

## «ARCHEOLOGIA E CALCOLATORI»: NUOVE STRATEGIE PER LA DIFFUSIONE DI CONTENUTI IN RETE SULLA BASE DELL'OAI-PMH

### 1. INTRODUZIONE

La rete Internet rappresenta un canale di straordinaria efficacia per la diffusione dei risultati della ricerca. Tra le sue applicazioni il web, che ha avuto enorme fortuna, è oggi largamente usato come strumento di prima informazione e di ricerca bibliografica non solo dagli scienziati della fisica e della chimica, tra i primi a sfruttare le potenzialità offerte da questo mezzo per velocizzare e migliorare la comunicazione scientifica, ma anche dagli archeologi. Tutte le principali Istituzioni nell'ambito degli studi archeologici, internazionali e nazionali, sono oggi convinte dell'importanza di questo canale divulgativo e iniziano a pubblicare sul web parte dei propri lavori, cercando di migliorare in questo modo la visibilità scientifica delle attività svolte e dei progetti di ricerca.

La continua crescita dell'informazione disponibile in rete crea tuttavia alcune difficoltà al funzionamento dei motori di ricerca (Google, Altavista, Yahoo, etc.), che spesso, in risposta a interrogazioni poco dettagliate, forniscono troppi risultati. Per trovare soluzioni a questo problema, sembra necessario sperimentare nuovi modi di catalogare le informazioni da condividere, per renderle più facilmente rintracciabili. Nell'ambito dell'informatica archeologica gli standard per la descrizione e la catalogazione elettronica dei documenti sono oggi, in questo scenario, al centro di un grande interesse. Nella prospettiva del web semantico, sono state proposte nella ricerca archeologica alcune complesse ontologie, che promettono di affermarsi quali standard condivisi del web di domani. Le ontologie e il web semantico suscitano tuttavia alcune perplessità. Si tratta infatti di concetti molto complessi e spesso la complessità di un modello, per quanto esso sia valido, è un grande ostacolo alla sua diffusione.

Che il web di oggi sia basato sulla semplicità lo mostra esemplarmente la scelta di Google. Il motore di ricerca più diffuso, che utilizza algoritmi complicatissimi e gestisce database dalle dimensioni inimmaginabili, si presenta all'utente con un'interfaccia che mostra un minimalismo grafico e una semplicità d'utilizzo estrema: l'homepage di Google è composta sostanzialmente da un logo, una casella di testo e un bottone su fondo bianco. Sembra inoltre che Google sia assai poco propenso a credere nel futuro del web semantico e piuttosto che invitare gli utenti del web a cambiare il modo di trattare i propri documenti, adottando le tecnologie OWL e RDF, continua ad impegnarsi nel migliorare i suoi algoritmi di indicizzazione basati sui link e sui termini

contenuti nelle pagine web. Bisogna riconoscere che di web semantico si parla ormai da sette anni (la prima nota di BERNERS-LEE è del 1998) – un tempo lunghissimo nella cronologia dell'informatica – ma che esso è ancora relegato in ambiti sperimentali, come dimostra un certo scetticismo che inizia a circolare anche in rete (cfr. ad esempio NEE 2005).

Il web semantico appare un buon terreno accademico per gli specialisti dell'Intelligenza Artificiale, ma ad alcuni sembra utopistico rivoluzionare un web che cresce a ritmi esponenziali, muovendo masse di denaro enormi, rinnegando le ragioni di questo successo, che dipendono per la più gran parte proprio dalla semplicità delle tecnologie impiegate.

Nel frattempo, per migliorare la circolazione delle informazioni scientifiche potrebbe essere più proficuo, almeno nel breve periodo, affidarsi a schemi di metadati semplici, renderne facile la creazione e la diffusione in rete, creare sistemi di Information Retrieval basati su questi schemi. A questo occorrerebbe aggiungere servizi per facilitare l'indicizzazione corretta dei documenti da parte dei grandi motori di ricerca, poiché essi rappresentano per gli utenti del web uno standard operativo de facto.

Sempre per quanto attiene la semplicità, relativamente ai metadati ci sono ulteriori considerazioni da fare. Gli schemi troppo complessi diventano autoreferenziali e sono applicati solo da comunità ristrette. In questo modo non possono contribuire alla formazione di un sapere interdisciplinare realmente condiviso. La condivisione dell'informazione presuppone ambiti e strumenti comuni, che nascono quali risposte tecnologiche alla disponibilità di dati reali e significativi. I dati, tuttavia, diventano disponibili quando gli standard sui quali essi devono essere modellati sono applicabili dagli utenti con uno sforzo ragionevole, non quando la complessità dei modelli blocca sul nascere ogni interesse per la loro applicazione.

Come appare dalla storia della cultura informatica, molti processi di definizione di protocolli e linguaggi partono dalla complessità iniziale (che deve recepire tutte le istanze delle varie Istituzioni che partecipano alla loro stesura) per muovere poi verso un deciso sfrondamento del superfluo. Si pensi allo SGML, che prima ha generato il semplice e ristretto HTML, lo standard forse più fortunato della storia dell'informatica, e oggi è soppiantato dal più agile XML; si pensi alla DTD TEI, presto circolata in versioni ridotte: la TEILite e la Bare Bones TEI. Si pensi anche a certi schemi di metadati, oggi sostituiti in molti campi dal più semplice Dublin Core. L'osservazione del nostro recente passato tecnologico ci dovrebbe suggerire di intraprendere per tempo la via verso ciò che appare l'obiettivo di ogni processo evolutivo: la semplicità.

Su questo scenario e con queste premesse, appare plausibile che nel web di domani il linguaggio XML avrà sì una maggior diffusione e permetterà di aggiungere semantica ai contenuti, ma questo avverrà principalmente sotto forma di semplici metadati descrittivi. Il successo avuto da RSS (*RDF Site*

*Summary* o *Really Simple Syndication* che sia), il più popolare e semplice formato per l'esportazione di contenuti web, usato da milioni di utenti, appare esemplare nel chiarire come XML potrà continuare a produrre nuovi schemi e forse nuove tecnologie per la rete, che probabilmente porteranno all'affermarsi di una realtà web dinamica, svincolata da complessità imposte dall'alto, ma comunque indirizzata verso migliori soluzioni. Prima di affermarsi universalmente, queste soluzioni dovranno trovare applicazione nell'uso quotidiano della gran moltitudine di siti web piccoli e medi, amministrati da webmaster pragmatici, concreti, sostanzialmente interessati a migliorare il funzionamento dei propri siti e la loro rintracciabilità in rete.

Una soluzione semplice per la diffusione e la reperibilità della letteratura archeologica sulla base dei metadati, senza considerare l'applicazione di architetture *peer to peer* (forse ancora premature per il mondo della ricerca: comunque cfr. PAPADOPOULOS, MAVRIKAS 2004 e CANTONE 2005), appare oggi quella degli *Open Archives* (OA), che usano per il momento lo schema di metadati Dublin Core, ma incoraggiano e non ostacolano lo sviluppo di altri standard. La filosofia degli OA (che può essere approfondita nelle sue idee di base anche sul sito italiano <http://www.aepic.it/risorse.php>) si fonda sulla distinzione dei ruoli dei soggetti partecipanti, definiti *data providers* e *service providers*. I primi hanno il compito di produrre risorse digitali e i loro metadati, i secondi di usarle per creare servizi. Lo scopo di questa distinzione è duplice: semplificare la creazione e la gestione dei metadati e permetterne l'interrogazione attraverso servizi centralizzati.

L'idea di lasciare la produzione dei metadati descrittivi agli organi istituzionali che possiedono e rendono disponibili le risorse, ma di assemblare poi le metainformazioni di più soggetti produttori in sistemi di interrogazione tematici, la cui amministrazione scientifica sia delegata ad un'authority riconosciuta, sembra oggi abbastanza praticabile e al centro di attente riflessioni anche nel campo degli studi umanistici (ABBATTISTA 2004; TAMMARO, DE GREGORI 2004).

Gli archivi aperti possono avere un'importanza strategica per le risorse elettroniche sviluppate in seno alle discipline storiche ed archeologiche (cfr. ad esempio FOULONNEAU, DAWSON 2003), anche perché la loro tecnologia si basa su standard aperti e linguaggi di marcatura dichiarativi non proprietari, ed è perfettamente in linea con quanto espresso nel documento emanato dalla Comunità Europea nell'aprile del 2001 a Lund ([ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund\\_principles-it.pdf](ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund_principles-it.pdf)), in cui si auspica la creazione e la diffusione di contenuti da collocare sulle reti globali attraverso tecnologie in grado di garantire la disponibilità e l'accessibilità dei dati sul lungo termine.

Alla luce di queste riflessioni, la redazione di «Archeologia e Calcolatori» ha scelto di sperimentare l'implementazione del protocollo OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting: <http://www.openarchives.org/OAI/openarchivesprotocol.html>), con l'obiettivo di creare

un repository che contenga le versioni elettroniche e i metadati descrittivi degli articoli pubblicati, interrogabile attraverso strumenti di ricerca locali ma anche collegato a meta-archivi interdisciplinari, indicizzabile per il web dai principali motori di ricerca.

## 2. LA RIVISTA «ARCHEOLOGIA E CALCOLATORI» E L'OAI-PMH

Nel corso dell'attività editoriale ormai quindicennale della rivista «Archeologia e Calcolatori» 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 (cfr. da ultimo MOSCATI 1999, 2002). La vivace attività scientifica che emerge dalle pagine di ogni volume ha consentito 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.

Il problema della pubblicazione in rete dei risultati delle ricerche non poteva restare avulso, già 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 è nato il progetto che qui si presenta, basato sull'adozione degli standard vigenti e su tecnologie ormai consolidate e affidabili, nella prospettiva di fornire una soluzione che sia scalabile su dimensioni maggiori, quale modello e stimolo per nuove e più impegnative applicazioni.

«Archeologia e Calcolatori» ha inteso in questo modo aggiungere al suo ruolo istituzionale di polo editoriale per il confronto delle teorie e delle applicazioni dell'archeologia computazionale un suo diretto impegno sperimentale, volto all'adozione di tecnologie d'avanguardia per la diffusione di contenuti sui nuovi media. Tra queste tecnologie, l'OAI-PMH, proposto nel 2001 quale protocollo standard per la condivisione dei metadati di risorse all'interno di comunità interdisciplinari (LAGOZE, VAN DE SOMPEL 2001; cfr anche LYNCH 2001) – oggi in rapida affermazione – è apparso quello più adeguato. Il protocollo OAI-PMH usa il paradigma di *harvesting* e si basa su XML e HTTP.

La filosofia degli OA, come si è già anticipato, distingue i fornitori di dati dai fornitori di servizi. Mentre i primi debbono produrre e rendere disponibili le risorse e i loro metadati descrittivi, i secondi devono aggregare i metadati in archivi centralizzati (mediante la procedura di *harvesting*, ripetuta di solito su base mensile) e fornire 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 repertori di risorse elettroniche OAI: 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 onerosi in termini di costi di configurazione e richiedono, almeno inizialmente, consulenti professionali esperti.

Per molti potenziali data provider, le spese necessarie per dare avvio ad un progetto OA appaiono insostenibili. Una 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.

### 3. 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", ideata da HOCHSTENBACH, JEREZ, VAN DE SOMPEL nel 2003 (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>).

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*. Quest'ultimo, 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 harvester (service provider).

L'Istituto di Studi sulle Civiltà Italiane e del Mediterraneo Antico del CNR ha adottato questo modello, realizzando attraverso le sue strutture informatiche tutte le parti e i servizi costituenti per il data provider, sulla scorta della piattaforma Windows 2000, delle tecnologie ASP, DOM, VBscript e XML-XSL. Per gestire i metadati è stato usato il database Access. La generazione del file XML del repository, conforme allo schema richiesto dalla OAI, è automatizzata da uno script inserito nel database.

Con questo sistema, l'operatore che amministra 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. L'applicazione sviluppata dal nostro Istituto usa oggi un gateway esterno, secondo la filosofia

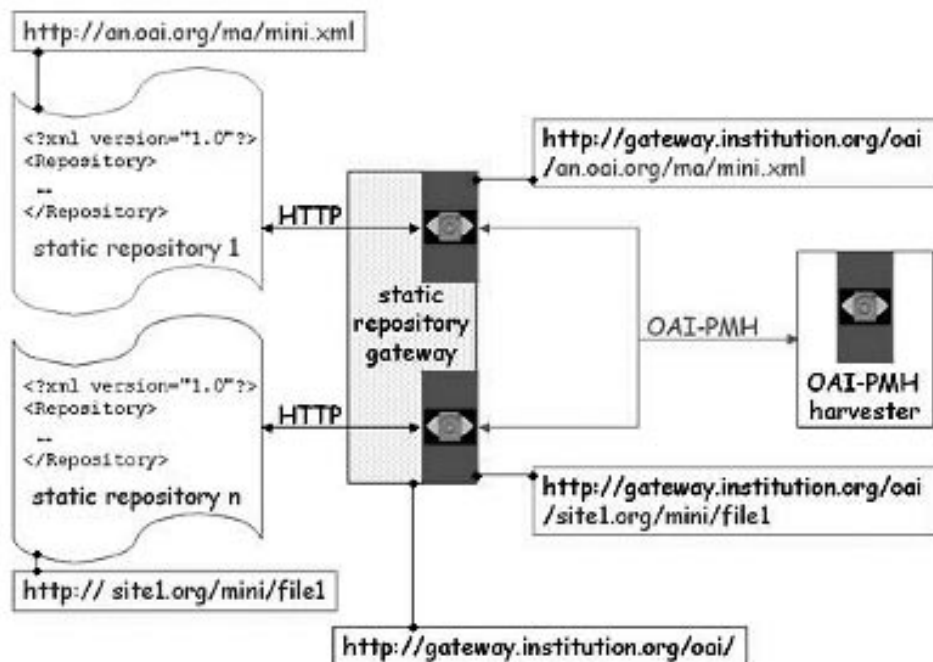


Fig.1 – Modello Static Repository – Static Repository Gateway – Harvester (from OAI website).

originale degli autori del modello, amministrato dall'Università di Ghent (Belgio). Occorre però segnalare che è in corso di studio la realizzazione di un gateway proprietario, grazie ad una collaborazione con il Servizio Reti del CNR.

Ci è gradito evidenziare come il nostro Istituto sia stato il primo ad utilizzare direttamente il modello di P. Hochstenback per generare un'applicazione funzionante. 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» è online con tutti i metadati descrittivi degli articoli pubblicati, e offre già una prima collezione di postprint full text in PDF. L'editore e la redazione mirano a rendere disponibile la versione PDF di un selezionato numero di articoli pubblicati dal 1991 ad oggi, privilegiando quelli di carattere metodologico e quelli ritenuti più significativi dal Comitato Scientifico della rivista, anche in considerazione della rapida obsolescenza di quelli a carattere più specificata-

mente applicativo. La rivista dovrebbe così assumere una politica di post-print elettronico delayed della versione cartacea, che resterà comunque centrale nelle strategie editoriali della redazione.

Il repository di «Archeologia e Calcolatori» è già sottoposto ad harvesting dai service provider di OAI: in particolare segnaliamo 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.

#### 4. LA REALIZZAZIONE DEL REPOSITORY OAI DI «ARCHEOLOGIA E CALCOLATORI»

##### 4.1 Dublin Core Iniziative

Il protocollo OAI-PMH usa lo schema di metadati Dublin Core (<http://dublincore.org/>), che è nato in ambito bibliotecario ed è gestito dalla 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 loro ripetibilità, risorse documentarie di ogni genere. 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 (BONINCONTRO 2001; BARCHESI, CECCARELLI 2004) – lo schema Dublin Core appare perfettamente adeguato per la creazione di metadati descrittivi efficaci, è facile da implementare (HILLMANN 2003), chiaro da comprendere e adottato da OAI-PMH come schema *mandatory* del protocollo.

Nella nostra applicazione i metadati degli articoli pubblicati sono allocati negli elementi Dublin Core secondo la schema illustrato nella Tab. 1.

##### 4.2 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 e un database può “mapparli” su tabelle diverse, legate da relazioni.

Nella nostra applicazione lo schema relazionale è molto semplice, perché la tipologia omogenea dei dati (schede bibliografiche) non richiede una ripetizione di elementi tale da suggerire uno schema complesso e normalizzato. 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,

Elemento DC	Descrizione DC	Istanze	Uso in A&C	Formato
<b>Creator</b>	La persona o l'organizzazione che ha la responsabilità principale della creazione del contenuto intellettuale della risorsa	Si (senza limite)	Autore/i dell'articolo	
<b>Title</b>	Il nome dato alla risorsa dal Creator	No	Titolo dell'articolo	
<b>Subject</b>	Il soggetto può essere espresso da parole chiave o frasi che descrivono il contenuto della risorsa	Si (2 istanze)	Catalogazione per tipologia I istanza: classificazione informatica II istanza: classificazione archeologica	
<b>Relation</b>	Lo scopo di questo elemento è di fornire un mezzo per esprimere relazioni formali fra risorse	Si (2 istanze)	I istanza: utilizzato nel caso di articoli presentati a convegni i cui atti sono stati pubblicati dalla rivista. Registra i riferimenti cronotopografici e onomastici dell'evento. II istanza: ISBN della rivista	
<b>Source</b>	Una sequenza di caratteri alfabetici o numerici, usata per identificare univocamente l'opera dalla quale è derivata la risorsa	No	Numero della rivista (esempio: "Archeologia e Calcolatori, n. 15-2005")	
<b>Date</b>	La data in cui la risorsa è stata resa disponibile nella sua forma presente	No	Data di pubblicazione	Formato: ISO 8601 (YYYY-MM-DD)
<b>Format</b>	Il formato della risorsa, usato per identificare il software e l'eventuale hardware che potrebbe essere necessario per visualizzare o elaborare la risorsa	No	Formato in cui si presenta la risorsa: (PDF)	MIME <a href="http://www.iana.org/assignments/media-types/">http://www.iana.org/assignments/media-types/</a>
<b>Type</b>	La categoria della risorsa, ad esempio internet page, testo, immagine, video	No	Tipo di documento: (Text)	DCMI Type Vocabulary <a href="http://dublincore.org/documents/2004/06/14/dcmi-type-vocabulary">http://dublincore.org/documents/2004/06/14/dcmi-type-vocabulary</a>
<b>Contributor</b>	La persona o l'ente che ha dato significativi contributi intellettuali alla creazione della risorsa, ma il cui contributo è secondario rispetto ad un autore	No	Curatore o responsabile di edizione che apporta contributi critici al testo (esempio: il curatore dell'edizione degli atti di un convegno)	
<b>Publisher</b>	La persona o l'ente responsabile della produzione della risorsa, disponibile nella sua forma presente, come ad esempio una casa editrice, un dipartimento universitario o un ente	No	Editore di A&C: "Edizioni all'Insegna del Giglio-Firenze"	
<b>Language</b>	La/le lingua/e del contenuto intellettuale della risorsa. Il valore di questo campo deve coincidere con quelli definiti dall'RFC 1766	No	Lingua in cui è scritto l'articolo (A&C è una rivista che supporta il multilinguismo)	ISO639-1
<b>Identifier</b>	Una sequenza di caratteri alfabetici o numerici usata per identificare univocamente la risorsa	No	URL della risorsa PDF accessibile dalla rete	
<b>Coverage</b>	Le caratteristiche spaziali e/o temporali della risorsa	Si (2 istanze)	Riferimenti topografici e cronologici del contesto archeologico. I istanza: ambito topografico II istanza: contesto cronologico	
<b>Description</b>	Una descrizione in forma testuale del contenuto della risorsa	Si (2 istanze)	I istanza: contiene l'abstract in inglese dell'articolo. II istanza: numeri di pagina	
<b>Rights</b>	Copyright		Non usato	

Tabella n.1 – Uso degli elementi Dublin Core nei metadati di «Archeologia e Calcolatori».



del quale sono previste 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 il codice ISO639-1: <http://www.loc.gov/standards/iso639-2/englangn.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.

#### 4.3 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> (Fig. 2). Il file deve contenere sia i metadati, sia le informazioni di supporto necessarie all'harvesting sulla base del protocollo OAI-PMH.

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 (Fig. 3). Lo script è eseguito a richiesta, premendo un bottone presente sul form di gestione principale del database 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 (Fig. 4) è salvato con l'estensione.xml ed è well-formed e valido secondo lo schema emesso dalla OAI; la verifica può essere compiuta con *XMLspy*, il software più diffuso per lo sviluppo di applicazioni XML.

Se il file è salvato direttamente in una cartella del sito web istituzionale, esso assume immediatamente anche uno specifico indirizzo statico URL (definito Static Repository URL). Per trasformare questo file in un archivio OAI è necessario affidarsi all'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 (?initiate) al suo indirizzo URL (definito Static Repository Gateway URL) nel quale sia specificato l'indirizzo URL del repository da gestire:





Fig. 4 – Il database è esportato in XML.

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 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 entra ufficialmente in OAI come data provider e si è pronti a richiedere l'harvesting ai service provider sui quali si vuole comparire (è sufficiente inviare una e-mail agli amministratori). Più semplicemente si può aspettare che essi aggiornino le loro liste di harvesting sulla base di quelle ufficiali dei data provider OAI. La Fig. 5 mostra i metadati

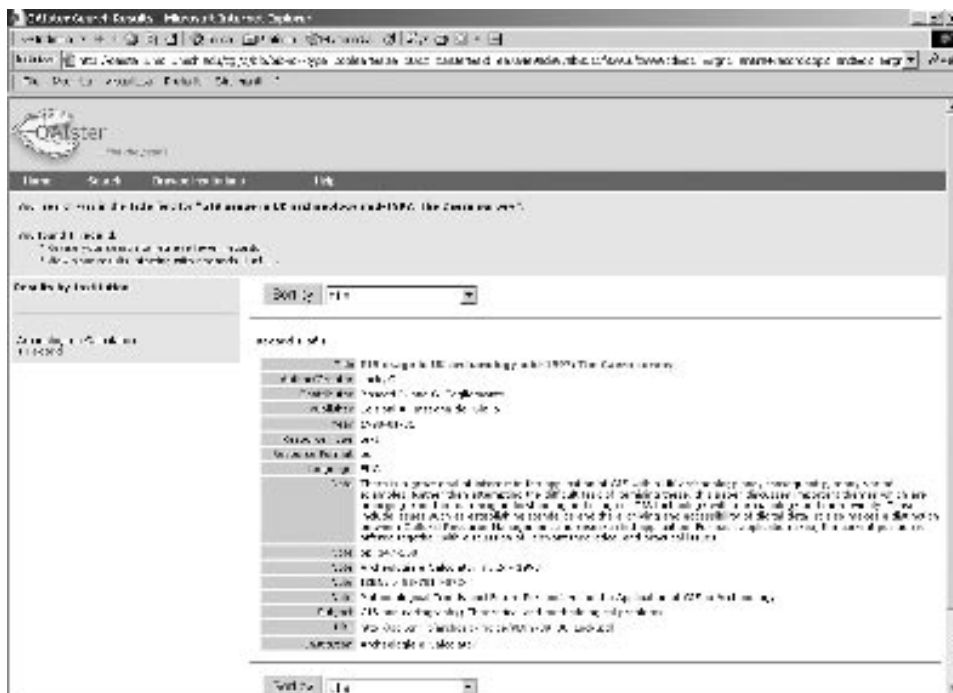


Fig. 5 – I metadati di «Archeologia e Calcolatori» sono presenti in OAISTER.

di «Archeologia e Calcolatori» così come appaiono in OAISTER (<http://oai-ster.umdl.umich.edu/o/oaister/>), il più grande servizio di ricerca cross-archive OAI del web.

## 5. 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 lo spider di Google ed è in grado di fornire sul sito web della rivista URL persistenti per tutti i record contenuti nel repository.

### 5.1 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 hyper-

link, con un indirizzo persistente. È stata richiesta l'indicizzazione a *Scholar Google*, la sezione "scientifica" di Google, che indicizza gli articoli scientifici e che ha mostrato interesse nel sostenere il progetto.

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 della Dublin Core Initiative (KUNZE 1999). Ogni record ha dunque un URL persistente, raggiungibile in soli tre iperlink dalla 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 (<http://arc.cs.ou.edu:8080/dp9/index.jsp>; LIU 2002), un progetto pensato per permettere a Googlebot l'indicizzazione di repository 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.

## 5.2 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: 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. È possibile cercare per autore, soggetto, anno e testo libero contenuto nell'abstract. Data la natura dei documenti gestiti, che si riferiscono spesso a precisi contesti storico-archeologici, è prevista, e sarà presto implementata, la possibilità di fare ricerche in base a un filtro topografico e cronologico, per il quale sono in corso di definizione adeguati vocabolari. 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).

Nella Fig. 7 è rappresentato l'intero progetto, definito complessivamente *OAI sistema* (BARCHESI 2005). 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.

## 6. CONCLUSIONI

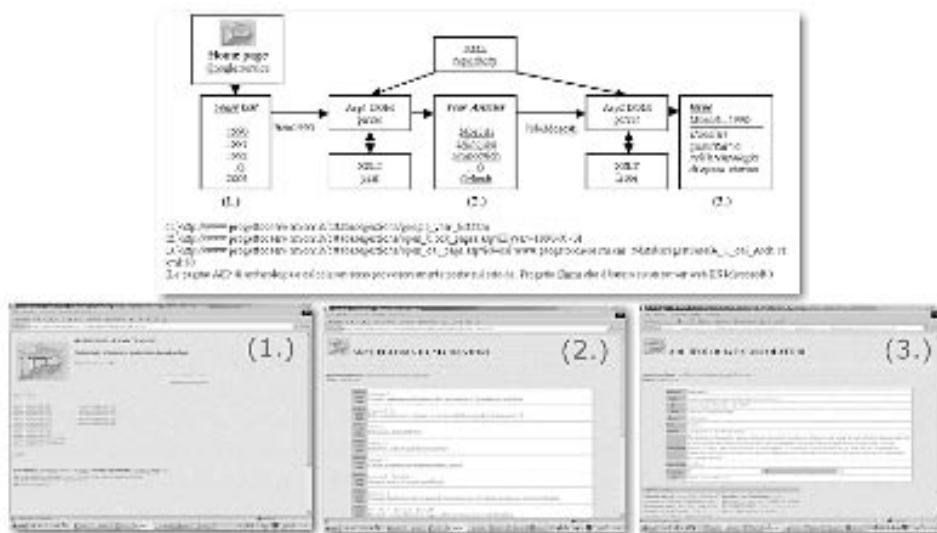


Fig. 6 – Offrire a Google i propri record, una soluzione XML-DOM-XSLT.

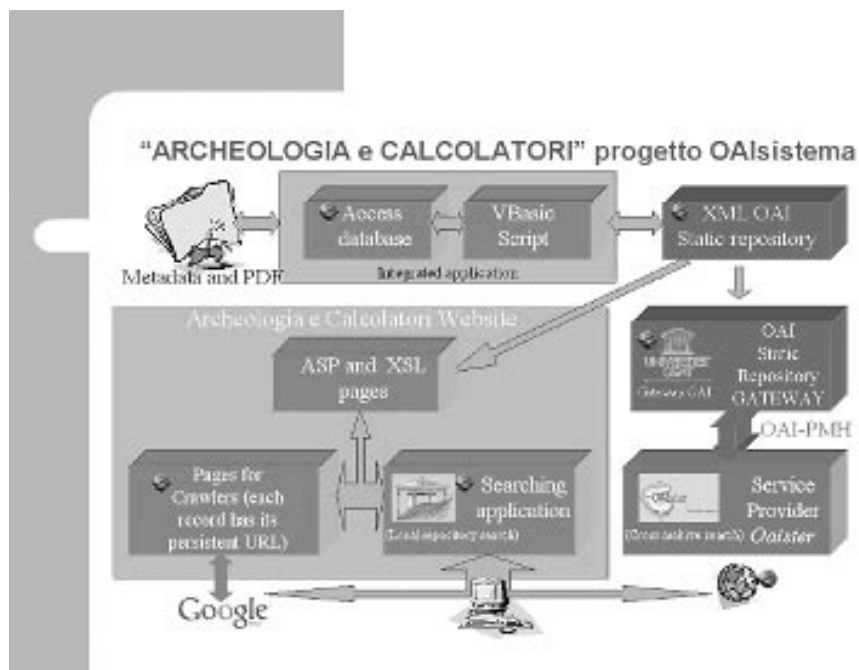


Fig. 7 – Lo schema complessivo del progetto OAI sistema.

Un gran numero di giornali elettronici e di collezioni di articoli digitali sono disponibili oggi sul web. Tra questi, solo pochi si organizzano per permettere l'interoperabilità degli archivi. In attesa degli sviluppi del Semantic Web, appare invece sempre più necessario ricorrere a tecnologie in grado di descrivere e rendere rintracciabili le risorse web.

Il protocollo OAI-PMH è sempre più diffuso e fornisce la normalizzazione necessaria per poter collegare riviste elettroniche che utilizzano database e formati diversi in motori di ricerca centralizzati. Questa tecnologia può essere facilmente adottata per realizzare portali disciplinari, dotati di servizi di ricerca di documenti che operino su più archivi. Gli utenti del web potrebbero cercare le informazioni in database virtuali di risorse disciplinari (si pensi a qualcosa di simile al modello di Humbul Humanities Hub: <http://www.humbul.ac.uk>) dove tuttavia i contenuti non sarebbero prodotti da un servizio di redazione unico, ma da diversi soggetti indipendenti.

In quest'articolo è stata presentata una soluzione software per un repository di risorse digitali, che assicura l'interoperabilità con i service provider OAI ed è basata su un database Access, tecnologie Microsoft per il web dinamico e XML. In sintesi, il progetto è costituito da un database Access e da alcuni file XML ed XSL; le funzionalità OAI-PMH sono delegate a un server gateway esterno; il supporto ai motori di ricerca è fornito attraverso dei servizi specifici realizzati in ASP (*Active Server Pages*), che permettono l'indicizzazione dei contenuti del repository; il software comprende un motore di interrogazione locale da implementare nel proprio sito web. Il software, inoltre, può essere facilmente configurato da qualunque webmaster con minime conoscenze di XML e dell'ambiente Windows, così come all'utente compilatore delle schede bibliografiche non è richiesta nessuna competenza di web engineering.

Il sito Internet della rivista è in rete presso l'indirizzo <http://soi.cnr.it/archcalc>. All'insieme dei servizi descritti in questo articolo si accede seguendo il link "Bibliografia". Nell'ambito di una convenzione stipulata con l'Università degli Studi di Roma "Tor Vergata", è attualmente in corso di realizzazione, nell'ambito di uno stage formativo presso l'ISCIMA, un completo restyling del sito da parte di Riccardo Pasquini, che sarà sviluppato in aderenza alle norme emesse dal Ministero per i Beni e le Attività Culturali nell'ambito del Progetto MINERVA (FILIPPI 2004). Il nuovo sito introdurrà, quale ulteriore strumento per la diffusione dei contenuti, il feed RSS di ogni nuovo numero della rivista.

CLAUDIO BARCHESI  
Istituto di Studi sulle Civiltà Italiane  
e del Mediterraneo Antico  
CNR – Roma

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, 2004)* (<http://www.uiniv.trieste.it/~storia/docenti/Abbattista/editoria/editoria.htm>, ultimo accesso 15/4/2005).
- BARCHESE C. 2001, *Progetto Caere: un'applicazione Internet attiva per l'Information Retrieval di documenti SGML*, «Archeologia e Calcolatori», 12, 170-189.
- BARCHESE C. 2005, *OAI sistema: A simple solution for implementing OAI-PMH on small size archives*, in *Proceedings of CAA2005 (Tomar 2005)*, in corso di stampa.
- BARCHESE C., CECCARELLI L. 2004, *Linguaggi dichiarativi per la ricerca archeologica*, in P. MOSCATI (ed.), *Nuove frontiere della ricerca archeologica. Linguaggi, Comunicazione, Informatica*, «Archeologia e Calcolatori», 15, 95-114.
- BERNERS-LEE T. 1998, *Semantic Web Road Map* (<http://www.w3.org/DesignIssues/Semantic>, ultimo accesso 01/08/2005).
- BONINCONTRO I. 2001, *Progetto Caere: prospettive di applicazione degli standard internazionali per la codifica dei dati testuali*, «Archeologia e Calcolatori», 12, 55-69.
- CANTONE F. 2005, *Shared technologies in archeologia: nuove prospettive di gestione e condivisione di dati in rete*, «Archeologia e Calcolatori», 16, 2005, [????-???](#).
- FILIPPI F. 2004, *Manuale per la qualità dei siti Web pubblici culturali*, Roma, Ministero per i Beni e le Attività Culturali, "Progetto MINERVA".
- 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), ultimo accesso 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>, ultimo accesso 15/4/2005).
- HILLMANN D. 2003, *Using Dublin Core* (<http://dublincore.org/documents/usageguide>, ultimo accesso 15/4/2005).
- HUSSEIN S. 2001, *Using the repository explorer to achieve OAI protocol compliance*, in *Joint Conference on Digital Libraries (JCDL) '01 (Roanoke, Virginia, 2001)*, New York, ACM press, 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), ultimo accesso 15/4/2005).
- KUNZE J. 1999, *Encoding Dublin Core in HTML*, IETF RFC (2731) (<http://www.ietf.org/rfc/rfc2731.txt>, ultimo accesso 15/4/2005).
- LAGOZE C., VAN DE SOMPEL H. 2001, *The Open Archives Initiative: Building a low-barrier interoperability framework*, in *Joint Conference on Digital Libraries (JCDL) '01 (Roanoke, Virginia, 2001)*, New York, ACM press, 54-62 (<http://www.openarchives.org/documents/jcdl2001-oai.pdf>, ultimo accesso 15/4/2005).
- LIU X., K.MALY, M.ZUBAIR, M.L.NELSON 2002, *DP9: an OAI gateway service for Web Crawlers*, in *Joint Conference on Digital Libraries (JCDL) '02 (Portland, Oregon, 2002)*, ([http://www.cs.odu.edu/~liu\\_x/dp9/dp9.pdf](http://www.cs.odu.edu/~liu_x/dp9/dp9.pdf), ultimo accesso 15/4/2005).
- LYNCH C. A. 2001, *Metadata harvesting and Open Archives Initiative*, ARL Bimonthly Report 217, Washington D.C., 1-9 (<http://www.arl.org/newsltr/217/mhp.html>, ultimo accesso 15/4/2005).
- MOSCATI P. 1999, *Archeologia e Calcolatori: dieci anni di contributi all'informatica archeologica*, «Archeologia e Calcolatori», 10, 343-352.
- MOSCATI P. 2002, *Archeologia informatica: tra tradizione e rinnovamento*, "Bollettino ICR" n.s. 5, 21-27.
- NEE E. 2005, *Web future is not Semantic, or overly orderly*, «CIO Insight» (<http://www.cioinsight.com/article2/0,1540,1815338,00.asp>, ultimo accesso 15/4/2005).



«Archeologia e Calcolatori»: nuove strategie per la diffusione di contenuti in rete  
sulla base dell'OAI-PMH

- PAPADOPOULOS D.C., MAVRIKAS E.C. 2004, *Peer-to-Peer ways to Cultural Heritage*, in AA.VV., *Enter the Past. Proceedings of CAA2003 conference (Vienna 2003)*, Bar International Series 1227, Oxford, Archaeopress, 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, 7-19 (<http://www.bibliotecheoggi.it/2004/20041000701.pdf>, ultimo accesso 15/4/2005).
- VAN DE SOMPEL H., LAGOZE C. 2000, *The Santa Fe Convention of the Open Archives Initiative*, «D-Lib Magazine» (<http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>, ultimo accesso 15/4/2005).