

## Verso la costituzione di un portale OAI per gli studi sulle Civiltà italiche e del Mediterraneo antico

CLAUDIO BARCHESI

Istituto di Studi sulle Civiltà Italiche e del Mediterraneo Antico  
Consiglio Nazionale delle Ricerche

### 1. Introduzione

L'Istituto di Studi sulle Civiltà Italiche e del Mediterraneo Antico del CNR (ISCIMA) è nato il 15 ottobre 2001 ed è operativo dal luglio 2002. Diretto da Francesco Roncalli, esso riunisce il personale, le esperienze e i progetti di due preesistenti Istituti di ricerca: l'Istituto per l'Archeologia Etrusco-Italica (IAEI) e l'Istituto per la Civiltà Fenicia e Punica "Sabatino Moscati" (IFP).

Tra le attività di ricerca dell'ISCIMA, lo sviluppo di metodi e di strumenti di lettura ed elaborazione informatizzata dei dati archeologici, filologici e storici rappresenta una consolidata tradizione, che ha portato alla pubblicazione della rivista «Archeologia e Calcolatori» (<http://soi.cnr.it/archcalc>) e alla definizione di un settore informatico sotto la responsabilità scientifica di Paola Moscati (cfr. BARCHESI C. 2004A).

Molto si discute oggi, nella comunità archeologica, sui metodi di codifica e catalogazione delle informazioni, e si assiste ad un significativo fiorire di schemi di codifica XML, di ontologie e modelli semantici, tanto potenti quanto complessi. L'obiettivo finale cui questa temperie normativa e descrittiva sembra mirare è una migliore condivisione dei dati, stante la sempre più difficile realtà del web, particolarmente critica per gli studi storico-archeologici. La crescita esponenziale delle pagine in rete causa una diluizione delle informazioni realmente significative nel rumore di fondo prodotto da cultori della materia, da studiosi improbabili, da agenzie di turismo.

Le soluzioni informatiche attualmente in corso di studio per fronteggiare questi problemi appaiono per la gran parte indirizzate verso paradigmi complessi che creano ragionevoli perplessità sulle loro future possibilità di affermazione. Una soluzione praticabile, senza considerare l'applicazione di architetture *peer to peer* (forse ancora un po' premature per il mondo della ricerca: comunque si vedano PAPADOPOULOS D.C., MAVRIKAS E.C. 2004 e CANTONE F. 2005), appare quella fornita dalla tecnologia usata negli Open Archives ove l'idea di delegare la creazione di metadati descrittivi delle risorse digitali agli stessi organi istituzionali che le producono, e di assemblare poi le informazioni in sistemi di interrogazione tematici e centralizzati la cui amministrazione scientifica sia delegata ad una authority riconosciuta, sembra oggi abbastanza praticabile e al centro di attente riflessioni (ABBATTISTA G. 2004; TAMMARO A.M., DE GREGORI T. 2004).

Questa soluzione permetterebbe di creare *repositories* istituzionali di strutture di ricerca decentrate interrogabili da un motore centralizzato. Per migliorare la circolazione delle informazioni scientifiche è necessario a nostro avviso affidarsi a schemi di metadati semplici, rendere facile la creazione e la diffusione in rete dei metadati, creare sistemi di *information retrieval* basati su questi schemi. Appare ormai un dato di fatto che la complessità, specie nei sistemi che svolgono funzioni sussidiarie alla ricerca applicata, può dare lustro ai modelli teorici, ma sancisce in modo sistematico il fallimento di ogni speranza d'applicazione pratica.

L'accessibilità alle risorse elettroniche, siano esse pre o postprint di articoli scientifici, cartografia, archivi fotografici e testuali codificati nelle più diverse semantiche, può essere

notevolmente accresciuta, non facendo affidamento su tecnologie proprietarie o complesse, ma su metadati standard e protocolli di gestione semplici e realmente condivisi.

Stante la necessità di esporre all'interno del sito web istituzionale dell'ISCIMA versioni elettroniche e abstract di risorse documentali che siano accessibili attraverso locali strumenti di interrogazione, ma anche connessi a meta-archivi interdisciplinari e indicizzabili per il web dai principali motori di ricerca, si è scelto di sperimentare l'implementazione del protocollo OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html>). Gli archivi aperti possono avere infatti un'importanza strategica per le risorse elettroniche sviluppate in seno alle discipline storiche ed archeologiche (cfr. FOULONNEAU M., DAWSON D. 2003). La nostra sperimentazione si è applicata al repository della rivista "Archeologia e Calcolatori".

## 2. La rivista "Archeologia e Calcolatori"

Il primo numero di "Archeologia e Calcolatori" ha visto la luce nel 1990: la rivista è giunta oggi al quindicesimo volume. L'idea di fondare una rivista dedicata alle applicazioni informatiche in archeologia nacque nell'Istituto per l'Archeologia Etrusco-Italica del CNR, che era già da alcuni anni impegnato nella sperimentazione delle nuove tecnologie computerizzate in diversi settori degli studi archeologici. D'intesa con il Dipartimento di Archeologia e Storia delle Arti dell'Università di Siena, l'Istituto dette così vita, con il sostegno di un Comitato Scientifico internazionale, alla rivista annuale "Archeologia e Calcolatori".

Gli obiettivi inizialmente individuati sono stati nel tempo perseguiti e ampliati, con esiti certamente inaspettati rispetto alle previsioni, legate a un incontro tra tecnologia e ricerca tradizionale che negli anni Ottanta appariva sporadico, frutto di iniziative isolate, fonte di diffidenza, ma anche di soluzioni innovative. Alcuni punti fermi hanno consentito alla rivista di stare al passo con i tempi (MOSCATI P. 2002B): il costante contatto con il panorama internazionale; la scelta del multilinguismo; il rapporto dialettico tra riflessione teorica e sperimentazione; il rispetto dell'equilibrio tra gli spazi dedicati agli aspetti tecnologici e a quelli archeologici; l'aggiornamento bibliografico, mediante la classificazione sistematica delle informazioni; l'apertura verso le innovazioni, attraverso sperimentazioni intese a sviluppare progetti di ricerca concreti e nuove riflessioni teoriche.

La pubblicazione di numeri monografici dedicati a tematiche di carattere metodologico (n. 5, 1994: *Choice, Representation and Structuring of Archaeological Information*; n. 9, 1998: *Methodological Trends and Future Perspectives in the Application of GIS in Archaeology*; n. 15, 2004: *New Frontiers of Archaeological Research: Languages, Communication, Information Technology*), insieme all'edizione di Atti di Convegni internazionali (n. 4, 1993: *Data and Image Processing in Classical Archaeology* - Ravello 1992; n. 7, 1996: *III International Symposium on Computing and Archaeology* - Roma 1995; n. 11, 2000: *I Workshop di Archeologia Computazionale* - Napoli, Firenze 1999; n. 13, 2002: *XIV Congrès International UISPP, Proceedings of Commission IV Symposia: Data Management and Mathematical Methods in Archaeology* - Liège 2001), hanno consentito di coronare l'attività intrapresa, conferendo alla rivista quel ruolo, auspicato fin dal primo Editoriale (CRISTOFANI, FRANCOVICH 1990), di polo per la raccolta e la presentazione di progetti e di coordinamento per il superamento della frammentarietà delle esperienze.

Nel corso di questa attività editoriale più che decennale è stato possibile individuare i settori della ricerca archeologica più direttamente coinvolti nell'uso degli strumenti informatici ed evidenziare le applicazioni più diffuse, delineandone lo sviluppo nel corso del tempo e verificandone la diffusione nei singoli settori della ricerca (MOSCATI P. 1999C). L'enorme progresso delle tecnologie informatiche ha comportato un'evoluzione nel tipo di approccio da parte degli archeologi all'automazione dei propri dati. Alcuni settori applicativi, come quello dei metodi

di analisi statistica e delle banche-dati, di più tradizionale applicazione in ambito archeologico, hanno risentito di tale sviluppo.

Gli anni Novanta, del resto, hanno portato numerose altre novità: il potenziamento delle tecniche di trattamento delle immagini e della grafica computerizzata e soprattutto l'avvento e la diffusione dei GIS, piattaforma ideale per l'integrazione fra informazioni di natura diversa e per l'inserimento dell'elemento spaziale, dato imprescindibile in ogni forma di documentazione dei beni. In quei settori, in cui è apparso più forte il rischio di esaltare le potenzialità tecniche ed espressive del mezzo informatico e di rivolgersi all'aspetto della comunicazione visiva più che al contenuto della realtà archeologica rappresentata e "animata", l'esistenza di una sede editoriale attenta a privilegiare gli aspetti metodologici della ricerca ha consentito di rivalutare l'intervento fondamentale dell'archeologo.

La vivace attività scientifica che emerge dalle pagine di ogni volume di "Archeologia e Calcolatori" consente inoltre di evidenziare nuovi percorsi nell'utilizzazione dell'informatica in archeologia e di dimostrare la profonda incidenza delle tecnologie informatiche sulle metodologie di ricerca tradizionali. La via seguita in questi anni dimostra che un approccio interdisciplinare seriamente perseguito consente di superare gli ostacoli incontrati e di giungere alla definizione di vie innovative di ricerca, in cui lo strumento informatico costituisce un supporto metodologicamente integrante all'indagine archeologica.

Gli elementi che hanno contribuito a questo incontro sono molteplici e si possono individuare all'interno di problematiche sia informatiche sia archeologiche. Da un lato la trasparenza, la trasportabilità dei dati, l'uso di standard riconosciuti internazionalmente per la codifica delle informazioni; l'integrazione e la costituzione di reti; la realizzazione di modelli dinamici di archiviazione e interrogazione della documentazione. Dall'altro lato, la descrizione formalizzata dei dati archeologici e delle loro relazioni; l'attenzione alle fasi di documentazione e di rappresentazione del record archeologico; la diffusione di nuove forme di presentazione delle conoscenze, più favorevoli a una consultazione interattiva.

Il problema della pubblicazione dei risultati delle ricerche, non poteva restare avulso, a partire dalla metà degli anni Novanta e in modo ancor più pressante oggi, dal crescente fenomeno della diffusione dei dati in linea e dal riconoscimento delle potenzialità delle reti telematiche per la costituzione di archivi e banche dati. Da queste considerazioni ed esperienze nasce il progetto che qui si presenta, costruito sugli standard vigenti, su tecnologie ormai consolidate e affidabili, nella prospettiva di fornire una soluzione che sia scalabile su dimensioni maggiori, avente valore di modello, di punto di riferimento, di stimolo per nuove e più impegnative applicazioni.

### *3. Open archive protocol for metadata harvesting (OAI-PMH)*

OAI-PMH è stato proposto nel 2001 quale protocollo standard per la condivisione dei metadati di risorse all'interno di comunità interdisciplinari (LAGOZE C., VAN DE SOMPEL H. 2001) (si veda anche LYNCH C.A. 2001). Il protocollo usa il paradigma di harvesting e si basa su XML e HTTP.

La filosofia degli Open Archives distingue i fornitori di dati (*data providers*) dai fornitori di servizi (*service providers*). Mentre i primi debbono produrre e rendere disponibili le risorse e i loro metadati descrittivi, i secondi aggregano i metadati in archivi centralizzati (mediante la procedura di *harvesting*, ripetuta di solito su base mensile) e forniscono servizi verso l'utenza di rete, usualmente costituiti da indici e applicazioni di ricerca cross-archive.

Per i data provider sono disponibili alcuni software per l'implementazione di archivi aperti istituzionali (*repositories*): tra quelli che hanno avuto buona diffusione i più noti sono Dspace (<http://www.dspace.org/>, realizzato dal MIT e dalla Hewlett-Packard) ed Eprints (<http://www.eprints.org/> della University of Southampton, UK). Tuttavia essi non appaiono di facile utilizzo. Questi software, pur essendo Open Source, sono abbastanza costosi in termini di costi di configurazione e gestione e richiedono consulenti professionali esperti.

La creazione di archivi OAI è quindi ancora un problema di difficile soluzione. Per molti potenziali data provider, le spese e i tempi necessari per dare avvio ad un progetto OA appaiono insostenibili e non rapportati all'entità delle informazioni da diffondere. Uno delle più grandi barriere deriva dall'utilizzo abbastanza generalizzato della piattaforma Unix. I software per l'implementazione di OAI-PMH sono per la maggior parte realizzati per questo sistema operativo, che richiede specifiche competenze, diffuse nei dipartimenti universitari più grandi, ma non altrettanto nei piccoli centri di ricerca e nei centri di studio dell'ambiente umanistico. Questo problema è stato riconosciuto dagli autori di OAI-PMH e appare la vera barriera da infrangere per una più ampia diffusione degli OA.

#### 4. Il modello "static repository" per i data provider

Una soluzione per abbattere le difficoltà di realizzazione dei repository OAI è stata recentemente proposta. Essa si basa su una differente architettura, definita "Static Repository/Static Repository Gateway". È stata ideata da HOCHSTENBACH P., JEREZ H., VAN DE SOMPEL H. nel 2003 (vd. Fig. 1) e rapidamente sostenuta da OAI, che l'ha elevata al rango di *Implementation Guidelines* (<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>).

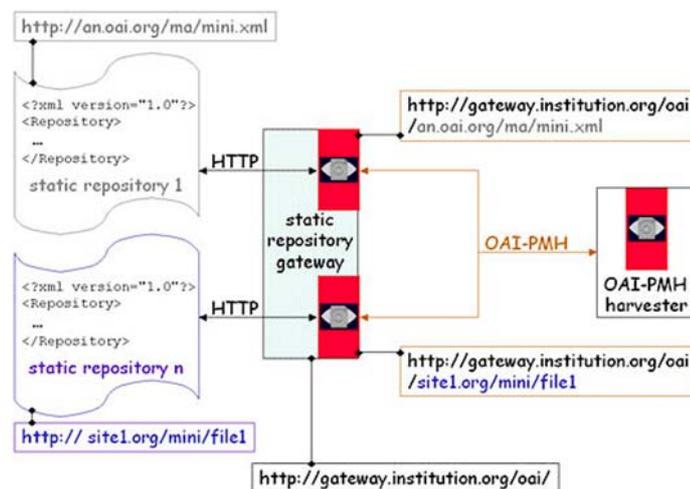


Fig.1 Modello Static Repository – Static Repository Gateway – Harvester (from OAI website)

Secondo questo modello, il data provider genera soltanto un file XML contenente i metadati (static repository file) e gli assegna poi un URL statico ponendolo su un website. Tra il data provider e il service provider è quindi inserita l'intermediazione di un server, definito *gateway*.

Il gateway, gestito da una terza parte, ha il compito di raccogliere le sottoscrizioni di più static repository file (prodotti da più data provider) e fornire supporto al protocollo OAI-PMH verso gli *harvesters* (service provider).

ISCIMA ha adottato questo modello, realizzando attraverso le sue strutture informatiche tutte le parti e i servizi costituenti per il lato data provider, sulla scorta della piattaforma Windows 2000, delle tecnologie ASP, DOM, VBscript e XML-XSL. Per gestire i metadati, è usato l'RDBMS Access (Microsoft). L'esportazione del File XML del repository, conforme allo schema richiesto da OAI, è automatizzata da uno script. L'operatore che crea e inserisce i metadati non deve avere nessuna conoscenza di XML né di tecnologie di rete. I processi sono eseguiti alla pressione di semplici bottoni di comando.

Ci è gradito evidenziare come il nostro Istituto sia stato il primo ad utilizzare direttamente questo nuovo modello per produrre un'applicazione pratica. Il repository di Archeologia e Calcolatori è stato segnalato da OAI nella mailing list dei data provider come il primo al mondo ad apparire nella veste di *direct static repository implementation*.

Attualmente l'archivio della rivista "Archeologia e Calcolatori", realizzato con questo software, è on-line con tutti i metadati descrittivi degli articoli pubblicati e offre già una parte di postprint full text in PDF.

Il repository è sottoposto ad harvesting dai service provider di OAI, in particolare segnaliamo (per doveri istituzionali) l'harvester SAIL del CNR (un progetto sperimentale: <http://eprints.bo.cnr.it/>) e OAISTER dell'Università del Michigan (<http://oaister.umdl.umich.edu/o/oaister/>) che è il più grande servizio di ricerca cross-archive OAI del web.

L'applicazione usa oggi un gateway esterno, secondo la filosofia originale degli autori del modello, amministrato dall'università di Ghent (Belgio), ma è in corso di studio la realizzazione di un gateway proprietario ISCIMA. La realizzazione di un portale gateway/harvester, basato sul paradigma OAI-PMH, indirizzato agli studi storico archeologici del Mediterraneo antico, in grado di proporre un'interfaccia unica per l'interrogazione di più archivi connessi, appare l'obiettivo futuro verso cui questo progetto può evolvere.

## 5. Dublin Core Initiative

Il protocollo OAI-PMH usa lo schema di metadati Dublin Core (<http://dublincore.org/>), che nasce in ambito bibliotecario ed è gestito da DCMI (Dublin Core Metadata Initiative). Dublin Core (DC) è basato su un set ristretto di elementi - quindici - in grado di descrivere adeguatamente, anche in virtù della ripetibilità, risorse documentarie di ogni genere.

Il ristretto numero di elementi non permette di sostituire senza rinunce il Dublin Core a schemi più complessi di descrizione documentaria, quali l'EAD, etc., né di considerarlo sempre e comunque adeguato ad ogni classe di risorse; tuttavia la sua semplicità offre significativi vantaggi, che hanno attratto molti progetti, sia culturali sia tecnologici.

In particolare, per quanto riguarda i documenti testuali, mentre il formato PDF sembra essere universalmente adottato per la diffusione dei documenti - e la codifica XML strumento potente per la strutturazione semantica di testi (Cfr. BONINCONTRO I. 2001; BARCHESI C., CECCARELLI L. 2004) - lo schema Dublin Core appare perfettamente adeguato per la creazione di metadati descrittivi efficaci. Lo schema Dublin Core è facile da implementare (HILLMANN D. 2003) chiaro da comprendere ed è adottato da OAI-PMH come schema *mandatory* del protocollo.

## 6. La realizzazione del repository OAI di "Archeologia e Calcolatori"

### 6.1 Il database di amministrazione dei metadati

Per gestire i 15 elementi del Dublin Core è comodo usare un database relazionale; gli elementi sono quasi tutti ripetibili; un database può mapparli su tabelle diverse, legate da legami relazionali attraverso l'ID.

Nella nostra applicazione lo schema relazionale è molto semplice, perché la tipologia omogenea dei nostri dati (schede bibliografiche) non richiedeva ripetizioni di elementi tali da richiedere uno schema complesso. Il modello relazionale è stato denormalizzato per semplificare gli script di gestione. Si è utilizzata una tabella collegata 1-M solo per l'elemento DC-Creator (autori). Tutti gli altri elementi DC sono integrati nei campi di un'unica tabella, nella quale sono inclusi anche gli elementi ripetuti (come il DC-subject, del quale erano richieste 4 istanze). Altre tabelle, non legate da relazioni, sono state utilizzate per contenere valori predefiniti per gli elementi DC-Format (popolato con valori MIME, <http://www.iana.org/assignments/media-types/>), DC-type (popolato con valori DCMI Type Vocabulary, <http://dublincore.org/documents/2004/06/14/dcmi-type-vocabulary>) e DC-language, popolato con NISO Z39.53 Language Codes (<http://www.sil.org/sgml/nisoLang3-1994.html>). Un'ulteriore tabella è stata utilizzata per gestire la sezione header (intestazione) del repository. In essa si definiscono i dati di amministrazione, gli

URL del gateway e del repository, le caratteristiche di granulazione dei dati e altri parametri d'ambiente.

## 6.2 Esportazione del database in un file XML

L'implementazione del protocollo OAI sulla base del modello Static Repository/Static Repository Gateway richiede al data provider di esporre i metadati in un file XML conforme allo schema pubblicato su <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm> (nella Fig. 2 riportiamo un esempio tratto da HOCHSTENBACH P. et al. 2003). Il file deve contenere sia i metadati, sia le informazioni di supporto necessarie all'harvesting sulla base del protocollo OAI-PMH.

```
<?xml version="1.0" encoding="UTF-8"?>
<Repository>
  <Identify>
    <oai:repositoryName>Demo</oai:repositoryName>
    <oai:baseURL>http://an.oai.org/ma/mini.xml</oai:baseURL>
    <oai:protocolVersion>2.0</oai:protocolVersion>
    <oai:adminEmail>jondoe@oai.org</oai:adminEmail>
    <oai:earliestDatestamp>2002-09-19</oai:earliestDatestamp>
    <oai:deletedRecord>no</oai:deletedRecord>
    <oai:granularity>YYYY-MM-DD</oai:granularity>
  </Identify>
  <ListMetadataFormats>
    <oai:metadataFormat>
      <oai:metadataPrefix>oai_dc</oai:metadataPrefix>
      <oai:schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd </oai:schema>
      <oai:metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</oai:metadataNamespace>
    </oai:metadataFormat>
    <oai:metadataFormat>
      <oai:metadataPrefix>oai_rfc1807</oai:metadataPrefix>
      <oai:schema>http://www.openarchives.org/OAI/1.1/rfc1807.xsd</oai:schema>
      <oai:metadataNamespace>http://info.internet.isi.edu:80/in-
notes/rfc/files/rfc1807.txt</oai:metadataNamespace>
    </oai:metadataFormat>
  </ListMetadataFormats>
  <ListRecords metadataPrefix="oai_dc">
    <oai:record>
      <oai:header>
        <oai:identifier>oai:an.oai.org:0112017</oai:identifier>
        <oai:datestamp>2003-01-17</oai:datestamp>
      </oai:header>
      <oai:metadata>
        <oai_dc:dc>
          <dc:title>Structural Metadata</dc:title>
          <dc:creator>Smith, Hector</dc:creator>
          <dc:subject>Digital Libraries</dc:subject>
          <dc:date>2001-12-14</dc:date>
        </oai_dc:dc>
      </oai:metadata>
    </oai:record>
  </ListRecords>
  <ListRecords metadataPrefix="oai_rfc1807">
    <oai:record>
      <oai:header>
        <oai:identifier>oai:an.oai.org:0112017</oai:identifier>
        <oai:datestamp>2002-01-15</oai:datestamp>
      </oai:header>
      <oai:metadata>
        <oai_rfc1897:rfc1807>
          <rfc1807:bib-version>v2</rfc1807:bib-version>
          <rfc1807:id>0112017</rfc1807:id>
          <rfc1807:entry>January 15, 2002</rfc1807:entry>
          <rfc1807:title>Structural Metadata</ rfc1807:title>
          <rfc1807:author>Hector Smith</rfc1807:author>
          <rfc1807:date>December 14, 2001</rfc1807:date>
        </oai_rfc1897:rfc1807>
      </oai:metadata>
    </oai:record>
  </ListRecords>
</Repository>
```

Fig. 2 XML Static Repository file (da Hochstenbach P. et al, 2003)

Le dimensioni massime del repository XML sono definite da OAI (questo modello è applicabile ad archivi di piccole e medie dimensioni). La massima dimensione è di 2Mbyte; quale ulteriore limitazione OAI stabilisce che l'archivio non ecceda i 5000 record.

L'esportazione dei record di un database relazionale in un file XML può essere facilmente automatizzata. Nel nostro progetto abbiamo utilizzato una procedura Visual Basic for Application sviluppata appositamente e inserita all'interno di un modulo di programmazione di un semplice database Access di Microsoft (Fig. 3). Lo script è eseguito a richiesta, premendo un bottone evidenziato dall'etichetta: "Avvia processo di trasformazione in XML". Lo script preleva i dati dalle tabelle e assembla un file testuale aggiungendo ai valori dei campi la marcatura XML richiesta. Il file generato è salvato con l'estensione ".xml". Il file (Fig. 4) è well-formed e valido secondo lo schema emesso da OAI; la verifica può essere compiuta con XMLspy, il software più diffuso per la sviluppo di applicazioni XML.

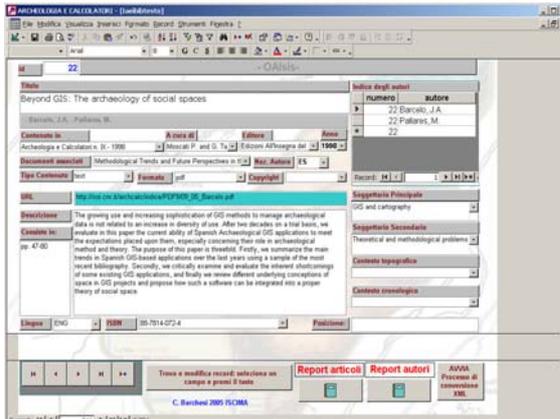


Fig. 3 Un database Access gestisce i metadati

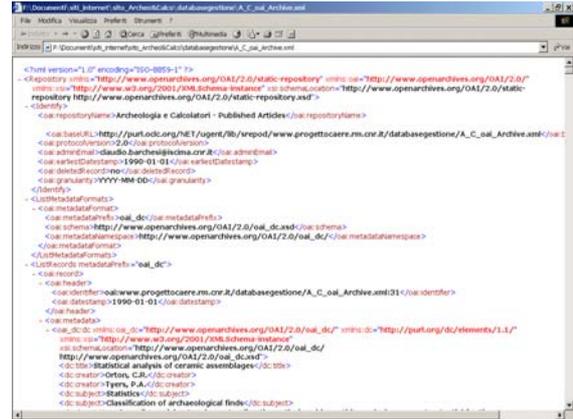


Fig. 4 Il database è esportato in XML-OAI

Se il file è salvato direttamente in una cartella del sito web istituzionale assume immediatamente anche uno specifico indirizzo statico URL (definito Static Repository URL). Per trasformare questo file in un archivio OAI è necessario istanziare l'intermediazione di un gateway esterno in grado di rispondere ai service provider sulla base del protocollo OAI-PMH. Per avviare l'intermediazione di un gateway è richiesto l'invio di uno specifico comando al suo indirizzo URL (definito Static Repository Gateway URL) nel quale sia specificato l'indirizzo URL del repository da gestire:

```
Static Repository Gateway URL?initiate=Static Repository URL
```

Il gateway esegue delle verifiche di correttezza sul file sottoposto e se queste hanno successo avvia immediatamente l'intermediazione. OAI mette a disposizione uno strumento per verificare il funzionamento dell'implementazione: attraverso il sito web "Repository Explorer" (<http://re.cs.uct.ac.za/>) si può collaudare il funzionamento del gateway e dello static repository interconnessi (HUSSEIN S. 2001). Il superamento del test permette di procedere con la registrazione del repository - ormai sicuramente privo di errori - sul sito OAI, nella lista degli Official Data Provider (<http://www.openarchives.org/data/registerasprovider.html>); anche qui sono previsti alcuni test di validità.

Dopo aver effettuato la registrazione si appare finalmente in OAI come data provider e si è pronti a richiedere l'harvesting ai service provider sui quali si vuole comparire (è sufficiente inviare un mail agli amministratori).

Più semplicemente si può aspettare che essi aggiornino le loro liste di harvesting sulla base di quelle ufficiali di OAI. Nella Fig. 5 i metadati di "Archeologia e Calcolatori" così come appaiono in OAISTER, il più grande servizio di ricerca cross-archive del web.

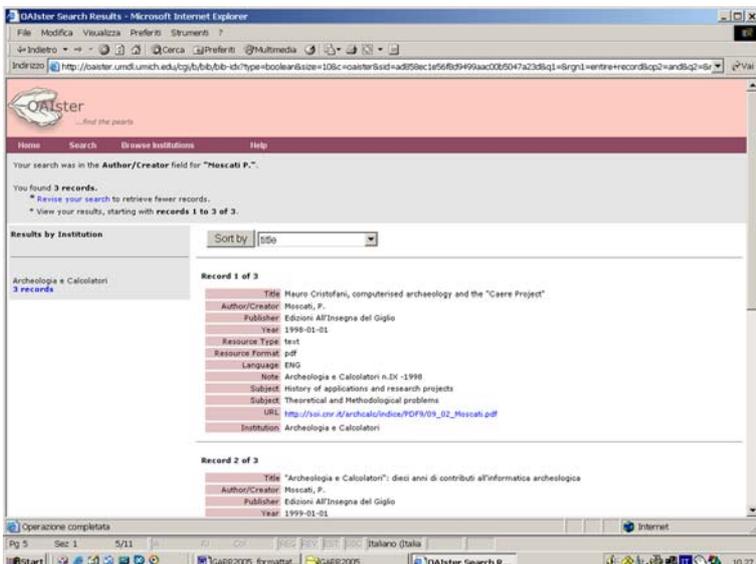


Fig. 5 I metadati di "Archeologia e Calcolatori" sono presenti in OAISTER

## 7. Servizi web aggiuntivi

Oltre all'implementazione del repository OAI, l'applicazione sviluppata per "Archeologia e Calcolatori" comprende due ulteriori servizi. Il primo è creato appositamente per Googlebot (spider di Google) ed è in grado di fornire sul sito web della rivista URL persistenti per tutti i record contenuti nel repository.

### 7.1 Permettere l'indicizzazione di Google

Il servizio crea dinamicamente per ogni record del repository una pagina HTML con i metadati *embedded*. Le pagine sono associate a semplici hyperlink, con un indirizzo persistente. E' stata richiesta l'indicizzazione a *Scholar.Google*, la sezione "scientifica" di Google, che indicizza gli articoli scientifici.

Data la facilità con cui un file XML può essere trasformato in adeguate forme di rappresentazione mediante i fogli di stile XSL parametrizzati abbiamo utilizzato le possibilità offerte dall'MSXML parser su DOM all'interno di pagine ASP compilate in Vbscript. Con questa tecnologia è stato possibile assemblare dinamicamente restituzioni XSLT-HTML adeguate per ogni record contenuto nel repository. La pagina HTML che è generata contiene tutti i metadati DC, sia nel body del documento (rappresentati in forma di scheda) sia nell'header, in forma di metatag espressi secondo i dettami di Dublin Core Initiative (KUNZE J). Ogni record ha dunque un URL persistente, raggiungibile in soli tre iperlink dall'homepage della rivista. (→ Years List → Year Articles → Issue). In questo modo tutto il repository può essere indicizzato da Google. Questo servizio è analogo a DP9 (LIU X. 2002) (<http://arc.cs.odu.edu:8080/dp9/index.jsp>) un progetto pensato per permettere a Googlebot l'indicizzazione di repositories OAI, ma mentre questo gestisce più data provider, il nostro è dedicato ad un solo static repository. Nella Fig. 6 si rappresenta lo schema logico del servizio per Google.

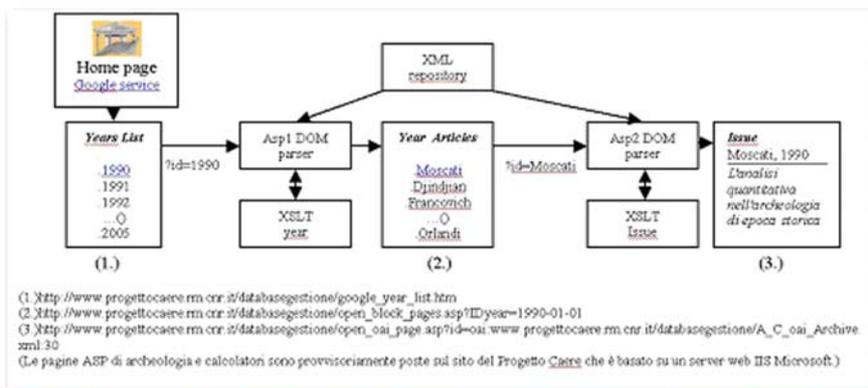


Fig. 6 Offrire a Google i propri record, una soluzione XML-DOM-XSLT

## 7.2 Servizio di ricerca locale

Oltre all'implementazione del repository OAI e del servizio per Google il progetto comprende anche, sul sito web di "Archeologia e Calcolatori", un motore di ricerca locale (Fig. 7 a,b,c). Questa è la terza via attraverso cui un utente può accedere alle risorse messe in rete. L'applicazione opera direttamente su una copia replicata del database originale, attraverso un collegamento ODBC gestito da pagine ASP attraverso i metodi di ADO. E' possibile cercare per autore, soggetto, anno. Data la natura dei documenti gestiti, che si riferiscono spesso a precisi contesti storico-archeologici, è prevista, e sarà presto implementata, la possibilità di cercare articoli in base a un filtro topografico e cronologico. L'applicazione per l'interrogazione locale del repository è accessibile presso ([http://soi.cnr.it/archcalc/Bibliografia\\_new.htm](http://soi.cnr.it/archcalc/Bibliografia_new.htm)).



Fig. 7a Ricerca locale per autori

Fig. 7b Ricerca per Soggetti e Data

Fig. 7c Risultati

Nella Fig. 8 è rappresentato l'intero progetto, definito complessivamente *OAI sistema*. L'obiettivo perseguito è quello della massima visibilità e della più ampia condivisione dei dati. Ogni risorsa elettronica inserita nel database può essere rintracciata attraverso tre differenti percorsi web: Google, OAI e website della rivista.

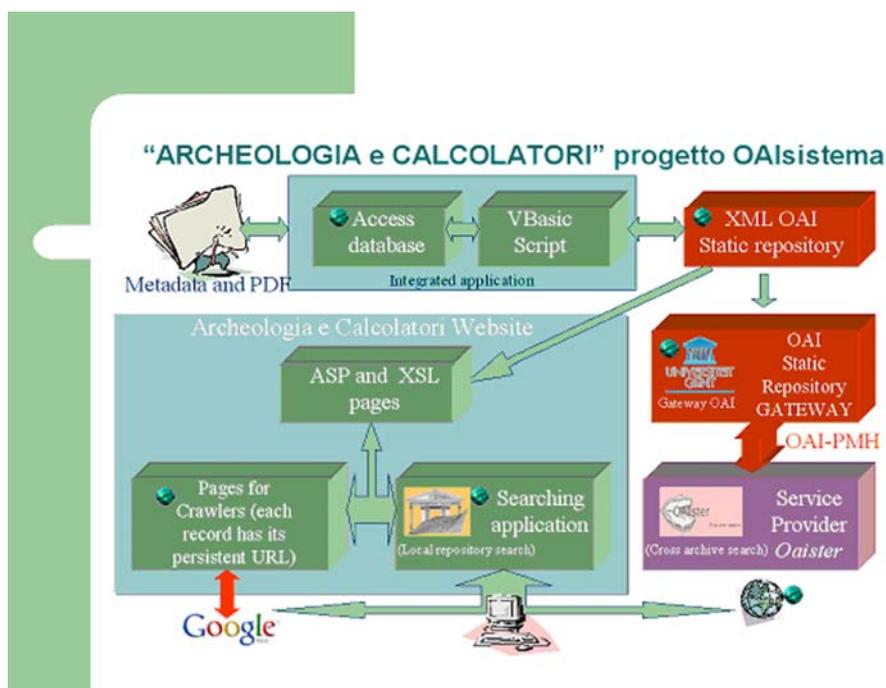


Fig. 8 Lo schema del progetto

## 8. *Sviluppi previsti*

Questo progetto offre una soluzione a basso impatto di difficoltà per i data provider OAI, ma prevede di evolvere verso la realizzazione di servizi di tipo service provider.

Il gateway utilizzato da “Archeologia e Calcolatori” è gestito dall’Università di Ghent (Belgio). Si prevede di implementare presto un gateway ISCIMA. In rete, all’interno del circuito OAI, ne esistono alcuni open source (l’elenco dei software open source per gli OA è presso <http://www.openarchives.org/tools/tools.html>) tuttavia, mentre i tentativi di implementazione del gateway OAI-PMH per la piattaforma IIS di Microsoft (in ASP) realizzato da UIUC (University of Illinois, Urbana-Champaign), anche a fronte di considerevoli sforzi, non ha dato risultati positivi, si sta valutando la possibilità di sperimentare il gateway originale del modello Static repository/Static repository gateway (scaricabile presso <http://sreprod.sourceforge.net/>).

L’implementazione di un gateway istituzionale permetterà la costituzione di un portale OAI tematico, aperto ai centri di ricerca che svolgono studi sulle civiltà italiche e del mediterraneo antico. Il gateway sarà in grado di intermediare i repositories di tutte le pubblicazioni scientifiche e della ‘letteratura grigia’ prodotta per quest’ambito disciplinare. L’uso della nostra applicazione di gestione dei metadati, o di una analoga, potrebbe rendere la creazione dei contenuti semplice. In questa prospettiva, si potrebbe procedere all’implementazione di un servizio di harvesting e alla creazione di un website per la ricerca cross-archive. Per gli archeologi si aprirebbe la possibilità di condividere informazioni, articoli scientifici, comunicazioni, immagini, multimedia all’interno di un portale ‘OAI’ il cui sviluppo sarebbe delegato nei contenuti ai centri scientifici aderenti, ma che apparirebbe aperto a tutto il mondo della ricerca per la consultazione. Un’adeguata politica di accesso al portale per i data provider delegherebbe la funzione di *peer review* ai direttori delle Istituzioni afferenti, garantendo in tal modo anche l’autorevolezza delle informazioni contenute nell’harvester centrale.

## 9. *Ringraziamenti*

Si ringrazia Paola Moscati, Direttore Responsabile di “Archeologia e Calcolatori”, per una serie di indicazioni relative al paragrafo 2.

## **BIBLIOGRAFIA**

- ABBATTISTA G. 2004, *Comunicazione telematica di risorse scientifiche in ambito umanistico*, in “Editoria Elettronica, Open Archives e comunità digitali – Convegno sui nuovi scenari dell’editoria elettronica”, Trieste, SISSA, 13 maggio 2004, (<http://www.uiniv.trieste.it/~storia/docenti/Abbattista/editoria/editoria.htm> (risorsa accessibile in data 15/4/2005))
- BARCHESI C. 2001, *Progetto Caere: un'applicazione Internet attiva per l'Information Retrieval di documenti SGML*, “Archeologia e Calcolatori”, 12, 2001, pp. 170-89. ([http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:297](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:297))
- BARCHESI C. 2004A, «*Il settore informatico dell'ISCIMA. Un percorso ventennale nell'informatica archeologica*», “Proceedings of CAA2004 – Italia”, Aprile 2004, Prato, in “Journal for the Intercultural and Interdisciplinary Archaeology”, in corso di edizione
- BARCHESI C., CECCARELLI L. 2004B, *Linguaggi dichiarativi per la ricerca archeologica*, “Archeologia e Calcolatori”, 15, 2004, pp. 95-114. [http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:361](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:361)
- BARCHESI C., 2005, *OAI sistema a Simple solution for Implementing OAI-PMH on Small size archives*, Proceedings of CAA2005, Tomar 21-25 marzo 2005, in corso di edizione.
- BONINCONTRO I. 2001, *Progetto Caere: prospettive di applicazione degli standard internazionali per la codifica dei dati testuali*, “Archeologia e Calcolatori”, 12, 2001, pp. 55-69. [http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:312](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:312)
- CANTONE F. 2005, *Shared technologies in Archeologia: nuove prospettive di gestione e condivisione di dati in rete*, “Archeologia e Calcolatori”, 16, 2005, in stampa.
- CRISTOFANI M., FRANCOVICH R. 1990, *Editoriale*, “Archeologia e Calcolatori”, 1, 1990, pp. 7-8
- FOULONNEAU M., DAWSON D., 2003 *Protocol For Metadata Harvesting, Practices of cultural heritage actors*, [http://www.oaforum.org/otherfiles/oaf\\_d48\\_cser3\\_foullonneau.pdf](http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf) (risorsa accessibile in data 15/4/2005)
- HOCHSTENBACH P., JEREZ H., VAN DE SOMPEL H. 2003, *The OAI-PMH Static Repository and Static Repository Gateway*, <http://lib-www.lanl.gov/~herbertv/papers/jcdl2003-submitted-draft.pdf> (risorsa accessibile in data 15/4/2005)
- HILLMANN D. 2003, *Using Dublin Core*, <http://dublincore.org/documents/usageguide/> (risorsa accessibile in data 15/4/2005)
- HUSSEIN S. 2001, *Using the repository explorer to achieve OAI protocol compliance*, in JCDL '01, June 24-28, 2001, Roanoke, Virginia, USA.. New York, NY:ACM press (2001):459. ([http://www.dlib.vt.edu/projects/OAI/reports/jcdl\\_2001\\_demo\\_repository\\_explorer.pdf](http://www.dlib.vt.edu/projects/OAI/reports/jcdl_2001_demo_repository_explorer.pdf)) (risorsa accessibile in data 15/4/2005)

KUNZE J., *Encoding Dublin Core in HTML*, <http://www.ietf.org/rfc/rfc2731.txt> (risorsa accessibile in data 15/4/2005)

LAGOZE C., VAN DE SOMPEL H. 2001, *The Open Archives Initiative: Building a low-barrier interoperability framework*, in JCDL '01, June 24-28, 2001, Roanoke, Virginia, USA.. New York, NY:ACM press (2001):54-62 (<http://www.openarchives.org/documents/jcdl2001-oai.pdf>) (risorsa accessibile in data 15/4/2005)

LIU X., K.MALY, M.ZUBAIR, M.L.NELSON 2002, *DP9: an OAI gateway service for Web Crawlers*, JCDL'02, July 13-17, 2002, Portland, Oregon, USA, [http://www.cs.odu.edu/~liu\\_x/dp9/dp9.pdf](http://www.cs.odu.edu/~liu_x/dp9/dp9.pdf) (risorsa accessibile in data 15/4/2005)

LYNCH C. A. 2001, *Metadata harvesting and Open Archives Initiative*, ARL Bimonthly Report 217. Washington D.C.:ARL (August 2001): 1-9. (<http://www.arl.org/newsltr/217/mhp.html>) (risorsa accessibile in data 15/4/2005)

MOSCATI P. 1999A, *The 'Caere Project': Methodological and Technical Considerations*, in A. Guarino (ed.), *Proceedings of the II International Congress "Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin (Paris 1999)*, Paris, CNR-CNRS, 119-128.

MOSCATI P., MARIOTTI S., LIMATA B. 1999B, *Il "Progetto Caere": un esempio di informatizzazione dei diari di scavo*, «Archeologia e Calcolatori» 10, 1999, pp. 165-188.  
[http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:257](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:257)

MOSCATI P. 1999C, *Archeologia e Calcolatori: dieci anni di contributi all'informatica archeologica*, «Archeologia e Calcolatori», 10, 1999, pp. 343-352.  
[http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:256](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:256)

MOSCATI P. 2001, *Progetto Caere: questioni di metodo e sperimentazioni*, «Archeologia e Calcolatori», 12, 2001, pp. 47-53.  
[http://www.progettocaere.rm.cnr.it/databasegestione/open\\_oai\\_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A\\_C\\_oai\\_Archive.xml:296](http://www.progettocaere.rm.cnr.it/databasegestione/open_oai_page.asp?id=oai:www.progettocaere.rm.cnr.it/databasegestione/A_C_oai_Archive.xml:296)

MOSCATI P. 2002A, *From an Etruscan town to modern technologies: new advancements in the "Caere Project"*, in F. Djindjian, P. Moscati (eds.), *XIV Congress of the UISPP, Commission IV Data Management and Mathematical Methods in Archaeology (Liège 2001)*, «Archeologia e Calcolatori», 13, pp. 135-149.

MOSCATI P. 2002B, *Archeologia informatica: tra tradizione e rinnovamento*, «Bollettino ICR» n.s. 5, 2002, pp. 21-27.

PAPADOPOULOS D.C., MAVRIKAS E. C. 2004, *Peer-to-Peer Ways to Cultural Heritage*, *Proceedings of CAA2003 conference "Enter the Past"*, Aprile 2003 – Vienna, Bar International Series 1227, 2004, pp. 389-392

TAMMARO A.M., DE GREGORI T. 2004, *Ruolo e funzionalità dei depositi istituzionali: uno spazio virtuale al servizio della comunità scientifica*, «Biblioteche Oggi», 10, pp. 7-19.  
<http://www.bibliotecheoggi.it/2004/20041000701.pdf> (risorsa accessibile in data 15/4/2005)

VAN DE SOMPEL H., LAGOZE C. 2000, *The Santa Fe Convention of the Open Archives Initiative*, D-Lib Magazine, 2000. <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html> (risorsa accessibile in data 15/4/2005)