

Open Archives Initiative (OAI) in Europa

Workshop al CERN di Ginevra

Dal 22 al 24 marzo 2001 si è svolto a Ginevra, presso il CERN, il Workshop on The Open Archives Initiative (OAI) and Peer Review journals in Europe, promosso dalla Access Division of LIBER (Ligue des Bibliothèques Européennes de Recherche).

Il comitato organizzativo del workshop composto da Raf Dekeyser (LIBER, Access Division; Katholieke Universiteit Leuven), Herbert Van de Sompel (OAI Executive; Cornell University), Corrado Pettenati (CERN, Library) aveva previsto, nel dicembre 2000, la convocazione di scienziati e bibliotecari di tutto il mondo disposti a svolgere un ruolo attivo all'interno del workshop europeo.

Nella pagina in agenda sul sito del CERN è disponibile il programma con foto dei partecipanti, slide delle presentazioni e alcuni testi (in vari formati), <<http://documents.cern.ch/AGE/current/fullAgenda.php?id=a01193>>.

La nascita di OAI risale all'ottobre 1999 al meeting di Santa Fe, nel Nuovo Messico, meeting organizzato da Council on Library and Information Resources (CLIR), Digital Library Federation, Scholarly Publishing and Academic Resources Coalition, Association of Research Libraries e Los Alamos National Laboratory Research Library.

L'attività di OAI è supportata principalmente dalla Digital

Library Federation (DLF) e dalla Coalition for Networked Information (CNI), e in secondo luogo dalla National Science Foundation (NSF).

Il meeting di Santa Fe era focalizzato sull'interoperabilità degli archivi di preprint, collezioni di articoli e preprint posti su server di dipartimenti, università, centri di ricerca ed eventualmente organizzati a livello disciplinare, <<http://www.openarchives.org>>.

Al fine di preparare e promuovere il lancio di centri OA in Europa è stato perciò organizzato il workshop a numero chiuso a Ginevra, presso il CERN, centro internazionale per la fisica delle alte energie, che gestisce uno dei server di preprint più importanti del mondo, attraverso il servizio CDS (CERN Documents Server). Il workshop ginevrino, in un certo senso, aveva l'ambizione di aprire la strada ad un'accettazione più ampia dei fondamenti e della filosofia di OAI, estendendola anche al contesto europeo.

OAI risulta essere un'organizzazione in crescita, esplicitamente in transizione, aperta all'esplorazione di nuove tecnologie e di nuove forme di applicazione di esse. In particolare gli sforzi maggiori sono concentrati non solo sulla conoscenza delle tecnologie che sottendono le varie applicazioni e sulla compatibilità agli standard che sono stati sviluppati, e che si svilupperanno, ma sono anche orientati a

comprendere la struttura e la cultura delle varie comunità che adottano OAI.

La mission e l'organizzazione di OAI devono perciò necessariamente essere in continua evoluzione per adattarsi ai cambiamenti che di volta in volta si presentano.

Dalla Convenzione di Santa Fe ci sono state alcune tappe che hanno condotto alla stesura del documento noto come OAMH, The Open Archives Initiative Protocol for Metadata Harvesting (protocol version 1.0, document version 2001-01-21), recuperabile all'indirizzo, <<http://www.openarchives.org/OAI/openarchivesprotocol.htm>>.

Il documento che descrive le specifiche tecniche di compatibilità è fondamentale per l'implementazione del protocollo stesso, ma anche per comprendere in dettaglio la filosofia che permea l'architettura OAI.

I partecipanti al workshop erano poco più di sessanta. Per l'Italia era presente l'autrice del presente testo in qualità di partecipante ai gruppi del workshop e come bibliotecaria in rappresentanza dell'Università degli studi di Padova. Nella prima giornata c'è stata la partecipazione di tecnici informatici della SISSA di Trieste che hanno presentato il servizio TIP per i fisici, non compatibile con la piattaforma OAI.

Il compito principale degli inviati al workshop era di risolvere i punti cruciali messi in luce durante le sessioni "espositive" dove alcuni dei partecipanti (metà di parte statunitense e metà per l'Europa) esponevano esperienze e opinioni anche assai diverse tra loro.

I metadati sono stati il tema cruciale di tutto il workshop, quale parola chiave implicita o esplicita: incardinati in protocolli, espressi in standard (anche draft) e/o configurati in formati differenti, ma soprattutto oggetto di grande atten-

zione è stato il loro trasporto tra server comunicanti.

Per lo scambio dei metadati OAI adotta un set minimo del formato Dublin Core.

Data provider e service provider, le due componenti nello scambio di metadati tipiche della filosofia di OAI, sono state un altro dei punti cardine di cui si è discusso a lungo in vari interventi.

In particolare OAI distingue nettamente tra due aree, le quali si connotano per differenti peculiarità e funzioni, e precisamente:

- a) *data providers*;
- b) *services providers*.

L'approccio, noto come *metadata harvesting*, consente di mettere in relazione questi due settori distinti attraverso un colloquio di scambio di informazioni, da una parte i server contenenti le "informazioni sulle informazioni", ovvero i "dati sui dati" (metadati), dall'altra i servizi (*services providers*) che raccolgono i dati "esposti" dai *data providers* e li organizzano in servizi a valore aggiunto da offrire alle comunità scientifiche a seconda delle necessità espresse da quella particolare comunità di utenti.

La raccolta dati in OAI avviene all'interno di un modello di tipo distribuito che si fonda sull'architettura Harvest, in un meccanismo di scambio tra *gatherer* e *brokers*.

Harvest, nato da un progetto sulla ricerca distribuita condotto nel 1996 da Michael Schwartz, allora all'Università del Colorado, si fondava su un'idea di architettura che divideva le principali funzioni di un sistema di ricerca centralizzato in sottosistemi separati. Tale idea concettuale è rimasta come nocciolo nella filosofia di OAI. Il progetto Harvest definiva i formati e i protocolli per la comunicazione lungo questi sottosistemi. Il software implementato dimostra la centralità del concetto di *gatherer*, un

programma che colleziona informazioni indicizzate dalle varie collezioni all'interno di diverse biblioteche digitali. Ogni *gatherer* estrae informazioni indicizzate dalle collezioni e le trasmette in un formato standard, attraverso un protocollo standard, a programmi chiamati *brokers*. Il *broker* costruisce un indice combinato di informazioni sulle varie collezioni.

Al fine di trovare una soluzione all'attuale crisi dell'editoria scientifica in relazione alle pubblicazioni periodiche, i metadati sono stati visti all'interno del workshop come utensili per gli autori e per la comunità scientifica, utilizzabili all'interno dei processi di *submission* automatici offerti da alcuni server.

Tale logica prevede che i metadati siano visti come strumenti in grado di conformarsi all'interno di una piattaforma, in questo caso OAI, adatta alla regolazione di processi di validazione dei contenuti sottomessi dagli stessi autori e offerti dai server. Al fine di offrire documenti validati a vario livello di certificazione, le biblioteche si pongono nel ruolo di *services provider*. Di questo si è discusso durante la seconda sessione del workshop. Il meeting aveva principalmente due obiettivi:

- l'*obiettivo concreto* era quello di assemblare un gruppo di esperti (scienziati e bibliotecari) che volesse prendere parte immediatamente ad azioni guida nello sviluppo di OAI, obiettivo rivolto soprattutto a implementatori in Europa di e-print server già operativi o nuovi, ma che comunque si rendessero disponibili ad aderire agli standard di OAI;
- l'*obiettivo esplorativo* del meeting puntava ad una fase già successiva e precisamente alla certificazione dei lavori sottomessi ai server da parte degli autori. Ma come attuare questo processo in modo eco-

nomico ed efficace? quale il ruolo degli editori? Dato per certo il vasto consenso attorno alla questione della necessità di trovare forme di certificazione dei lavori scientifici che ne garantiscano la qualità, la domanda era: come può essere possibile implementare, in una infrastruttura di nuova concezione, funzioni che prevedano tale certificazione attraverso sistemi accettabili dalla comunità scientifica?

Il programma si è composto di brevi introduzioni tematiche e sessioni di lavoro attraverso gruppi di discussione. Gli argomenti trattati sono stati i seguenti:

- Introduzione generale all'architettura tecnica di OAI
- Il ruolo della peer-review nella comunicazione scientifica
- Implementazione tecnica e gestione di un archivio di pre-print
- Presentazione degli archivi esistenti che implementano l'architettura OAI
- Implementazione tecnica e gestione di sistemi di peer-review
- Il ruolo della certificazione nella comunicazione scientifica
- Ipotesi per nuove implementazioni delle funzioni della certificazione
- Implementazione tecnica e gestione dei sistemi di certificazione
- Integrazione di una rete di archivi e-print con sistemi di certificazione
- Modelli economicamente sostenibili
- Prototipi

Nell'ambito di tali linee il workshop ha individuato alcuni nodi cruciali, riassumibili in cinque domande chiave per l'obiettivo concreto:

- 1) come devono essere organizzati gli archivi? devono riferirsi ad una base istituzionale o piuttosto essere orientati su base disciplinare?
- 2) a chi la responsabilità per il lancio, la gestione di tali server?
- 3) a chi la responsabilità di as-

sicurare le funzioni "archiviazione a lungo termine" dei documenti contenuti nei server?

4) quali saranno le parti coinvolte? le università, le società scientifiche, i centri di ricerca internazionali o gli editori e/o aggregatori?

5) quali parti risulteranno vincenti nell'implementare archivi prototipo in questa fase iniziale? L'obiettivo esplorativo del workshop, forse anche il più ambizioso, è stato quello di provare a identificare e discutere le proprietà e le caratteristiche principali, richieste agli OA, di materiale non soggetto a peer-review. La proposta è stata quella di individuare differenti meccanismi di certificazione dei documenti, entro blocchi da costruire all'interno di un nuovo modello di comunicazione scientifica tale da indirizzare i bisogni della scienza, della comunità scientifica e della società pubblica in generale. Propongo di seguito una rapida carrellata relativa alle introduzioni tematiche che hanno preceduto i *panels* dei gruppi di lavoro.

In apertura della prima giornata, dopo le introduzioni di Corrado Pettenati, Raf Dekeyser e Elmar Mittler, Van de Sompel ha fatto una panoramica generale su storia, significato e nuovi orizzonti di OAI. Ha sottolineato che obiettivo dell'incontro era non tanto sentire esperienze "diverse" non orientabili al protocollo, quanto piuttosto quello di creare una comunità OAI forte.

Ci sono stati relatori più o meno incisivi. Il matematico tedesco Roland Schwaenzl di MPPRESS ha fatto scattare un dibattito leggermente sopra le righe, tenuto a freno da Heberhard Hilf che dirige il servizio analogo per la fisica, PhysNet. Tale servizio è descritto in un articolo sulla rivista online internazionale "High Energy Physics Libraries Webzine" (issue 4 del 2001), <<http://library.cern.ch/HEPLW/>>.



Schwaenzl ha parlato del ruolo dei MetaData Dublin Core nel servizio per la matematica MathNet.

MPPRESS contrariamente a PhysNet di Hilf (i servizi sono fratelli e viaggiano quasi in parallelo) non ha ancora aderito a OAI.

Per la verità la compatibilità di PhysNet con OAI non è propriamente completa in quanto si tratta di un progetto che si fonda su Harvest per la raccolta di metadati, ma da server distribuiti che per il momento non adottano OAI. PhysDoc, all'interno di PhysNet, adotta lo standard internazionale Dublin Core.

Come spiegato sopra relativamente a Harvest, molti dei benefici dell'architettura di Harvest vanno persi se il *gatherer* non è installato anche localmente, con la collezione digitale. Il lavoro di adattamento su Harvest per la compatibilità OAI di PhysNet è un compromesso di cui MPPRESS dovrebbe tener conto.

Michael Jost di FIZ Karlsruhe, ha svolto il suo intervento sulla Electronic Library of Mathematics. Jost collabora al progetto europeo matematico EULER diretto da Bernd Wegner, editor-in-chief del Zentralblatt für Mathematik e responsabile della commissione editoria elettronica (Electronic Publishing Committee) in seno all'European Mathematical Society (EMS).

Simeon Warner, per i laboratori di Los Alamos, ha esposto The Los Alamos arXiv, ex server <xxx.gov> nato ancora dieci anni fa per idea di Paul Ginsparg. Warner, assai giovane e assai geniale per i contenuti proposti, ha saputo es- ➤

porre in modo chiaro ed efficace il ruolo che tali e-print server giocano nell'ambito della comunicazioni scientifica, un effetto trainante, a catena, che sta coinvolgendo altre discipline.

Il CERN, che giocava in casa, ha portato uno splendido lavoro sulla validazione effettuata all'interno dei loro server sui documenti depositati dagli autori, attualmente oltre 170.000.

Attraverso il processo di *submission* da parte degli stessi autori, processo tutto automatizzato e altamente personalizzato (la nuova versione è stata messa a punto nel settembre scorso) è possibile attuare alcune strategie orientate alla validazione dei documenti sottomessi. Questo è possibile utilizzando differenti percorsi tra cui anche avvalersi di tag entro i metadati.

Sono stati illustrati cinque differenti processi di validazione e varie ipotesi correlate. Thomas Baron e Tibor Simko, insieme a Jean-Yves Le Meur (non presente al workshop) fanno parte dello staff CDS (CERN Document Server) che già dalla scorsa estate ha messo a punto un motore di ricerca *ad hoc* che consente l'interrogazione a testo pieno sugli oltre 170.000 documenti contenuti in CDS. La ricerca full-text è resa possibile attraverso un'interfaccia dinamica costruita con WebLib. Lo staff CDS lavora a stretto contatto con i bibliotecari.

Poi è stata la volta di Leslie Carr che ha parlato di Scientometric e del sito e-print, con una presentazione ricchissima di contenuti dal titolo *Distributed eprints archives and scientometrics*.

Carr lavora all'Università di Southampton con Stevan Harnad.

Thomas Krichel ha illustrato RePEC, il servizio per gli economisti. Krichel è un economista, ma attualmente è professore di biblioteconomia alla Palmer School of Library and

Information Science at Long Island University. Nel 1993 ha fondato NetEc, un consorzio di progetti Internet per economisti accademici. Nel 1997 in Gran Bretagna ha fondato la rete RePEC, con l'archivio dei documenti di ambito economico in cooperazione con Giappone, Russia, Stati Uniti, Canada e vari altri paesi. Attualmente assieme a Simeon Warner di LANL (Los Alamos National Laboratories) sta mettendo a punto un documento in ambito OAI dal titolo *Vocabulary for the Academic Metadata Format (AMF)*, che definisce, descrive e codifica sia i documenti accademici creati e utilizzati dalla comunità accademica sia le organizzazioni stesse, quali università, editori accademici scientifici, società scientifiche, ecc.

E-BioSci, a platform for e-publishing and information integration in the life sciences è stato il titolo presentato da Les Grivell di EMBO, il quale ha esposto il modello europeo di e-print server per la biomedicina e scienza della vita, contrapposto allo statunitense PubMed Central, voluto da Harold Varmus di NIH (National Institutes of Health) di Bethesda. Due modelli concettualmente differenti che riflettono due culture profondamente diverse.

J. Weeks ha presentato il neonato Chemistry Preprint Server, di grande effetto scenico, ma che ad oggi contiene pochissimi documenti.

Krichel ha coordinato anche la mattina successiva; nel suo ruolo di *chairman* ha saputo condurre in modo vivace ma al contempo rigoroso le due sessioni, i cui titoli erano:

– Agire assieme per supportare la comunicazione scientifica attraverso gli e-print (obiettivo concreto).

– Esplorare i confini della certificazione e gestione del sistema di comunicazione scientifica (obiettivo esplorativo).

La giornata è stata ancora più interessante della precedente, con vari interventi di notevole calibro.

La prima sessione, che proseguiva dalla giornata precedente, ha avuto termine con una comparazione delle opinioni di quattro eminenti scienziati e bibliotecari: Raf Dekeyser, Lorcan Dempsey, Hans-Joachim Wätjen e Tony Mayer sul tema *Towards a concerted action for a european network of e-print servers*.

Il focus della seconda parte era incentrato sul processo di referaggio (peer-review) e sul ruolo delle biblioteche nell'offerta di servizi correlati ai server di preprint. La seconda fase ha offerto gli interventi sulla questione della certificazione.

In questo contesto è stato presentato il progetto olandese Roquade, da parte di B. Save-nije.

Simon J. Buckingham Shum (bibliotecario) e Gary Li (informatico) del Knowledge Media Institute, Open University (UK) hanno esposto il loro lavoro su *Open archives meet open peer-review*, facendo esempi assai pragmatici.

Ronald Schmidt della Hochschulbibliothekszen-trum di Osnabruck ha portato un lavoro assai interessante che ci interessa da vicino: *Can libraries be intermediaries for the acquisition, production and presentation of scholarly information?* Nel suo intervento Schmidt ha offerto la visione di un ambiente di editoria elettronica dinamico, come valore aggiunto a portali per la biblioteca elettronica.

Stevan Harnad, arrivato da Firenze giusto in tempo per dirigere il gruppo di lavoro sulle peer-review, ha esposto la sua teoria nell'intervento dal titolo *Peer review reform: an empirical matter* di cui si trovano in linea sia i *trasparen-cies* sia il testo pieno del documento. Nel suo stile acceso, tipo lista-di-discussione,

ha sostenuto la centralità e l'importanza del processo tradizionale del sistema peer-review.

Harnad ha utilizzato varie metafore per spiegare come non si possa prescindere da una peer-review di qualità e come questo processo non possa essere abbattuto nei costi, pena un decadimento della qualità dei documenti.

Sul versante opposto il grande Bill, cioè William Y. Arms, che ha fatto la parte del leone con un intervento mirabile, ricco di osservazioni estremamente interessanti a favore degli open archives e della non sempre buona qualità del tradizionale processo di referaggio offerto dalle costosissime riviste scientifiche. *Quality control in scholarly publishing. What are the alternatives to peer-review?* ha catturato l'attenzione generale. Arms con esempi assai divertenti, esattamente al polo opposto di Stevan Harnad, ha proposto alternative decisamente percorribili, tra le quali la divisione dei costi di produzione di una rivista, scorporando la parte relativa all'attività di referaggio.

Sia Harnad che Arms sono convinti che una validazione dei lavori sia presupposto fondamentale per un'offerta di contenuti di qualità, e le loro tesi hanno percorsi differenti probabilmente solo nella forma in cui questi sono tracciati. La teoria di Arms (e gli esempi da lui citati sono assai pertinenti oltre che convincenti) è quella di dividere le due attività: le università mettono a disposizione sui server i lavori scientifici e i *publishers* si occupano del referaggio (dentro o fuori le banche dati), quali agenzie specializzate nella recensione scientifica.

In questo modo la comunità scientifica ha subito ciò che le serve per la propria crescita intellettuale, cioè informazione immediatamente fruibile.

le. Gli autori rimangono i detentori dei diritti e le università non sono obbligate a sottostare ai perversi meccanismi degli editori che attualmente acquisiscono i diritti economici (e perciò li detengono) che gli autori spesso cedono gratuitamente pur di veder pubblicato il proprio lavoro (meccanismo delle carriere). In sostanza anche Harnad dice esattamente le stesse cose. Gli editori che sostengono che i prezzi sono alti perché il referaggio costa possono in questo modo sganciare tale attività dai costi produttivi, concentrando i propri sforzi (questo dovrebbe essere il vero compito di un periodico scientifico) per creare nuova produzione intellettuale sulla produzione intellettuale di altri. La loro attività di referaggio diverrebbe così altro prodotto intellettuale, altro valo-

re aggiunto, su cui gravano ovviamente dei diritti editoriali, sganciati da quelli degli autori. Tale nuovo meccanismo non può però appoggiarsi all'attività gratuita di referaggio da parte dei cervelli degli atenei, a meno che il periodico non agisca senza scopo di lucro. Arms ha poi esposti altri modelli, come per esempio il meccanismo di vari livelli di certificazione attraverso l'utilizzo di metadati di qualità. Un intervento epocale.

Altri interventi sono stati quelli di alcuni responsabili commerciali, Mark Doyle di APS, *Peer-review's future in a world of open archives*, IOP e Interscience Wiley.

Eberhard Hilf e Hans-Joachim Waetjen di Onsbruck hanno illustrato il sistema di referaggio del servizio PhysNet, di cui ho parlato all'inizio.

Nella parte relativa ai *panels*, sono state indirizzate cinque linee strategiche, corrispondenti a cinque *panels* che riflettevano i nodi cruciali emersi durante le due sessioni. Dei sette proposti dal comitato organizzatore è stato possibile costituirne solo cinque.

L'obiettivo era un brainstorming su alcune questioni strategiche che riguardano OAI, per cui ciascun gruppo ha prodotto un proprio *panel*.

All'interno di ogni gruppo si è tentato di rispondere, con soluzioni articolate, alle problematiche proposte in una serie di quesiti posti dal comitato organizzatore.

I gruppi (*rapporteur* tra parentesi) erano:

1) Supporto organizzativo per un'azione a livello di piattaforma europea a sostegno degli sforzi innovativi all'interno della nuova architettura della

comunicazione scientifica (Raf Dekeyser);

2) Meccanismi alternativi di peer-review per i preprint (Mark Doyle);

3) Protocolli per la certificazione degli archivi (Herbert Van de Sompel);

4) Economia per una nuova architettura della comunicazione scientifica, chi paga cosa? (Corrado Pettenati);

5) Ruoli della biblioteca (Lorcan Dempsey).

Dopo la presentazione dei *panels* da parte dei gruppi, si è proceduto alla stesura di un documento finale e composto dai punti principali emersi dalla discussione dopo la presentazione dei cinque *panels*.

Il documento finale, emerso da queste giornate intense e denominato "raccomandazioni", fungerà da base per i prossimi sviluppi di OAI.

Antonella De Robbio