

4th Open Archives Forum Workshop:

'In Practice, Good Practice: the Future of Open Archives'

4th-5th September 2003, University of Bath, Bath, UK

Susanna Mornati - CILEA

Il quarto *workshop*¹ dell'Open Archives Forum (OAF)², ospitato nella bella sede dell'Università di Bath³, è stato aperto da **Rachel Heery**, responsabile del settore Ricerca e Sviluppo di UKOLN⁴, che ha dato il benvenuto ai partecipanti (oltre 70, provenienti per la maggior parte da paesi Europei e in particolare dal Regno Unito) e ha introdotto le prime due relazioni riguardanti iniziative dell'OAF.

Birgit Matthaei, dall'Università di Humboldt, Berlino, ha presentato i contenuti dell'Information Resource Database⁵, che ha l'intento di censire progetti e iniziative collegate allo sviluppo di Open Archives in Europa. Matthaei ha sottolineato come un fatto positivo la crescita, anche se lenta, del numero dei siti, sia Data Provider sia Service Provider, non solo in ambito accademico ma anche museale. In particolare sono aumentate le installazioni che utilizzano *software* condiviso rispetto a quelle che impiegano *software* sviluppato in proprio (il 50% nel 2003, contro l'80% nel 2002), fatto positivo per la cooperazione, come pure si evidenzia l'evoluzione dei metadati dal Dublin Core *simple* al *qualified*, verso MARC21 e UNIMARC. Si rileva inoltre come lo sforzo di normalizzazione dei dati risieda ancora in gran parte nei Service Providers, mentre sarebbe auspicabile l'estendersi del modello tedesco (con accordi sui formati a livello dei Data Provider). Un modello intermedio potrebbe essere rappresentato dallo sviluppo di *middleware* condiviso. La relazione pone in evidenza l'importanza dei prodotti dell'Open Archives Initiative (OAI)⁶ in termini di scambio di dati, recupero più efficiente e servizi aggiuntivi.

La seconda relazione, presentata da **Leona Carpenter**, ora *free-lancer* dopo aver fatto parte dello staff di UKOLN, ha illustrato i contenuti dell'OAF Online Tutorial⁷, uno strumento introduttivo per chi desidera implementare un *database* di risorse elettroniche che adotti lo standard OAI-PMH (Protocol for Metadata Harvesting)⁸. Lo strumento è nato sulla scorta di precedenti iniziative didattiche tenutesi in concomitanza dei precedenti *workshop* OAF, e nasce dall'esigenza di fornire una solida base e un'opportunità formativa anche a chi si trova all'inizio del percorso. Il materiale incluso è molto ricco e varia da un'introduzione all'OAI alla storia del movimento, dalle specifiche tecniche strutturali a quelle relative all'implementazione del protocollo standard, da Dublin Core a XML. Contiene infine un glossario della terminologia indispensabile.

Nella seconda parte della mattinata i partecipanti si sono divisi fra tre distinte **breakout sessions**, che hanno approfondito le tematiche delle tesi digitali, della qualità e della sostenibilità degli Open Archives (OA). La **prima** ha approfondito le questioni aperte legate all'esigenza di definire formati e metadati condivisi, e di fornire strumenti per la gestione del copyright e la conservazione dei dati. La **seconda** ha sottolineato l'importanza di fornire servizi a valore aggiunto senza limitare i Service

¹ http://www.oaforum.org/workshops/bath_invitation.php

² <http://www.oaforum.org/>

³ <http://www.bath.ac.uk/>

⁴ <http://www.ukoln.ac.uk/>

⁵ http://www.oaforum.org/oaif_db/index.php

⁶ <http://www.openarchives.org/>

⁷ <http://www.ukoln.ac.uk/metadata/oa-forum/tutorial/>, indirizzo provvisorio in attesa di essere reso ufficialmente disponibile sul sito dell'OAF.

⁸ <http://www.openarchives.org/OAI/openarchivesprotocol.html>

4th Open Archives Forum Workshop:

'In Practice, Good Practice: the Future of Open Archives'

4th-5th September 2003, University of Bath, Bath, UK

Susanna Mornati - CILEA

Provider alla semplice raccolta (*harvesting*) dei metadati. La **terza** ha invocato il coinvolgimento delle biblioteche (ad esempio per praticare il cosiddetto *self-archiving-by-proxy*), lo sviluppo di servizi attraenti (statistiche di accesso ai documenti, estrazione di dati per pagine web individuali) per garantire una più rapida diffusione delle iniziative in ambito OA, ed auspicato una cooperazione a livello europeo sia per la soluzione dei problemi meno sentiti negli ambienti degli sviluppatori (quali il multilinguismo e l'accesso al patrimonio storico e culturale comune) sia per la possibilità di accedere a finanziamenti dell'Unione Europea per il sostegno dei progetti.

La sessione pomeridiana è stata aperta da **Timothy Cole**, dall'Università dell'Illinois di Urbana-Champaign, con una relazione sui problemi incontrati e le soluzioni adottate per la costruzione di un portale⁹ di accesso al patrimonio culturale dell'Illinois, focalizzata in particolare sui problemi derivanti dalla varietà degli oggetti rappresentati e dei metadati associati, nonché dalle ambiguità derivanti dal livello di rappresentazione, che in un'unica descrizione può racchiudere un'intera collezione o un singolo oggetto, descrivere l'oggetto "analogico" o una sua rappresentazione "digitale". Fra le soluzioni indicate Cole ha dedicato ampio spazio ai *gateway* per il *caching* dei depositi statici¹⁰, che forniscono direttamente *file* XML alle applicazioni, semplificando la raccolta dei dati stabili e provenienti da piccole istituzioni con problemi di accesso ai dati.

Muriel Foulonneau, consulente per Relais Culture Europe del Ministero della Cultura francese, ha presentato le tematiche affrontate nel settore dell'accesso al patrimonio culturale, sottolineando l'importanza e le difficoltà del lavoro sulla condivisione dei metadati e sulle politiche di erogazione dei servizi, dalla frequenza di raccolta dei dati alla granularità del livello di informazione, ai problemi di gestione posti da un servizio di aggregazione che non ha controllo sulla fonte delle informazioni, in particolare sul tasso di crescita, ma con conseguenze anche sulle scelte relative all'impiego di una determinata lingua, alla deduplicazione dei dati provenienti da archivi diversi, alla normalizzazione dei contenuti. Riutilizzando l'espressione "metadata gardening" (mutuata dal contesto della Mellon Metadata Harvesting Initiative) in contrapposizione al semplice "harvesting", Foulonneau ha posto l'accento sulle responsabilità dei Data Provider, e sulle difficoltà che i servizi *downstream* (ad esempio il controllo di autorità) incontrano nell'applicazione ad aggregazioni di dati disomogenei.

E' stata poi la volta di **Elizabeth Gadd**, dall'Università di Loughborough, che ha fatto il punto sul progetto RoMEO¹¹, Rights METadata for Open Archiving, appena concluso, che ha investigato le questioni aperte in merito alla protezione del diritto d'autore dei lavori depositati negli OA e dei relativi metadati. Il progetto si è articolato in una serie di indagini volte a determinare la percezione degli autori e le intenzioni degli editori in merito alla cessione del copyright, nonché le politiche dei depositi e dei servizi di aggregazione di dati, e nella produzione di una serie di studi e una relazione finale, disponibili sul sito del progetto. I risultati ottenuti descrivono la situazione più volte denunciata nell'ambiente OA, ossia la penalizzazione della diffusione delle ricerche,

⁹ <http://oai.grainger.uiuc.edu/>

¹⁰ Una versione alfa delle linee guida è già disponibile sul sito OAI: <http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>

¹¹ <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>

4th Open Archives Forum Workshop:

'In Practice, Good Practice: the Future of Open Archives'

4th-5th September 2003, University of Bath, Bath, UK

Susanna Mornati - CILEA

sacrificata alla protezione del copyright ceduto all'editore nel 90% dei casi, ed in maniera perlopiù inconsapevole da parte degli autori. L'auspicio è una maggiore pressione istituzionale nei confronti degli editori perché gli autori siano autorizzati a ritenere i propri diritti sulle opere pubblicate, incluso il deposito sugli archivi istituzionali (ora concesso solo nel 49% dei casi).

Ha chiuso la prima giornata una vivace relazione di **Carl Lagoze**, dalla Cornell University, che sul tema "dove siamo e dove andiamo?" ha sottolineato il ruolo dello standard OAI-PMH nella diffusione degli archivi e ne ha dichiarato la stabilità. Con questo strumento, l'interoperabilità (anni fa basata sulla federazione di risorse che adottavano protocolli descrittivi quali MARC) si situa oggi a livello di aggregazione (gli *harvester*) nella corsa verso un futuro di raccolta (*gathering*) operata dai *web crawlers*. Lagoze lancia una provocazione: i metadati sono troppo costosi da produrre¹², meglio affidarci a Google™? La conclusione è che occorrono sia buoni metadati, in grado anche di evidenziare le relazioni fra le differenti risorse, sia buoni motori di ricerca che li utilizzino: il futuro non sono i Service Provider bensì i portali evoluti che espongono contenuti in forma personalizzata.

La seconda giornata si è aperta con la presentazione di Theses Alive!¹³, un progetto per l'accesso alle dissertazioni prodotte dalle università del Regno Unito, che prevede l'adozione di DSpace (MIT) per la costituzione del deposito e lo sviluppo di software per la gestione del *workflow* legato ai processi di valutazione dei lavori. Il progetto, condotto dall'Università di Edinburgo, metterà a disposizione la piattaforma per il deposito entro la fine del 2003, e la piattaforma per l'invio e la valutazione delle tesi da discutere entro la primavera 2004. **Theo Andrew** ha illustrato le esigenze espresse e le soluzioni proposte, **Richard Jones** gli aspetti tecnici, dalla scelta di DSpace allo sviluppo di *software* aggiuntivo (disponibile sul sito) per l'invio e la supervisione dei lavori, per facilitare la comunicazione fra docenti e studenti, per gestire l'esame e la correzione dei lavori, per l'amministrazione del sistema, nonché funzionalità di *crosswalk* dei metadati per l'esportazione e lo scambio.

E' stata poi la volta di **Rasik Pandey**, sviluppatore di AJLSM, una ditta di consulenza che ha sviluppato un portale dedicato al patrimonio culturale dell'Aquitania e una piattaforma per le pubblicazioni elettroniche. Il portale¹⁴ integra le funzionalità del sistema francese SDX per la raccolta e il deposito dei metadati con altri pacchetti per la conversione, integrazione, normalizzazione e indicizzazione, e prevede un arricchimento dei metadati (Dublin Core +7) per consentire lo sviluppo di servizi specifici, ad esempio funzionalità di ricerca che prevedono anche un accesso cartografico tramite mappe attive. L'adozione di OAI-PMH e di SDX ha facilitato la costruzione del portale, consentendo l'integrazione di risorse diverse e la costruzione di servizi aggiuntivi. La piattaforma per le pubblicazioni elettroniche, Cyberdocs¹⁵, impiega *software open source* per la conversione di documenti prodotti da *word processor* nel formato XML TEI Lite, e fornisce moduli di gestione via web, e

¹² <http://www.well.com/~doctorow/metacrap.htm>

¹³ <http://www.thesesalive.ac.uk/>

¹⁴ <http://ajlsm-sdx.hopto.org/sdx-22h/pa-portail/>

¹⁵ <http://sourcesup.cru.fr/cybertheses/>

4th Open Archives Forum Workshop:

'In Practice, Good Practice: the Future of Open Archives'

4th-5th September 2003, University of Bath, Bath, UK

Susanna Mornati - CILEA

applicazioni dinamiche per la stampa e la pubblicazione di HTML statico e dinamico e PDF, con la possibilità di aggiungere metadati in formato OAI-PMH.

L'intervento successivo è stato condotto da **Henk Ellermann** dell'Università Erasmus di Rotterdam, allo scopo di presentare il progetto DARE¹⁶, un'iniziativa che coinvolge tutti gli atenei olandesi nell'implementazione di depositi istituzionali per la ricerca, e che ha implicato un'attività condivisa di definizione delle specifiche. Mentre gli archivi dovrebbero essere pronti entro l'autunno 2003, sono stati approvati alcuni progetti collaterali, da concludere nel corso del 2004, per lo sviluppo di funzionalità aggiuntive, dalla gestione della *peer-review* e del *copyright* al *print-on-demand*, dalla pubblicazione di *proceedings* alla costituzione di un *database* nazionale delle tesi. Ellermann ha sottolineato l'aspetto cooperativo dell'iniziativa e la necessità di coinvolgere docenti e bibliotecari attraverso attività estese di informazione e promozione.

Durante il coffee break viene presentata, fuori programma, l'iniziativa delle accademie belghe di area fiamminga per la costituzione di un deposito dedicato alle tesi di laurea. **Ivo Houbrechts**, della Katholieke Hogeschool Kempen, presenta le ragioni per lo sviluppo autonomo di un prototipo¹⁷, basato su *software open source* e Java, che consente di adottare set differenziati di metadati, di importare i dati anagrafici degli studenti dagli archivi delle istituzioni e di gestire l'invio e la revisione dei documenti. Entro la metà del 2004 l'installazione verrà resa compatibile con OAI-PMH.

La densa mattinata di presentazioni continua con una relazione a tre voci, dedicata alle problematiche di sviluppo di Ingenious¹⁸, un knowledge site in campo museale. **Ann Borda, Alpay Beler e Nick Wyatt** del Science Museum di Londra illustrano gli scopi del sito, i problemi affrontati, le questioni tecniche determinate dalla necessità di accoppiare metadati provenienti da sistemi che adottano standard descrittivi e di comunicazione diversi, nonché le soluzioni adottate per la catalogazione e il data mapping. Un accento particolare viene posto sulle Collection Level Descriptions (CLDs), che costituiscono una parte importante dei metadati del sito, formalizzandone la struttura editoriale e garantendo estese possibilità di ricerca. Il progetto sarà disponibile nel 2004, e pur non essendo compatibile OAI, l'impiego di XML e Dublin Core rende i dati uniformi e disponibili per lo scambio.

L'ultima presentazione di un progetto riguarda l'Art Museum dell'Università di Cambridge: "Harvesting the Fitzwilliam"¹⁹, illustrato da **Shaun Osborne**, nasce con lo scopo di fornire un catalogo elettronico unificato degli oggetti non solo per le esigenze di gestione della collezione e a fini didattici e di ricerca, ma anche come strumento qualificato per lo scambio in ambito OAI e *resource discovery*. L'implementazione come *data provider*, e l'allestimento da parte dello staff del museo di *record* descrittivi con l'impiego del protocollo OAI-PMH, renderà la collezione accessibile attraverso portali creati a partire dall'harvesting dei metadati disponibili nel settore artistico, umanistico ed archeologico. Si pone il problema di proteggere anche i diritti d'autore

¹⁶ <http://www.surf.nl/>, link: DARE

¹⁷ <http://doks.khk.be/>

¹⁸ <http://www.nmsi.ac.uk/nmsipages/nofdigitise.asp>

¹⁹ <http://www.fitzmuseum.cam.ac.uk/htf>

4th Open Archives Forum Workshop:
'In Practice, Good Practice: the Future of Open Archives'
4th-5th September 2003, University of Bath, Bath, UK

Susanna Mornati - CILEA

relativi ai metadati, il cui grado di unicità e ricchezza diviene elevato nel contesto museale.

La mattinata si conclude con due relazioni di argomento più generale. La prima, presentata da **Paul Child** dell'Università dell'East Anglia, anticipa i contenuti del secondo rapporto²⁰ sulle questioni organizzative previste dall'OAF. Il documento, dedicato al supporto delle istituzioni europee che intraprendono iniziative caratterizzate dall'approccio OA, ne prende in esame le varie implicazioni organizzative, dal *business model* alle questioni legate al diritto d'autore, dalla qualità (e l'attribuzione della relativa responsabilità ai vari *stakeholder*) ai metadati (a proposito dei quali si invocano le linee guida prodotte da UKOLN²¹). Una sezione è dedicata al tema dell'interoperabilità dei soggetti descrittivi e delle classificazioni, dove si sottolinea l'importanza dell'armonizzazione e della mappatura fra i diversi schemi, un'altra all'esigenza di dotarsi di sistemi di *content management* (CMS) e di pianificazione.

Infine, **David Prosser**, direttore di SPARC Europe²², presenta una carrellata delle argomentazioni a favore delle iniziative legate agli Open Archives e all'Open Access, dei problemi aperti e delle azioni da intraprendere, sostenendo la validità del modello dei depositi istituzionali affiancati a riviste *open access* per garantire la continuità del processo di validazione della ricerca.

Una breve sessione pomeridiana, dedicata alla discussione sul futuro degli Open Archives, ha chiuso il convegno.

²⁰ <http://www.oaforum.org/documents/>

²¹ <http://www.rdn.ac.uk/projects/eprints-uk/docs/simpledc-guidelines/>

²² <http://www.sparceurope.org/>