repository dei data provider** usando le richieste del protocollo OAI
 - ◆ **catturano i metadati esposti dai data provider**

Modello Tecnico

- I service provider interrogano i data provider da cui prelevano i metadati con il protocollo applicativo OAI-PMH (Protocol for Metadata Harvesting)
- I messaggi del protocollo (richieste e risposte) vengono attraverso il protocollo HTTP
- Le risposte sono documenti XML



Modello dei dati: "resource-item-metadata-record"



Parma 30/11/2006

Valdo Pasqui Università di Firenze

25

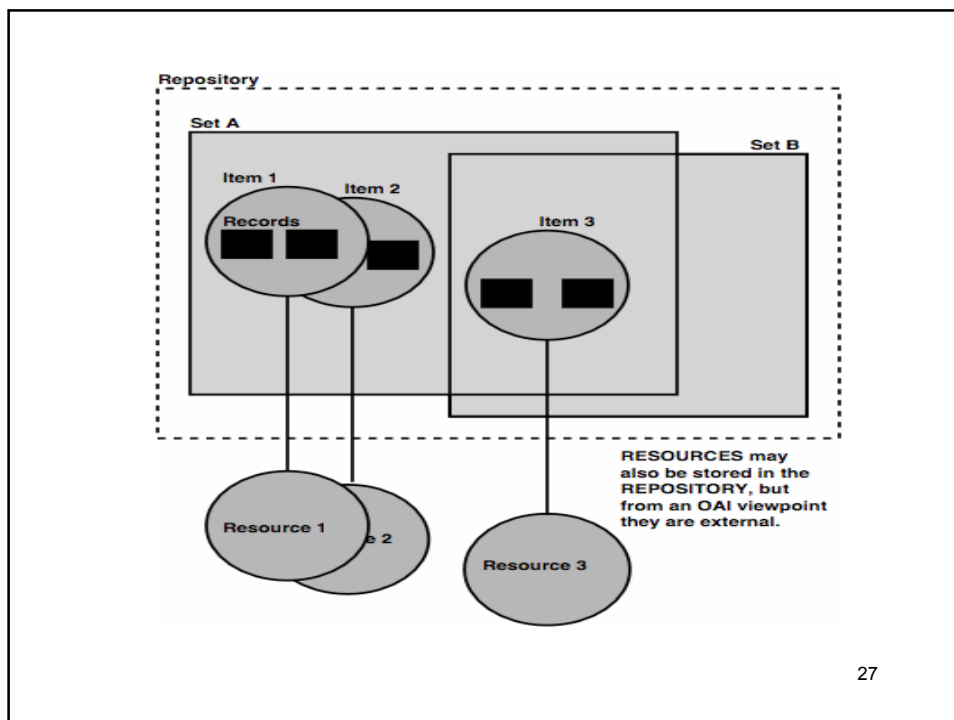
Modello dei dati: "resource-item-metadata-record"

- le **risorse** sono gli oggetti, contenuti nei digital repository mantenuti dai Data Provider
- a questi oggetti sono associati i **metadati**, esposti dai Data Provider tramite il protocollo OAI-PMH
- il punto di accesso logico ai record fisici di metadati costituisce un **item**
- i **record** sono i metadati espressi in sintassi XML secondo lo schema Dublin Core non qualificato tuttavia con OAI-PMH possono essere catturati anche metadati definiti in specifici domini di applicazione (vocabolari identificati attraverso i relativi namespaces).

Parma 30/11/2006

Valdo Pasqui Università di Firenze

26



27

Item e Record

- **Item:** “an item is now defined as a constituent of a repository from which metadata about a resource can be disseminated”
- **Record:**
 - formato XML
 - hanno un identificatore (sintassi URI) composto da `oai:identificatore_archivio:identificatore_record`

```

<record>
  <header>
    <identifier>oai:eg:001</identifier>
    <timestamp>2001-03-30</timestamp>
  </header>
  <metadata>
    <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/">
      <creator>.....</creator>
      <title>.....</title>
      <identifier>http://.....</identifier>
    </oai_dc:dc>
  </metadata>
  <about>
    <ea xmlns="http://www.arXiv.org/ea">
      <usage>No restrictions</usage>
    </ea>
  </about>
</record>

```

elementi specificati dal protocollo
 Metadati descrittivi almeno Dublin Core
 opzionale, non specificato: Contenuto = metadati definiti nell'ambito di specifiche comunità

OAI-PMH in sintesi

- **Risorsa** = oggetto cui sono associati i metadati
- **Item** = contenitore che genera/memorizza i metadati (ha un identificatore unico)
- **Record** = metadati espressi in sintassi XML (almeno Dublin Core)
- **Tipi di richieste:**
 - GetRecord
 - ListIdentifiers
 - ListRecords
 - Identify
 - ListMetadataFormats
 - ListSets
- **Harvesting selettivo:**
 - Datestamps
 - Sets(=>collezioni)

OAI-PMH in azione

- Caratteristiche di un archivio: **Identify**

<http://e-prints.unifi.it/perl/oai2?verb=Identify>

<http://dspace.unifi.it:8080/dspace-oai/request?verb=Identify>

- Cattura di un record: **GetRecord**

http://e-prints.unifi.it/perl/oai2?verb=GetRecord&identifier=oai:GenericEPrints.OAI2:00000137&metadataPrefix=oai_dc

Rilevanza del modello OAI

- ◆ Assicura la **qualità dei metadati diffusi**
- ◆ **Promuove lo sviluppo di servizi per l'aggregazione e per l'accesso** alle risorse di archivi aperti, ma può essere usato anche quando l'accesso alle risorse deve essere controllato
- ◆ **Modello per l'architettura della biblioteca digitale:** integrazione delle risorse nel contesto della biblioteca ibrida, sviluppo di portali tematici e istituzionali (con reference linking e citazioni)
- ◆ Favorisce lo **sviluppo di strumenti software open source**

OAI & PMH: Qualità dei dati

- Motori di ricerca WWW (es. Google)
 - indicizzazione pagine siti Web
 - rumore :: accuratezza, completezza
- OAI Protocol for Metadata Harvesting
 - metadati qualitativamente controllati dalle istituzioni che li pubblicano (espressi in sintassi XML) almeno secondo lo standard Dublin Core
 - risorse selezionate (i Data Repository)

OAI: software open source (1)

- Software per la creazione di archivi e-prints e digitali
 - eprints.org
 - DSpace
 - Fedora project
 - CDSware
 - aDORe
- Open Digital Libraries, Virginia Tech, <http://oai.dlib.vt.edu/odl>, estende OAI-PMH per creare una rete di OA interoperanti

OAI: software open source (2)

- **OAIHarvester2** (OCLC)
<http://www.oclc.org/research/software/oai/harvester2.htm>
- **Public Knowledge Project (PKP) Harvester**
<http://pkp.sfu.ca/pkp-harvester/>
- **ARC** harvester and search engine
<http://sourceforge.net/projects/oaiarc/>
- **OAIcat** (OCLC) metadata server che può essere interfacciato a repository esistenti
<http://www.oclc.org/research/software/oai/cat.htm>
- **DP9** (Old Dominion University) espone i metadati sotto forma di pagine HTML utilizzabili da web search-engines
<http://arc.cs.odu.edu:8080/dp9/index.jsp>

OAI- PMH: servizi

- ◆ **NCSTRL** Old Dominion University, University of Virginia and Virginia Tech; NASA
- ◆ **OAlster** University of Michigan Libraries Digital Library Production Service
- ◆ **SAIL-Eprints** CNR - Area della Ricerca di Bologna
- ◆ **citebaseSearch** Southampton University **Open Citation Project** (ArXiv, cogprints, Biomed Central)
- ◆ **Pleiadi** progetto CILEA e CASPUR
<http://www.openarchives.it/pleiadi/>
- ◆ **CYCLADES** European Research Consortium for Informatics and Mathematics (ERCIM)
- ◆ **Open Archives Forum** <http://www.oaforum.org/>

Registri di servizi OAI

- Open Archives registered Data Providers
<http://www.openarchives.org/Register/BrowseSites>
- Celestial <http://celestial.eprints.org/>
- Registry of Open Access Repositories ROAR
<http://archives.eprints.org/>
- OAI registry University of Illinois at Urbana-Champaign <http://gita.grainger.uiuc.edu/registry/Info.asp>

Parte II

- Il protocollo OAI per “veicolare” le risorse e oggetti complessi
- Confronto con altri modelli
- Gestione dei diritti

Harvesting delle risorse (1)

- Finalità

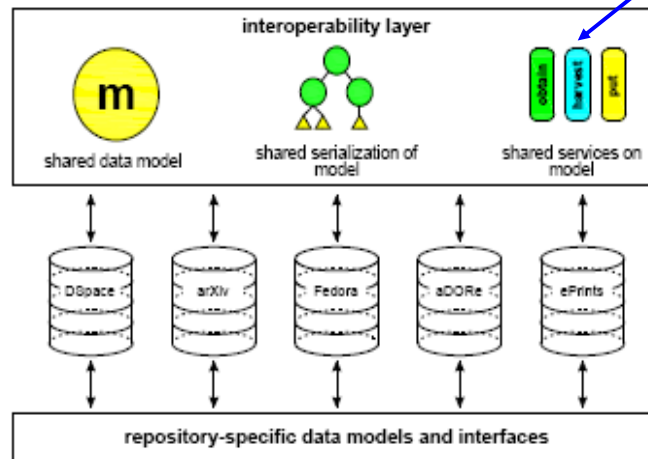
- **Preservazione**: “trusted digital repositories” da sincronizzare
- **Scoperta**: servizi di ricerca basate sui contenuti
 - Search engines che operano sul full text
 - Citation indexing services che estraggono le citazioni dal full text
 - Servizi esterni di browse sui thumbnails estratti da collezioni di immagini di elevata qualità

Harvesting delle risorse (2)

- **Interoperabilità**: tra repository accademici
 - Scambio di informazioni sui propri oggetti digitali
 - **Object Reuse and Exchange (ORE)** progetto di due anni lanciato nell'Ottobre 2006 nell'ambito OAI <http://www.openarchives.org/ore/>
 - Nuova generazione di “cross-repository services”
 - **obtain, harvesting e put**
 - Harvest basato su OAI-PMH per catturare lotti di oggetti digitali serializzati (XML)
 - Van de Sompel et al. prototipo di **overlay journal** <http://dlib.org/dlib/october06/vandesompe/10vandesompe.html>

Pathways project...

<http://www.infosci.cornell.edu/pathways/>



Parma 30/11/2006

Valdo Pasqui Università di Firenze

41

OAI-PMH e Complex Objects Formats....

- **Complex Objects Formats**
 - **MPEG-21 DIDL** Information Technology, Multimedia Framework
 - **METS** (Metadata Encoding & Transmission Standard) <http://www.loc.gov/standards/mets/>
 - **SCORM** (The Sharable Content Object Reference Model) in ambito distributed e-learning
 - **Pathways Core data model** (Van de Sompel et al)
 - **MAG** (Metadati Amministrativi Gestionali) <http://www.iccu.sbn.it/standardBD.html>

Parma 30/11/2006

Valdo Pasqui Università di Firenze

42

OAI-PMH eComplex Objects Formats....

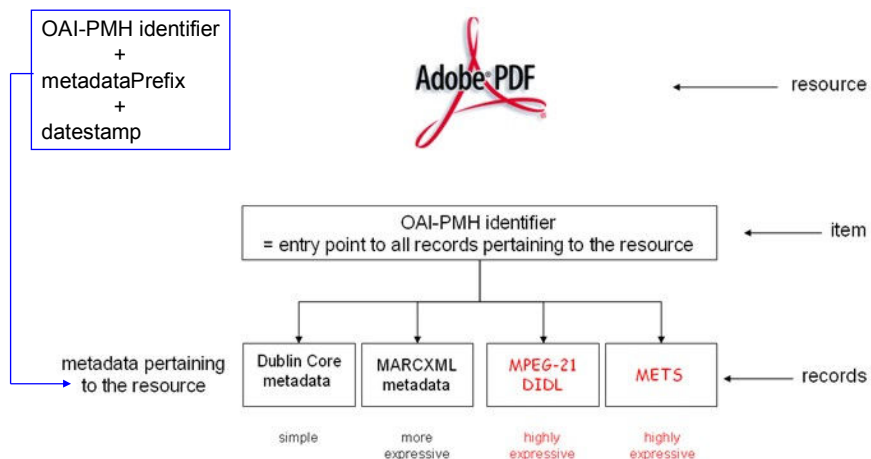
● Caratteristiche

- rappresentazione come documento XML
- possibilità di rappresentare oggetti semplici o complessi (i.e. 1 o più datastream)
- i datastream possono essere inclusi nel doc XML (come sequenze codificate in base-64) o vi può comparire solo il riferimento alla locazione di rete
- coesistenza dei metadati descrittivi, tecnici, di struttura e relativi alla gestione dei diritti
- distinzione senza ambiguità tra l'identificatore del "data object" e quelli dei datastream che lo compongono

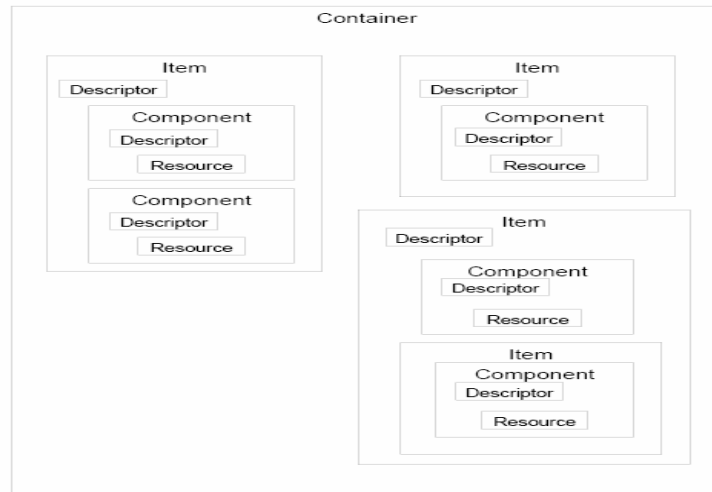
Resource Harvesting within the OAI-PMH Framework

Herbert Van de Sompel, Michael L. Nelson, Carl Lagoze, and Simeon Warner

<http://www.dlib.org/dlib/december04/vandesompel/12vandesompel.html>



Il modello MPEG-21



1

15

OAI e gli altri modelli

Confronto con altri modelli:

- ◆ Search engines
- ◆ Query distribuite (metacataloghi)
- ◆ OpenURL

Harvesting delle search engines

- ◆ Diverso dall'harvesting dei motori di ricerca del Web Google, AltaVista, HotBot etc. che "visitano" periodicamente i siti Web e ne indicizzano il contenuto (pagine HTML)
- ◆ L'harvesting di OAI opera su:
 - ◆ risorse selezionate
 - ◆ su metadati qualitativamente controllati dalle istituzioni che li pubblicano

OAI-PMH e Google

- Google prevede la definizione di **Sitemaps** per configurare l'accesso ai siti Web
<http://www.google.com/webmasters/sitemaps/docs/it/other.html>
- È possibile registrare come Sitemap l'URL base di accesso ad un repository OAI
- Quando Google interroga l'archivio aggiunge automaticamente *?verb=Identify* o *?verb=ListRecords*
- Nella risposta si aspetta metadati DC e che dc:identifier contenga l'URL della risorsa

Guida di Google Sitemaps (BETA) - Mozilla Firefox

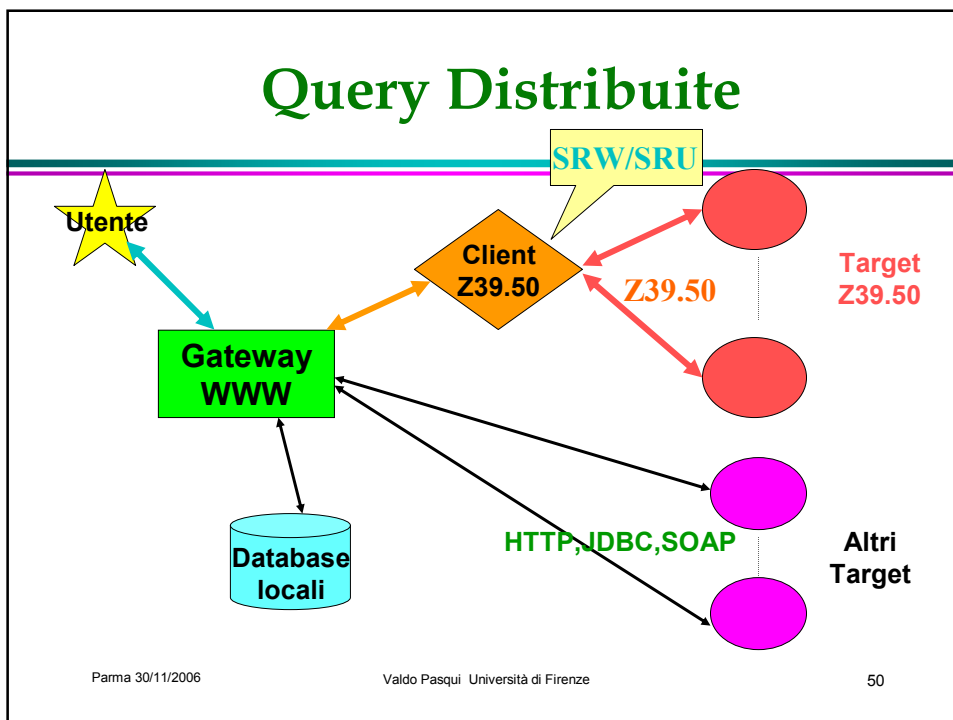
http://www.google.com/webmasters/sitemaps/docs/it/other.html#oai

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)

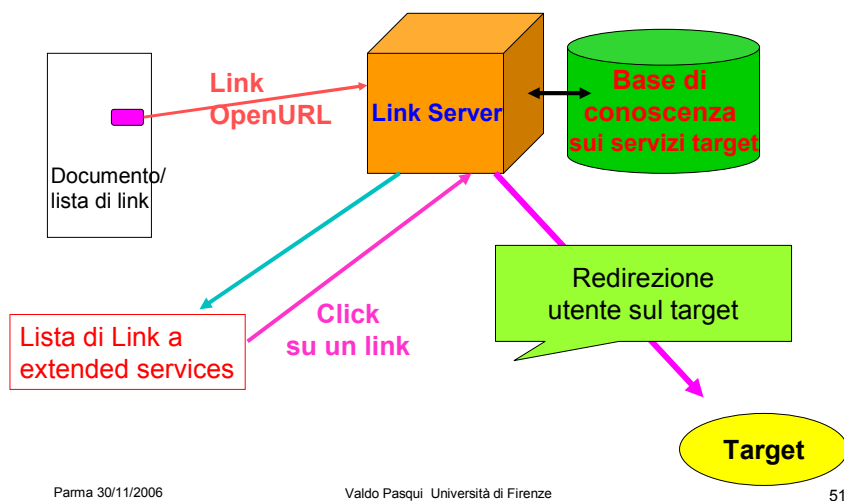
Google accetta il protocollo [OAI-PMH](#) versione 2.0. Non è possibile utilizzare questo formato per le [Sitemap mobili](#). Se utilizzi questo formato per il tuo sito, è sufficiente [aggiungere](#) il parametro baseURL del repository OAI (ad esempio <http://www.example.com/oaiserver>). Quando effettuiamo una query sul parametro baseURL, aggiungiamo automaticamente parametri di query (quali `?verb=Identify` o `?verb=ListRecords`), pertanto puoi limitarti ad aggiungere semplicemente il parametro baseURL. Quando procediamo all'estrazione degli URL per il tuo sito, presumiamo che i record nel repository siano formattati utilizzando lo standard [Dublin Core](#) e gli URL siano incorporati nei tag `<dc:identifier>`. Viene illustrato di seguito un record di esempio che include il tag `<dc:identifier>` in grassetto. L'URL indicato nel tag è la parte che estraiamo.

```
<oai_dc:dc
  xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  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 xml:lang="en">A title of extraordinary things</dc:title>
  <dc:creator>McCormack, Michael</dc:creator>
  <dc:subject>LCSH:Ausdehnungslehre; LCCN Q&A205.H99; Greatness:Amanda</dc:subject>
  <dc:publisher>J. Wiley & Sons</dc:publisher>
  <dc:date>Created: 1906; Available: 1991</dc:date>
  <dc:type>text</dc:type>
  <dc:identifier>http://example.com/physics/1796949</dc:identifier>
```

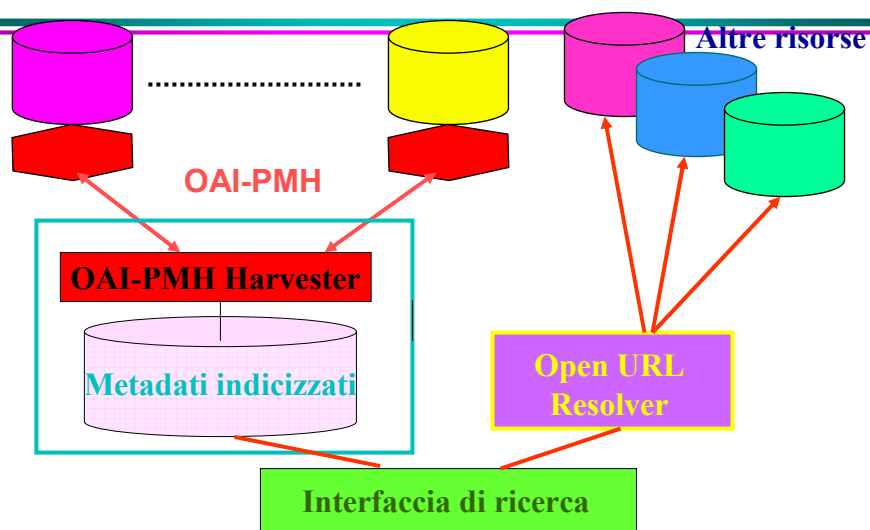
Done



Reference Linking



OAI + Reference Linking



Strumenti e Metadati

- **Cross-seaching**
 - interrogazioni distribuite sui metadati che descrivono gli oggetti digitali
- **Metadata Harvesting**
 - cattura, aggregazione e indicizzazione dei metadati associati agli oggetti digitali
- **Reference linking “context sensitive”**
 - i metadati trasportati sono usati dal resolver per creare dinamicamente riferimenti sensibili al contesto

Gestione dei Diritti

- **OAI non definisce né prescrive alcun schema di gestione dei diritti**
- La definizione (e pubblicazione) di restrizioni all'accesso e di vincoli relativi alla gestione della proprietà intellettuale nei metadati esposti è lasciata ai data provider
- **Il protocollo consente lo scambio di metadati gestionali in vari modi:**

◆ **description** nella risposta alla richiesta **Identify**, può contenere informazioni al livello di Repository sul:

- **contenuto del repository**
- **politiche di accesso ai metadati**
- **politiche di accesso ai dati**

◆ **about** nella risposta alla richiesta **GetRecord**, può contenere informazioni al livello di Record su:

- **diritti di accesso ai metadati del record**
- **termini e condizioni d'uso dei metadati del record**

◆ **metadata** nella risposta alle richieste **GetRecord** e **ListRecords**, può contenere:

- **formati di metadati che prevedono informazioni di tipo gestionale sui diritti** (es. elemento "rights" in Dublin Core) e di altro tipo

55

OAI-rights...

- Iniziativa lanciata nel Settembre 2003
- Le linee guida erano attese per il secondo quadrimestre del 2004
- Relazione con il Progetto RoMEO
- Obiettivo affrontare i problemi relativi:
 - **Diffusione dei metadati**
 - **Accesso alle risorse descritte**

...OAI-rights...

- **Espressione dei diritti associati con i metadati e con i contenuti (le risorse):**
 - ◆ Dichiarazioni sul copyright e restrizioni d'uso sui metadati perché sono il risultato di uno sforzo intellettuale considerevole
 - ◆ Restrizioni d'uso sull'accesso alle risorse per consentire ai Service Provider di fornire informazioni sull'accessibilità del contenuto degli archivi

...OAI-rights...

- **A quale livello associare i diritti ?**
 - ◆ Repository (tutti i Record e/o tutte le Risorse)
 - ◆ Item (tutti i Record raggruppati dall'item)
 - ◆ Set (tutti gli Item e quindi i Record aggregati dal set)
- **Come codificare queste espressioni ?**
 - ◆ Utilizzo dell'elemento "rights" di Dublin Core
 - ◆ Utilizzo dell'elemento "about" per contenere i diritti espressi in XML e conformi ad alcuni schemi come ODRL e XrML
 - ◆ Un metadata prefix "oai_rights" che prevede solo un elemento per includere i metadati espressi secondo i suddetti schemi

Schemi proposti

<http://www.openarchives.org/OAI/2.0/guidelines-rights.htm>

- uno schema XML per esprimere i diritti a livello di **record**, all'interno dell'elemento <about> con gli elementi <rights>, <rightsDefinition> e <rightsReference>
- uno schema XML per esprimere i diritti a livello di **repository** con l'elemento <rightsManifest> da includere in <description> nella risposta al comando Identity e per i diritti a livello di Set in <setDescription> in risposta al comando ListSets
- esempi basati su Creative Commons

Parte III

Il protocollo OAI-PMH

Protocollo di Harvesting V.2

- Item del repository
- Record restituiti dal repository
- Identificatori unici
- Metadati
- Richieste
- Trasporto e codifica



Item del repository

- "an item is now defined as a constituent of a repository from which metadata about a resource can be disseminated"
- concetto astratto e il protocollo non si occupa né della natura dell'item, né di quali metadati sono associati e se sia incluso il "full content"
- Ogni item può disseminare metadati in molteplici formati
- Gli item possono essere raggruppati per insiemi (es. per soggetto, istituzione etc.)

Record

- Quando un repository riceve una richiesta (del protocollo OAI) per ottenere i metadati di un item, restituisce una sequenza di byte codificata in XML che viene chiamata **record**
- I record OAI hanno tre parti:
 - ◆ **header**
 - ◆ **metadata**
 - ◆ **about**

Protocollo: Record

- ◆ header, composto da:
 - **identificatore unico**
 - **datestamp** (data di creazione, cancellazione o ultima modifica del record)
- ◆ metadata
 - una singola manifestazione(formato) dei metadati dell'item - **almeno Dublin Core deve essere supportato dal repository**
- ◆ about opzionale e non specificato
 - usato per contenere informazioni sui diritti e le condizioni d'uso dei metadati


```

<record>
  <header>
    <identifier>oai:eg:001</identifier>
    <timestamp>2001-03-30</timestamp>
  </header>
  <metadata>
    <oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/">
      <creator>.....</creator>
      <title>.....</title>
      <identifier>http://.....</identifier>
    </oai_dc:dc>
  </metadata>
  <about>
    <ea xmlns="http://www.arXiv.org/ea">
      <usage>No restrictions</usage>
    </ea>
  </about>
</record>

```

elementi specificati dal protocollo
 Metadati descrittivi almeno Dublin Core
 opzionale, non specificato: Contenuto = metadati definiti nell'ambito di specifiche comunità

Protocollo: Identificatori

- Sono le chiavi uniche che servono ad estrarre un record (i metadata di un item) da un repository
- Il formato segue la sintassi URI:

oai:identificatore_archivio:identificatore_record

es:

oai:arXiv:hep-th/9901001

oai:unifi:unifi00000010

Schema (fisso)

Registrato presso OAI

Univoco nel repository

Protocollo: Metadati

- Supporta **harvesting di molteplici formati di metadati a cui devono essere associati:**
- un **metadata schema** = uno schema XML di validazione pubblicamente accessibile tramite URL
 - ♦ <http://www.openarchives.org/OAI/dc.xsd>
 - ♦ <http://www.openarchives.org/OAI/rfc1807.xsd>
 - ♦ http://www.openarchives.org/OAI/oai_marc.xsd
- un **metadata prefix** = un nome che ogni repository mappa nello schema supportato e che è usato come argomento di alcune richieste
 - ♦ **oai_dc** è riservato per indicare lo schema Dublin Core senza qualificatori (obbligatorio)

Protocollo: Metadati

L'elemento radice dei metadati deve contenere:

- Una serie di attributi che cominciano con *xmlns* per stabilire il legame tra il metadata prefix e il namespace URI del relativo formato di metadati

```
xmlns:oai_dc=http://www.openarchives.org/OAI/2.0/oai_dc/  
xmlns:dc="http://purl.org/dc/elements/1.1/"
```

- Il riferimento al namespace URI dello schema XML

```
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

- L'attributo `xsi:schemaLocation` per specificare la corrispondenza tra il namespace URI dei metadati usati e l'URL dello schema XML che li valida

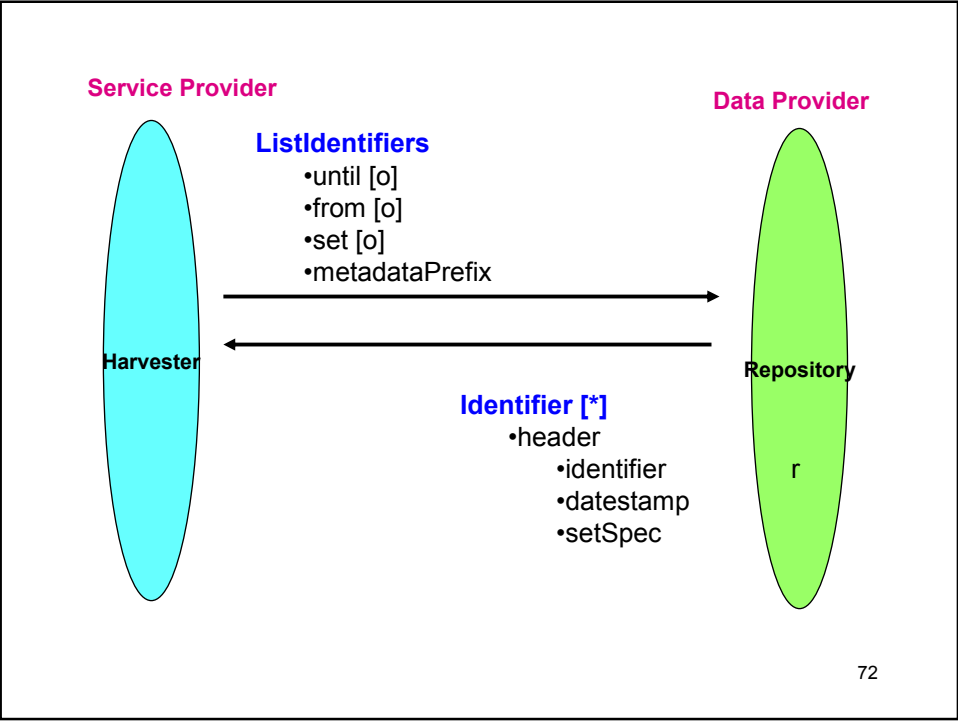
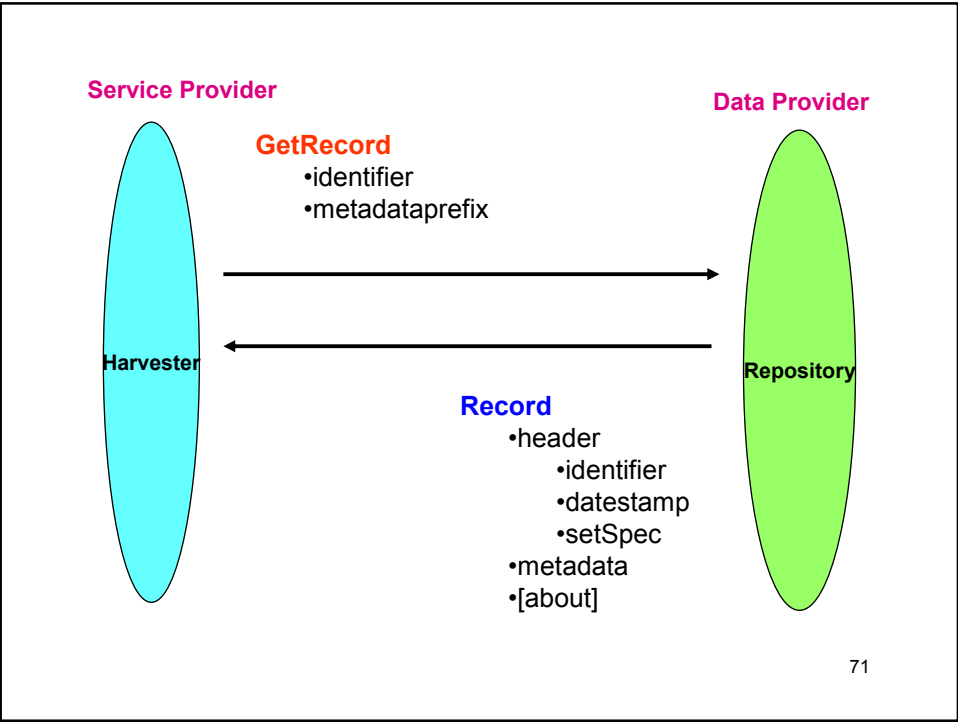
```
xsi:schemaLocation=  
  "http://www.openarchives.org/OAI/2.0/oai_dc/  
  http://www.openarchives.org/OAI/2.0/oai_dc.xsd"
```

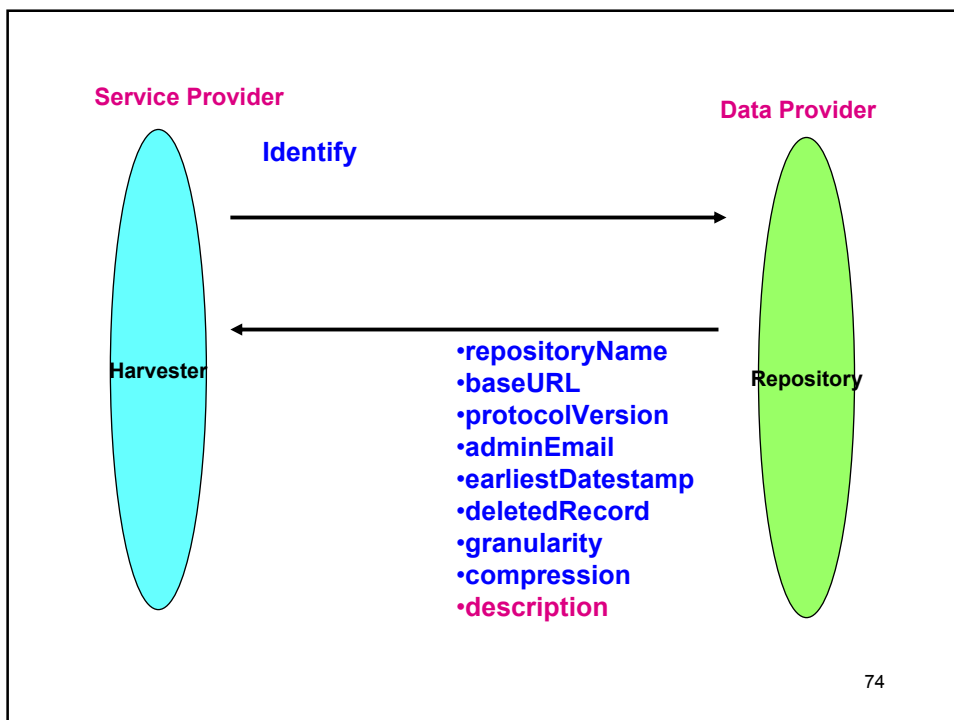
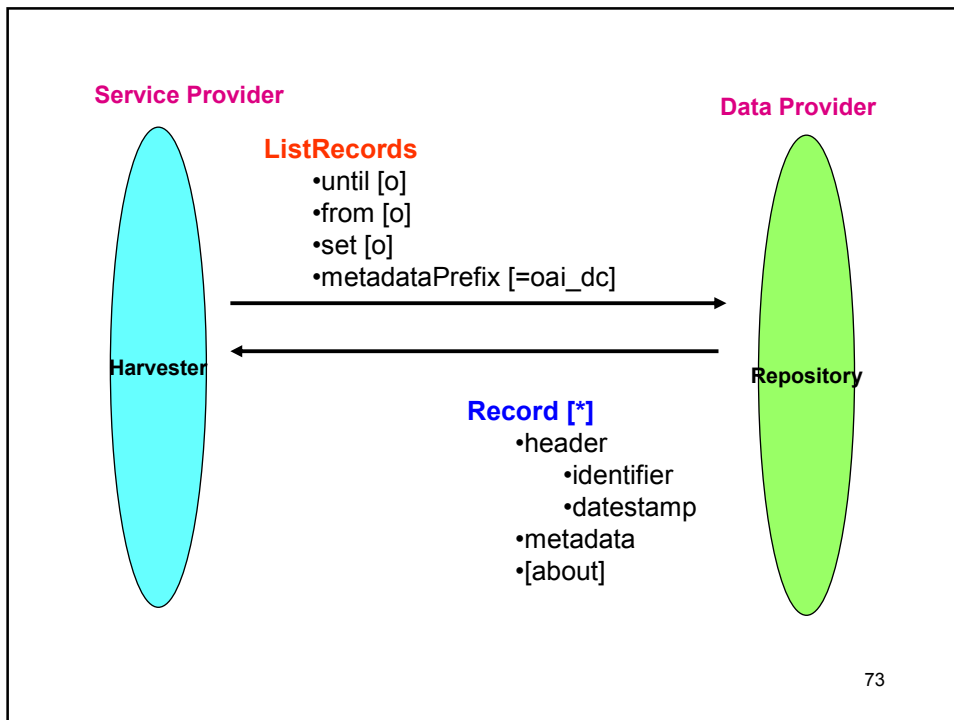
Protocollo: Metadati

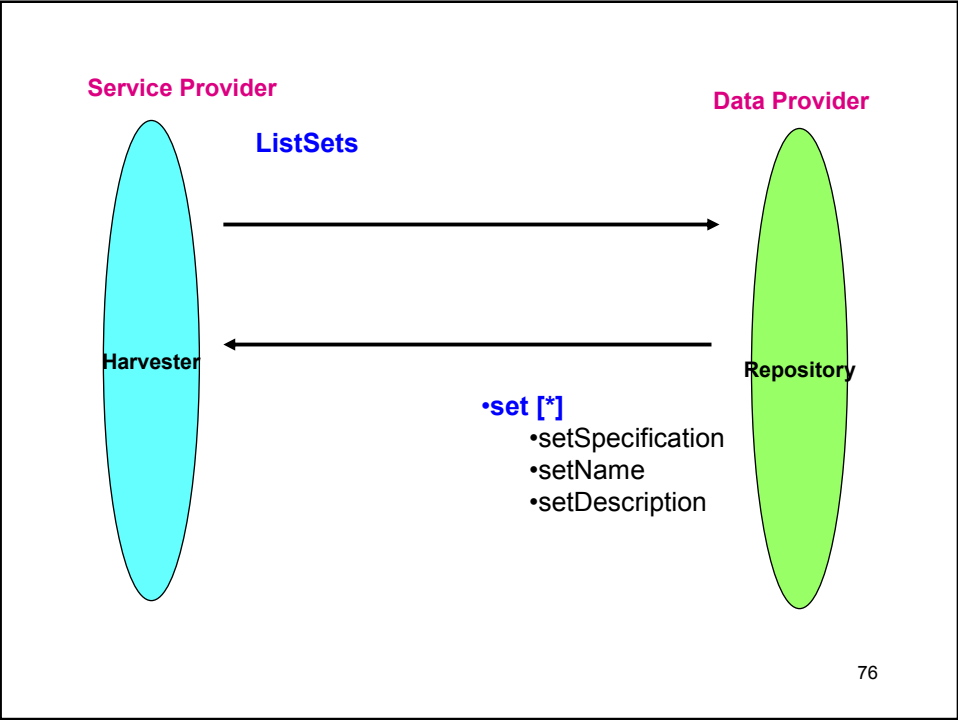
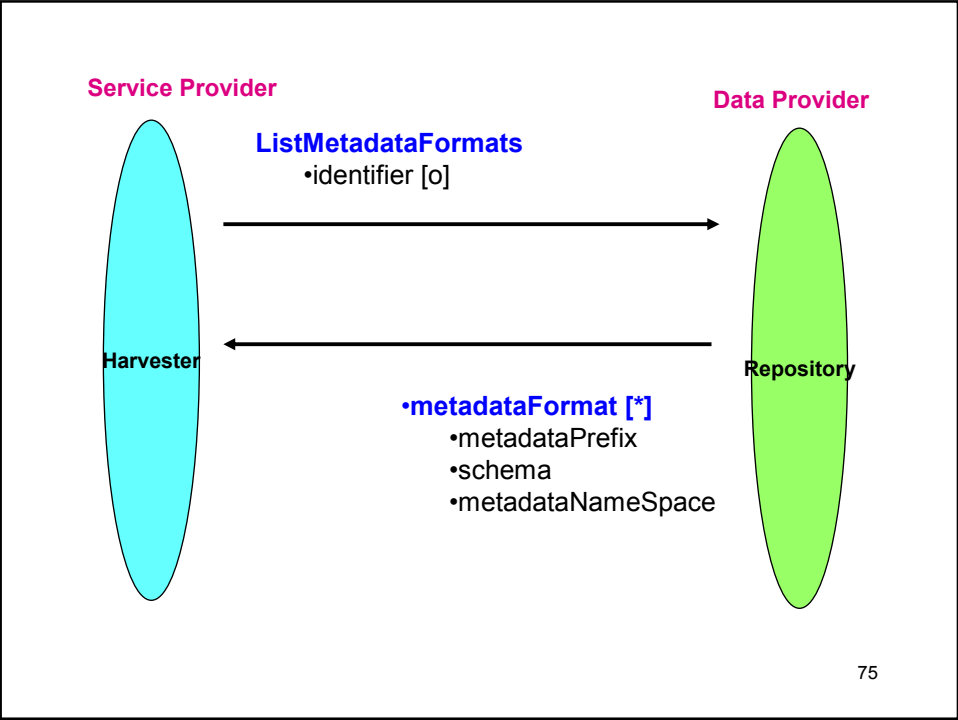
- ◆ È possibile estendere o includere altri set (schemi) di metadati.
- ◆ Altri schemi previsti sono rfc1807, marc21, oai_marc
- ◆ Esempi:
<http://www.oaforum.org/tutorial/english/page5.htm>

Protocollo: Richieste

- **GetRecord** : per ottenere uno specifico record (metadati) da un item del repository
- **ListIdentifiers**: per ottenere gli identificatori dei record che possono essere catturati dal repository
- **ListRecords**: per ottenere un insieme di record
- **Identify**: per ottenere informazione sul repository
- **ListMetadataFormats**: per ottenere i formati di metadati resi disponibili dal repository
- **ListSets**: per ottenere la struttura dei set (partizioni)







Implementazione del protocollo OAI-PMH

- **Richieste** espresse con i metodi Get o Post di HTTP
- **Risposte** HTTP con:
 - ◆ Content-Type:text/xml
 - ◆ Codifica con rappresentazione UTF8 di Unicode
 - ◆ il root element ha lo stesso nome della richiesta e fa riferimento al relativo XML schema definito da OAI