

Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

Open Archives in Italia: una piattaforma nazionale

Susanna Mornati* (mornati@cilea.it)

*CILEA – Consorzio Interuniversitario per le Tecnologie
dell'Informazione e della Comunicazione, Segrate (Milano)

Abstract

Le collezioni digitali rischiano di restare materiale isolato e non integrato nell'offerta documentaria delle istituzioni che le possiedono, se non vengono messe in atto strategie di integrazione per la costituzione di vere e proprie "biblioteche digitali", che offrano un accesso sinergico alle varie tipologie di materiale che oramai differenziano la vasta costellazione della documentazione disponibile in formato elettronico. La relazione illustra un progetto in corso di realizzazione al CILEA per la costituzione di una piattaforma nazionale di accesso agli *Open Archives*, una delle tipologie di collezione digitale di più recente diffusione. Mentre l'architettura si basa su tecnologie già esistenti e collaudate, restano aperti i problemi relativi alle politiche di condivisione dei documenti, al formato dei metadati, alla protezione del diritto d'autore. Per la soluzione di questi problemi è necessaria una iniziativa di forte cooperazione nazionale, alla quale collaborino tutti gli *stakeholder*, dai ricercatori agli amministratori, agli editori. Ma l'invito è soprattutto rivolto ai bibliotecari, in qualità di maggiori esperti delle tematiche più complesse che si agitano nel mondo dei documenti digitali.

1. Introduzione¹

Le collezioni digitali rischiano di restare materiale isolato e non integrato nell'offerta documentaria delle istituzioni che le possiedono, se non vengono messe in atto strategie di integrazione per la costituzione di vere e proprie "biblioteche digitali", che offrano un accesso sinergico alle varie tipologie di materiale che oramai differenziano la vasta costellazione della documentazione disponibile in formato elettronico.

L'esperienza in corso di maturazione sugli *open archives* (OA) nel mondo dimostra come non sia sufficiente predisporre depositi aperti per indurre la comunità scientifica ad utilizzarli. Non tratterò in questa sede delle problematiche legate alle modifiche in atto nelle modalità della comunicazione scientifica, con il relativo corollario della crisi dei prezzi dei periodici, degli atteggiamenti dei ricercatori rispetto all'open access - inclusa la riluttanza ad abbandonare le pratiche tradizionali - della valutazione dell'impatto della ricerca e del referaggio. Riguardo a queste tematiche la letteratura esistente è già sufficientemente ampia ed esaustiva².

¹ I link citati in questo documento sono stati visitati per l'ultima volta il 14 novembre 2003.

² Fra i contributi italiani, si vedano contributi di Valentina Comba, Antonella De Robbio, Roberto Delle Donne, Antonio Fantoni, Paola Gargiulo, Antonio Morselli Labate, Renato Spigler negli atti del convegno "Comunicazione

Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

Cercherò invece di soffermarmi sulla necessità di arricchire gli strumenti esistenti per il *self-archiving* ed integrarli con le altre risorse digitali, allo scopo di aumentarne l'impiego da parte degli autori di letteratura scientifica, a beneficio loro, delle istituzioni che li finanziano e di chi si occupa della gestione dei servizi informativi a loro necessari per la conduzione delle attività di ricerca: mi riferisco in particolare ai bibliotecari delle università e degli enti di ricerca, ed al ruolo che possono assumere nello scenario delle nuove modalità della comunicazione scientifica³.

Infatti, nonostante l'esistenza di una ricca messe di strumenti *software open source* per la costituzione di depositi aperti, trovare un documento in *full-text* in rete è ancora un'impresa accessibile quasi esclusivamente agli addetti ai lavori⁴. Talora la comparsa di nuovi strumenti non fa che aumentare la confusione dei lettori, avallando posizioni estremiste, dalla "*information illiteracy*" degli oltranzisti della carta, fino a coloro che impiegano solo Google per trovare qualunque informazione.

Cosa occorre perché gli OA si diffondano veramente, perché se ne diffonda la cultura e l'uso, perché diventino veramente un'alternativa per garantire l'accesso aperto alla letteratura scientifica prodotta dagli atenei e dagli enti di ricerca? Quale spinta verrà da decisioni drastiche come - notizia di questi giorni - la chiusura degli abbonamenti alle riviste edita da Elsevier annunciata dalla biblioteca della Cornell University⁵? Quale forza vincerà la riluttanza di autori ed istituzioni ad abbandonare il confortevole ma ingannevole sistema di valutazione dell'Impact Factor⁶ imposto dalle multinazionali dell'editoria STM, con il suo carico di conseguenze (prevedibili e controllabili dall'establishment della ricerca) sulle carriere e sui finanziamenti, per esplorare gli ignoti lidi della misurazione dei *download*⁷?

scientifica ed editoria elettronica: la parola agli Autori: L'Utente-Autore nel circuito della comunicazione scientifica: editoria elettronica e valutazione della ricerca", Milano, 20 Maggio 2003, URL:

<http://www.cilea.it/convegni/convegnoeditoria/presentazione.html>, in corso di pubblicazione, testi depositati nell'*open archive* E-LIS all'URL: <http://eprints.rclis.org/>.

³ Peter Suber, "Removing the Barriers to Research: an Introduction to Open Access for Librarians", in *College & Research Libraries News*, 64 (February 2003) pp. 92-94, 113, testo disponibile all'URL: http://www.ala.org/Content/NavigationMenu/ACRL/Publications/College_and_Research_Libraries_News/Back_Issues/2003/February1/Removing_barriers_to_research.htm.

⁴ Come ben rileva De Robbio in una recente comunicazione, molti degli strumenti disponibili per la costituzione di biblioteche digitali non colloquiano reciprocamente: Antonella De Robbio, "Le nuove tecnologie per l'accesso all'informazione: gli strumenti per la costruzione di biblioteche digitali", in *Proceedings La biblioteca digitale nelle università italiane: progetti e prospettive di sviluppo*, Roma, ottobre 2003, 50. Congresso nazionale AIB, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000393/>.

⁵ Le ragioni spiegate all'URL: <http://www.library.cornell.edu/scholarlycomm/>.

⁶ Si pensi solamente al rifiuto di ISI-Thomson di includere riviste *open access* nel calcolo dell'IF, falsandone di fatto i valori: testimonianza di A. Morselli Labate, cit.

⁷ Valentina Comba, "La valutazione delle pubblicazioni: dalla letteratura a stampa agli Open Archives", pre-print, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000095/>

Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

2. Strumenti e servizi

La disponibilità online ad accesso aperto dei risultati della ricerca scientifica ne aumenta visibilità ed impatto. Ma non basta. Occorrono nuovi strumenti e nuovi servizi⁸ che possano integrare e sostituire i modelli consolidati offerti dall'editoria commerciale, che ha costruito i propri profitti sulla gestione della letteratura scientifica e ne mantiene il controllo.

La comunità OAI produce senza sosta *software* nuovo ed aggiornato per offrire servizi basati sui documenti contenuti negli archivi aperti. La creatività non è ancora ripagata dalla quantità e gli archivi, soprattutto in Italia, non hanno ancora raggiunto la massa critica necessaria ad attrarre gli studiosi più delle pubblicazioni tradizionali, ma è in arrivo il punto di condensazione⁹. Per andare incontro alle nuove esigenze di modelli comunicativi aperti, il CILEA sta realizzando un progetto di piattaforma di servizi per la comunità nazionale, da concludersi entro la metà del 2004, che offra valore aggiunto agli archivi aperti e ai documenti depositati attraverso strumenti raffinati di ricerca, indicizzazione, analisi citazionale, integrazione con le altre risorse, cartacee o digitali, a disposizione degli studiosi. La piattaforma per gli Open Archives nasce nel contesto di AEPIC¹⁰, un progetto CILEA di editoria elettronica per la ricerca e la didattica rivolto agli atenei.

L'architettura della piattaforma, che si basa su strumenti recenti ma già collaudati dalla comunità OAI, è strutturata in due oggetti principali: un insieme di Service Provider e un portale. I metadati e, dove necessario, i documenti a testo pieno dei lavori scientifici, saranno raccolti dai Data Provider disponibili ed appropriati - possibilmente dedicati al deposito delle pubblicazioni accademiche e di ricerca prodotte a livello nazionale, sia in campo scientifico sia umanistico/giuridico - mediante l'impiego del protocollo OAI-PMH, quindi verranno sottoposti ad una serie di operazioni:

- attribuzione di una classificazione per soggetto a livello generale, basata sulla suddivisione in aree di ricerca operata dal Ministero¹¹;
- funzionalità centralizzate di caching e indicizzazione;
- funzionalità avanzate di ricerca, scorrimento e recupero delle informazioni;
- analisi citazionale operata nel testo del documento per la registrazione presso servizi di indicizzazione delle citazioni¹², nonché per l'estrazione automatica di OpenURL¹³, che possano essere utilizzati da servizi di *resolving* esistenti;
- allestimento di *crosswalk* per la conversione, ai fini dell'importazione/esportazione da/verso basi di dati che adottano differenti standard per i metadati;
- servizio di *gateway* per *web crawlers*;

⁸ Henk Ellermann e Hanna de Vries, "Setting up an Institutional Archive: Principles and Practice", in *Beyond the Network: Proceedings of the EUNIS 2003 conference*, testo disponibile all'URL: <http://web.eur.nl/ub/eeepi/pdf/Ellermann-deVries.pdf>.

⁹ Un'immagine derivata dalla fisica, da cui proviene anche quella di *momentum*: Peter Suber, Peter Suber, "Momentum for Eprint Archiving," Free Online Scholarship Newsletter, August 8, 2002, testo disponibile all'URL: <http://www.topica.com/lists/suber-fos/read/message.html?mid=1607391538&sort=d&start=38>.

¹⁰ Il progetto AEPIC e la piattaforma sono illustrati all'URL: <http://www.aepic.it/>.

¹¹ http://www.murst.it/atti/2000/alladm001004_01.htm.

¹² Ad esempio CiteSeer, noto anche come ResearchIndex, vedi: <http://citeseer.nj.nec.com/>

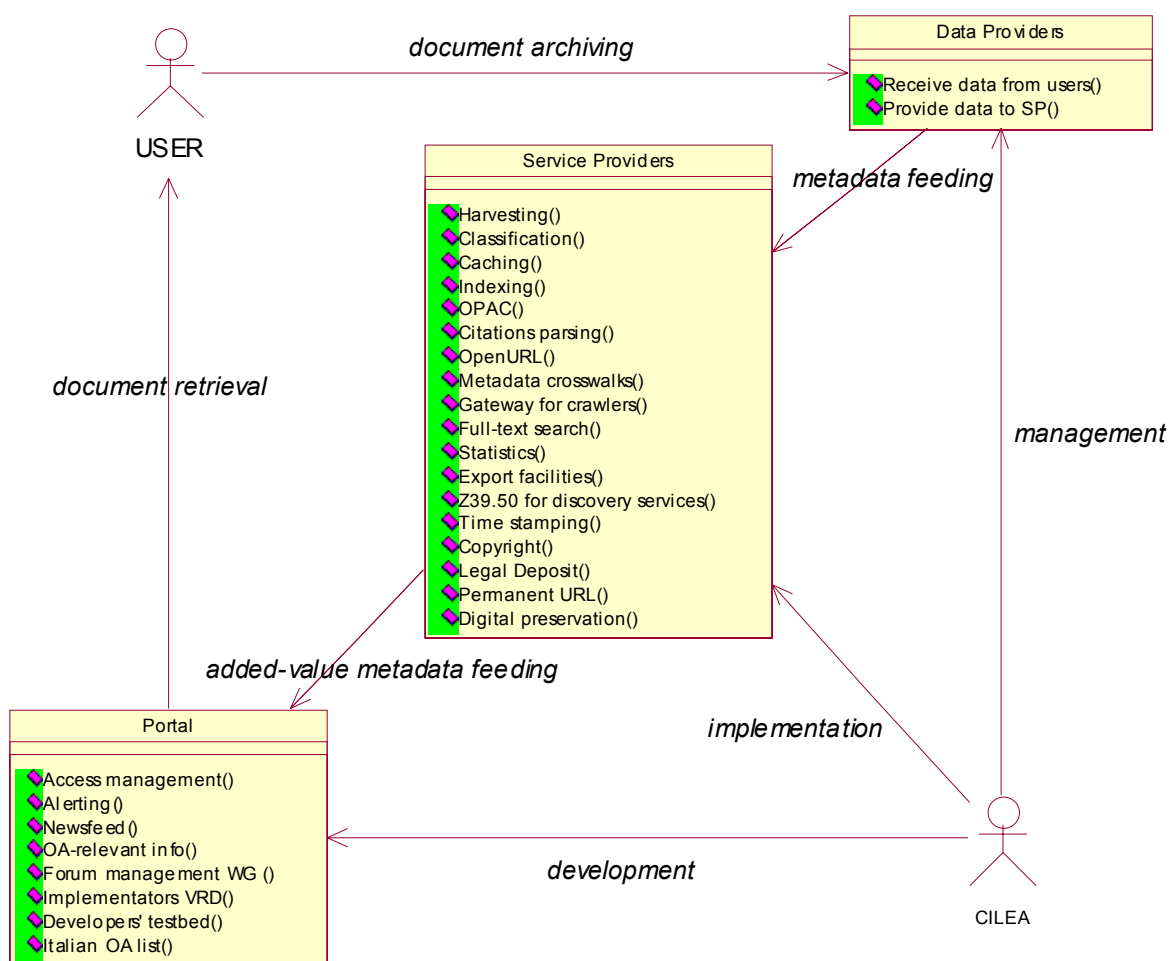
¹³ http://library.caltech.edu/openurl/Public_Comments.htm

Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

- funzionalità di ricerca nel testo pieno;
- statistiche di accesso;
- funzionalità di esportazione per la costruzione di report o pagine *Web* individuali o istituzionali, a fini di presentazione o valutazione;
- implementazione di un server Z39.50 per offrire la configurazione di servizi di *discovery* da inserire fra le risorse di sito, ad esempio cataloghi di biblioteche, delle istituzioni partecipanti;
- certificazione temporale, protezione dei diritti d'autore, deposito legale, indirizzo permanente e conservazione dei dati.

Attraverso un'interfaccia di ricerca, il portale offrirà un punto di accesso singolo ai dati centralizzati, e all'informazione rilevante legata all'ambiente degli archivi aperti, oltre ad una varietà di altri servizi personalizzati, basati su un sistema di gestione dei profili di accesso, quali *alerting*, *newsfeed*, e così via. Il sito del portale ospiterà anche una lista delle iniziative italiane in ambito OA, un *testbed* per gli strumenti ad uso degli sviluppatori, un *virtual reference desk* per gli implementatori, ed un *forum* per la condivisione a livello nazionale di linee guida relative agli archivi aperti.



Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

Per contribuire a garantire il massimo accesso alla letteratura scientifica prodotta in Italia, il CILEA fornisce servizi di implementazione e *hosting* di archivi aperti ed offre consulenza per la loro costituzione, il tutto utilizzando *software open source* e gli standard adottati dalla comunità OAI, garantendo la piena compatibilità dei dati a livello internazionale.

3. Problemi aperti e soluzioni

La scelta di Santa Fe¹⁴ di costituire archivi "leggeri" e caricare l'onere di mantenere servizi qualitativamente elevati a chi effettua la raccolta dei metadati lascia tuttavia aperti alcuni problemi.

Innanzitutto l'impiego nei depositi di *set* di metadati "poveri" rende difficile l'implementazione di servizi aggiuntivi, basati su aggregazioni di dati. Una soluzione parziale potrebbe essere quella di implementare la ricerca sul *full-text* dei documenti, ed individuare algoritmi di rilevanza (secondo lo stile dei motori di ricerca), tuttavia non sempre il testo è disponibile in un formato utilizzabile, né "harvestabile". In situazioni in cui l'oggetto descritto non è un documento testuale, ma ad esempio un'immagine, il *set* di metadati di provenienza deve necessariamente essere più ricco (si vedano le esperienze di musei e istituzioni di *cultural heritage* descritte a Bath al quarto *workshop* dell'OAF tenutosi nel settembre scorso¹⁵). La ragione per implementare metadati scarni è non scoraggiare con richieste eccessive gli autori che depositano, ma con l'intervento dei bibliotecari qualche progresso può essere previsto e diffuso trasversalmente fra i *Data Provider*.

Stesso discorso vale per il diritto d'autore: fintantoché non verranno definite le questioni relative ai diritti per le pubblicazioni scientifiche, stilati accordi espliciti con gli editori commerciali per consentire il *self-archiving*, descritti attraverso metadati esaustivi gli usi consentiti dei documenti nei depositi¹⁶, la raccolta dei dati in rete resterà un'operazione maldestra e non priva di rischi.

Un altro campo d'intervento dovrebbe riguardare le modalità di valutazione della ricerca, sia ai fini della progressione di carriera individuale dei ricercatori e docenti, sia dell'assegnazione di finanziamenti. Anche negli ambienti dove l'Impact Factor di ISI-Thompson detta legge, la proposta di correttivi basati sull'evidenza del maggiore impatto ottenuto dalle pubblicazioni in regime di *open access* ha buone possibilità di mostrare la propria *ratio*, sempre che vengano raffinati gli strumenti di raccolta e misurazione di questo tipo di informazioni, come il già citato CiteSeer.

Questi tre problemi possono essere risolti solo tramite accordi che, almeno a livello nazionale, riuniscano in un unico fronte compatto atenei ed enti di ricerca. Un gruppo

¹⁴ Herbert Van de Sompel e Carl Lagoze, "The Santa Fe Convention of the Open Archives", in D-LIB Magazine, 6 (2), 2000, testo disponibile all'URL: <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>.

¹⁵ Presentazioni disponibili all'URL: http://www.oaforum.org/workshops/bath_invitation.php.

¹⁶ A questo proposito l'Open Archives Initiative ha creato un gruppo tecnico denominato *oai-rights* per la stesura di linee guida, previste per la seconda metà del 2004:
<http://www.openarchives.org/documents/OAIRightsWhitePaper.html>.

Biblioteche digitali per la ricerca e la didattica: esperienze e prospettive

Parma, Casa della Musica, 22 novembre 2003

di lavoro italiano che esamini le soluzioni e le proponga alla comunità nazionale dovrebbe assumere un ruolo-chiave in questo periodo iniziale di incertezze.

Un altro ostacolo in cui inciampano le implementazioni di servizi basati su risorse informative digitali è l'integrazione: anche la costituzione di biblioteche digitali è resa faticosa dalla molteplicità di piattaforme e standard che nascono per gestire risorse diverse, dai periodici elettronici alle immagini, dagli strumenti per la didattica multimediale al cartaceo digitalizzato, dalle tesi agli OPAC. Ancora in linea con le raccomandazioni di Santa Fe, il progetto di piattaforma nazionale per gli Open Archives che il CILEA sta realizzando prevede la riesposizione dei metadati raccolti tramite il protocollo Z39.50, in modo da poterli integrare con le altre risorse che le biblioteche di ricerca mettono a disposizione dei propri utenti, che sono poi gli stessi produttori delle informazioni.

Infine, i costi a carico dei *Service Provider*, e di tutte le piattaforme di servizi quale quella qui illustrata, non sono indifferenti, sia in termini di progettazione sia di manutenzione e gestione. I servizi presentati potranno arricchirsi ancora se altre istituzioni accetteranno di allargare la collaborazione, presentando insieme proposte per i bandi di finanziamento. Il sostegno dei progetti di interesse collettivo, specie se condotti in *partnership*, dovrebbe comunque provenire dalle istituzioni che, in Italia ed in Europa, finanziano la ricerca e ne hanno a cuore le sorti.

4. Conclusioni

L'impegno sul fronte degli Open Archives va nella direzione della collaborazione: il progresso si basa sulla volontà di utilizzare strumenti e protocolli comuni, di uniformare quanto più possibile le politiche di accesso, di contribuire alla ricchezza di software *open source* già disponibile grazie alla generosità della comunità OAI. Il CILEA con il progetto qui illustrato intende perseguire questi obiettivi, e porre gratuitamente a disposizione della comunità nazionale ed internazionale tutto quanto verrà sviluppato per la piattaforma, invitando altre forze ad unirsi per la soluzione dei problemi aperti, in uno spirito collaborativo che ancora stenta ad affermarsi nel nostro paese. Ma l'invito è soprattutto rivolto ai bibliotecari, in qualità di maggiori esperti delle tematiche più complesse che si agitano nel mondo dei documenti digitali.