

PLEIADI, un portale per la letteratura scientifica “Open Access”

Ugo Contino*, Paola Gargiulo*, Susanna Mornati, Zeno Tajoli****

*CASPUR, Roma

**CILEA, Segrate

Abstract

Il progetto PLEIADI (acronimo per “Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali”) nasce dalla collaborazione tra CASPUR e CILEA, due dei maggiori consorzi universitari italiani, all’interno del progetto AEPIC. PLEIADI è un servizio nazionale che offre un accesso centralizzato alla letteratura scientifica presente negli archivi OAI istituzionali delle università e dei centri di ricerca italiani.

The PLEIADI Project (acronym for “Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali”, a portal for Italian scholarly e-literature in open archives and institutional repositories) originated from the collaboration between two major Italian university consortia, CASPUR and CILEA, within the framework of the AEPIC project. PLEIADI aims at building a national platform that offers centralized access to the scholarly literature archived in Italian repositories

Keywords: Open Archives, Open Access, OAI, PLEIADI, AEPIC

1. Introduzione: il contesto dell’Open Access e gli Open Archives

Nelle¹ ultime decadi siamo stati testimoni di un paradosso venutosi a creare nell’ambito dell’editoria scientifica: da una parte Internet ha profondamente innovato le modalità di comunicazione e scambio di informazioni nell’ambito della ricerca, dall’altra numerosi fattori hanno impedito un’ampia e libera distribuzione dei risultati della ricerca, l’avverarsi della “quarta rivoluzione” nei mezzi di produzione della conoscenza [10].

Molti di questi fattori possono essere riassunti nell’espressione ‘crisi dei periodici’, che è strettamente connessa ai limiti del modo tradizionale di comunicare risultati ufficiali della ricerca [8], ossia pubblicare: la scarsa capacità di innovazione dimostrata dagli editori,

le aspettative disattese nei confronti della rapidità di pubblicazione e ampiezza di disseminazione, le restrizioni d’accesso applicate alle edizioni online, il crescente costo di abbonamento ai periodici, il ruolo predominante di pochi editori commerciali che possiedono un numero elevato dei titoli più prestigiosi e hanno di fatto creato un monopolio dell’informazione scientifica, soprattutto nel campo STM - scienza, tecnologia e medicina [20]. Questa situazione crea insoddisfazione negli autori, che pubblicano alla ricerca dell’impatto e non per ritorni economici [9].

In questo contesto alcune iniziative si sono fatte portabandiera della necessità di un accesso libero alla letteratura scientifica² attraverso la proposta di due strategie complementari: l’auto-archiviazione e l’editoria ad accesso aperto.

La prima strategia in particolare prevede che siano gli autori stessi a depositare una copia elettronica (e-print) dei propri lavori, distinti in post-prints se già pubblicati, o pre-prints se

¹ Il presente lavoro costituisce un aggiornamento della relazione di Gargiulo, Paola e Mornati, Susanna e Contino, Ugo e Tajoli, Zeno, “PLEIADI, a portal solution for scholarly literature”, in Dobrev, Milena and Engelen, Jan, Eds. Proceedings *ELPUB2005. From Author to Reader: Challenges for the Digital Content Chain: Proceedings of the 9th ICC International Conference on Electronic Publishing*, pp. 277-281, Leuven-Heverlee (Belgium).

² Ad esempio la Budapest Open Access Initiative (BOAI), <http://www.soros.org/openaccess>, 17-10-2005

comunque sottoposti al peer-review presso riviste tradizionali, in un archivio online compatibile con gli standard OAI [15]. All'interno dell'architettura OAI sono stati realizzati diversi strumenti per la creazione e l'interoperabilità degli archivi aperti usando il protocollo OAI-PMH [17]. Gli archivi OAI si distinguono sostanzialmente in istituzionali e disciplinari: i primi raccolgono i lavori di ricerca e altri documenti prodotti da una specifica istituzione, i secondi raccolgono contributi limitati a uno specifico campo di ricerca o disciplina a prescindere dall'affiliazione dell'autore [7]. Obiettivo comune è mettere a disposizione dei ricercatori e del pubblico la maggior quantità possibile di contributi scientifici ad accesso aperto, senza la necessità del pagamento per la lettura ed il riutilizzo dei contenuti [11].

La seconda strategia prevede che le riviste scientifiche modifichino il proprio modello economico in modo da non imporre il pagamento ai lettori ma ricorrendo a diverse forme di sostentamento, quali il modello "author pays" [14], [22], anche in forma graduale [19]. Inoltre, anche le riviste ad accesso aperto possono utilizzare il protocollo OAI-PMH per rendersi più visibili ed utilizzabili dalla comunità scientifica.

Recentemente in diversi paesi università ed istituti di ricerca hanno aderito alle istanze del movimento Open Access [3], diverse istituzioni dei paesi più avanzati hanno reso obbligatorio l'accesso aperto ai risultati delle ricerche finanziate con denaro pubblico³, studi ed indagini sono stati intrapresi per dimostrare che le ricerche ad accesso aperto sono più citate [12],[13] e che gli autori sono disponibili a metterle liberamente a disposizione del pubblico [21].

Il risultato di questi sforzi è un numero crescente di archivi aperti operativi, una costante crescita di contenuti open access e dunque una maggiore diffusione della letteratura scientifica. In questo modo aumenta l'impatto dei risultati della ricerca, con benefici

per gli autori, le loro istituzioni, l'intera società⁴.

2. Gli archivi OAI italiani

L'implementazione di archivi aperti italiani è ancora agli inizi⁵. Poche istituzioni hanno già adottato politiche per incoraggiare l'auto-archiviazione e strategie per disseminare liberamente ed ampiamente i risultati delle proprie ricerche. Tre archivi sono pienamente operativi anche se la quantità di documenti di ciascuno non supera le poche centinaia: si tratta delle installazioni delle università di Bologna, Firenze e Trento. Altri sono ancora in fase di sperimentazione o si limitano a una specifica parte della produzione accademica (ad esempio le sole tesi di dottorato, come nel caso dell'Università di Roma "La Sapienza", o i soli materiali didattici a Parma). Alcuni contengono per ora una quantità di descrizioni di pubblicazioni maggiore rispetto ai documenti a testo completo liberamente accessibili (ad esempio SISSA e Università di Messina). Altri come l'Istituto Universitario Europeo o l'International Centre for Theoretical Physics hanno un archivio in Italia, ma non sono istituzioni italiane. L'Italia ospita inoltre alcuni archivi disciplinari, fra i quali E-LIS ha notevole consistenza, contando circa 3000 documenti di provenienza internazionale.

Negli ultimi mesi sono state avviate numerose iniziative, tra cui diversi nuovi archivi in fase di progettazione o di avvio. Con il supporto della Conferenza dei Rettori delle Università Italiane (CRUI), nel novembre 2004 a Messina si è tenuto un evento nazionale in occasione del quale più di 30 atenei si sono trovati concordi sul principio dell'accesso aperto alla letteratura scientifica ed hanno aderito alla "Dichiarazione di Messina" [2]. Questo evento ha diffuso anche a livello istituzionale in Italia una maggiore consapevolezza riguardo ai benefici dell'open access e le opportunità offerte dalle tecnologie connesse. Nei mesi successivi la quasi totalità degli atenei italiani ha firmato la "Berlin Declaration" [3] e ci si attende dunque che ciascuno dia seguito alle intenzioni espresse con azioni concrete, quali l'installazione di un archivio aperto e l'adozione di una politica di

³ Si veda il recente esempio dei NIH, National Institutes of Health, <http://www.nih.gov/about/publicaccess/>, 17-10-2005, e del Wellcome Trust, http://www.wellcome.ac.uk/doc_WTD002766.html, 17-10-2005. A breve è atteso anche il pronunciamento dei RCUK, Research Councils United Kingdom, <http://www.rcuk.ac.uk/access/index.asp>, 17-10-2005.

⁴ Cfr. il decimo rapporto 2004 della commissione parlamentare del Regno Unito sulla scienza e la tecnologia, <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39902.htm>, 17-10-2005

⁵ La lista completa degli archivi e repertori OAI italiani: <http://www.openarchives.it/>, 17-10-2005

ateneo a supporto dell'accesso aperto ai risultati delle proprie ricerche. (cfr. Institutional Self-Archiving Policy Commitment, <http://www.eprints.org/openaccess/policysignup/sign.php>, 17-10-2005).

Per incoraggiare gli autori a depositare le copie delle proprie pubblicazioni risulta senza dubbio utile disporre di una serie di servizi che aggiungano valore alla letteratura scientifica ad accesso aperto e ne aumentino la visibilità e le possibilità di utilizzazione.

3. PLEIADI: nuovi servizi per l'utente

Il rilascio del protocollo OAI-PMH nel 2001[17] ha dato il via alla creazione di vari strumenti per l'interoperabilità dei metadati tra diverse piattaforme, applicazioni e software. Quando un documento viene depositato in un archivio OAI, il suo potenziale lettore lo può localizzare con servizi centralizzati di ricerca conosciuti con il nome di service providers.

Il progetto PLEIADI (acronimo per "Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali") [18] nasce dalla collaborazione tra CASPUR e CILEA, due dei maggiori consorzi universitari italiani, all'interno del progetto AEPIC [1]. Lo scopo di PLEIADI è di costruire e gestire un portale per l'accesso centralizzato alla letteratura scientifica presente negli archivi OAI istituzionali delle università e dei centri di ricerca italiani (figura 1).

Il servizio è stato lanciato nel novembre 2004 durante il citato evento di Messina ed è in continua evoluzione per seguire gli sviluppi della tecnologia connessa agli Open Archives.

Le parole chiave legate a questa iniziativa sono: ricerca centralizzata, servizi personalizzati, visibilità e impatto, consapevolezza e disseminazione, deposito legale e preservazione.

The screenshot shows the PLEIADI website interface within a Mozilla browser window. The browser's address bar displays the URL: http://www.openarchives.it/pleiadi/index.php?sel_lang=italian. The website header includes the title "PLEIADI: Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali" and a navigation menu with items like "Home", "Ultime notizie", "RSS: Notizie dal mondo OA", "Forum", "Ricerca negli archivi", "Risorse Web", "Progetto PLEIADI", and "Chi siamo".

The main content area is divided into several sections:

- Login:** A section for user authentication with fields for "Nome utente:" and "Password:", and a "Login utente" button.
- Ricerca in OpenArchives:** A search bar with a "Cerca" button and a link to "Altre modalità di ricerca".
- Manifesto del progetto PLEIADI:** A text box containing the project's mission statement: "Il Progetto PLEIADI (Portale per la Letteratura scientifica Elettronica Italiana su Archivi aperti e Depositi Istituzionali) scaturisce dalla collaborazione fra due importanti consorzi interuniversitari italiani, CASPUR e CILEA, nell'ambito del progetto AEPIC. [continua...]"
- Notizie recenti:** A list of recent news items with dates, such as "Earth-prints, Roma 12 ottobre 2005 (10/10/2005)", "Il sapere Liberato : Gruppo LASER (22/9/2005)", and "Indisponibilità del portale PLEIADI - 22.9.05, h 8.30-9.30 (20/9/2005)".
- Discussioni recenti:** A section for recent discussions with buttons for "Forum", "Discussioni", "Risposte", "Visite", and "Ultimo messaggio".

The browser's taskbar at the bottom shows the Windows Start button, several open applications (Eudora, PLEIADI), and the system tray with the date and time (9.09).

Fig.1 – Home page di PLEIADI

Alcune di queste sono divenute dei servizi ora pienamente operativi, altri ne verranno implementati nel prossimo futuro, altri sono allo studio.

Lo scopo del progetto è di permettere alla comunità accademica di accedere alla ricerca italiana disponibile negli archivi aperti istituzionali aumentandone la visibilità e l'impatto. Inoltre PLEIADI offre alcuni servizi personalizzabili, che includono la creazione di un proprio profilo utente, l'*alerting*, ossia la possibilità di essere avvisati quando una nuova informazione viene immessa in un modulo del portale, il salvataggio delle proprie ricerche e citazioni, la disponibilità di fonti di informazioni differenziate (news, forum, RSS).

Altri servizi come l'interpretazione delle citazioni, informazioni statistiche sui risultati delle ricerche, export per costruire proprie pagine web o bibliografie verranno implementati nel prossimo futuro. Servizi come la certificazione del momento di deposito, la protezione del diritto d'autore, il deposito legale, la conservazione a lungo termine, la ricerca nel testo completo, verranno sviluppati più avanti nel tempo, quando verrà raggiunto un accordo fra gli archivi riguardo alle modalità di trasporto dei dati e metadati necessari.

Infine PLEIADI intende promuovere la consapevolezza dell'Open Access in Italia, contribuire al dibattito, facilitare l'avanzamento delle competenze tecniche, attraverso i propri forum come luogo di scambio di informazioni per i membri della comunità dell'Open Access.

PLEIADI è una piattaforma a più livelli. Sotto il livello visibile del portale c'è una struttura complessa di componenti. L'interfaccia utente è sviluppata con XOOPS [23], un software *open source* per la costruzione di portali. In PLEIADI è possibile identificare diversi blocchi funzionali:

- **area utente**, comprende il front-end dell'utente, il multilinguismo e l'autenticazione. In quest'area l'utente personalizza il proprio ambiente, e attiva i servizi personalizzati. In quest'area ci sono anche le news, i forum e l'RSS, link e documenti. In quest'area si raccolgono e si distribuiscono le informazioni sull'open access;
- **area delle ricerche**, gestisce la connessione Z39.50 con l'*harvester* e il modulo di indicizzazione dei record. Sono state implementate due modalità di ricerca,

semplice e avanzata. La seconda permette ricerche su campi differenziati, filtri, ordinamento, etc. ;

- **area della gestione**, con un intuitivo *content management engine*, quest'area gestisce tutto il sito web con un'interfaccia grafica;
- **area dei log e delle statistiche**, per le statistiche d'uso dei documenti e per le ricerche effettuate, in fase di progettazione.

Il portale è costruito con diversi moduli PHP realizzati dalla comunità di XOOPS. Il codice è stato adattato per ottenere l'accessibilità che caratterizza il sito di PLEIADI. Una GUI su web è disponibile per gestire la maggior parte dei moduli. XOOPS usa MySQL come database di appoggio e Apache come web server. Gli sviluppatori di PLEIADI hanno scelto XOOPS per diverse ragioni:

- disponibilità di funzioni di gestione per il portale e per i moduli;
- uso di template e fogli di stile per le pagine dinamiche;
- completa indipendenza dei moduli PHP dalla visualizzazione in HTML, come conseguenza dell'uso dei fogli di stile.

4. PLEIADI – Caratteristiche tecnologiche del service provider

PLEIADI come service provider⁶ effettua una serie di attività di back-end per scoprire, recuperare e rendere accessibili i contenuti open access conservati negli archivi aperti italiani.

I dati descrittivi riguardanti questi contenuti sono raccolti dagli archivi OAI istituzionali o disciplinari (chiamati anche data providers), dove sono depositati dagli autori. Una volta che le attività di raccolta, filtraggio, armonizzazione e indicizzazione dei metadati sono compiute, il risultato è un database collettivo, disponibile e ricercabile via HTTP e Z39.50. L'interfaccia Z39.50 permette che i metadati siano a disposizione anche di metamotori ed altri strumenti per la federated search e l'integrazione di risorse informative. L'interfaccia web è stata realizzata basandosi sullo Z39.50 con la libreria YAZ per PHP.

I metadati sono raccolti usando il protocollo OAI-PMH, attraverso l'*harvester* Celestial [4], a

⁶ L'architettura del sistema è simile a quella usata dall'ePrints UK Project. Si ringraziano UKOLN e RDN per il supporto dato nell'implementare questa soluzione, <http://www.miur.it/UserFiles/115.htm>, 17-10-2005.

cui sono state apportate alcune modifiche, quale la possibilità di raccogliere i metadati di ciascun archivio selezionando i singoli set, tramite interfaccia grafica. In questo modo è stato possibile raccogliere i metadati dei contributi italiani di archivi disciplinari internazionali.

Quando un nuovo archivio viene installato, è aggiunto alla lista di PLEIADI. I metadati sono raccolti nel formato OAI_DC [16], che ha la stessa semantica dello standard Dublin Core non qualificato [6]. PLEIADI utilizza anche i metadati presenti nelle sezioni 'header' e 'about' (descritte nello standard OAI-PMH) dei record XML raccolti.

Un'altra opzione aggiunta è l'integrazione dei convertitori (*crosswalks*) che potenzia Celestial come strumento di *caching* di metadati. Anziché riesporre i metadati originali, Celestial riespone i metadati convertiti, in modo da facilitare il loro riutilizzo da parte di altri *service providers*.

Tutti i metadati raccolti dai diversi archivi devono infatti essere convertiti ad un formato interno tramite specifici strumenti di mappatura. Numerosi campi hanno un significato identico o simile agli equivalenti campi del Dublin Core non qualificato. L'elemento che richiede più interventi è `dc:subject`, che contiene la classificazione del MIUR⁷. I `dc:subject` originali di ogni archivio sono tradotti, ove possibile, in un sistema di classificazione coerente, dopo un'analisi dei valori del campo.

Gli indici sono costruiti partendo dai metadati in formato XML, usando il software Cheshire II [5] come indicizzatore. Cheshire II può essere installato come server che lavora su due protocolli, HTTP e Z39.50, opportunità che ne ha determinato la scelta.

La ricerca per l'utente finale è implementata traducendo le richieste HTTP in Z39.50. Il risultato delle richieste viene passato in formato XML a un apposito modulo PHP, basato sulla libreria DOM, che provvede all'ordinamento.

I nuovi sviluppi riguarderanno servizi personalizzati di ricerca, come ad esempio la possibilità di salvare le proprie strategie di ricerca, rilanciarle e ricevere periodicamente

aggiornamenti via email grazie alla registrazione di un profilo sul portale.

5. Integrazione delle risorse bibliografiche: OAI-PMH e Z39.50

PLEIADI costituisce un tentativo di esplorare le potenzialità del nuovo protocollo OAI-PMH insieme con il ben conosciuto Z39.50, e crea un modello di ricerca centralizzata per diverse categorie di fonti informative. Pubblicazioni elettroniche accademiche, specialmente articoli ed atti di congresso open access, trarrebbero beneficio da un'integrazione con altre fonti di informazioni, quali i cataloghi delle biblioteche, i riferimenti bibliografici delle basi di dati, i periodici degli editori commerciali, che costituiscono il nucleo tradizionale della fornitura di informazioni nell'ambiente accademico e della ricerca. E-prints, dissertazioni, dati grezzi e tutto il materiale depositato negli archivi aperti, possono contribuire ad estendere la quantità di informazioni disponibile agli utenti accademici, particolarmente se potranno essere cercate e trovate con stessi sistemi di ricerca che sono già in uso. Inoltre la collaborazione inter-istituzionale può essere incoraggiata con lo scambio di metadati, aumentando la diffusione dei risultati della ricerca scientifica. La raccolta di metadati con l'OAI-PMH, l'uso di convertitori appositi e le ricerche via Z39.50, possono essere combinati al fine di offrire una possibile soluzione per integrazione delle risorse bibliografiche.

La collaborazione di CASPUR e CILEA per mantenere e migliorare questo servizio costituisce un esempio della sinergia utile a produrre un beneficio collettivo alla comunità scientifica italiana.

Bibliografia

- [1] AEPIC, <http://www.aepic.it>, 17-10-2005.
- [2] Gli Atenei italiani per l'Open Access: verso l'accesso aperto alla letteratura di ricerca, Workshop, 4-5 Novembre 2004, Università di Messina, URL: <http://www.aepic.it/conf/index.php?cf=1>, 17-10-2005.
- [3] Berlin Declaration on Open Access to Knowledge in Sciences and Humanities, URL: <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>, 17-10-2005.
- [4] Celestial, <http://celestial.eprints.org>, 17-10-2005.
- [5] Cheshire II Project URL: <http://cheshire.berkeley.edu/>, 17-10-2005.

⁷ MIUR- Il Ministero italiano dell'Istruzione, dell'Università e della Ricerca ha rilasciato uno schema di classificazione basato sulle discipline per definire piani di studio e carriere. Questo schema è usato da molti archivi istituzionali italiani., http://www.mur.st.it/atti/2000/alladm001004_01.htm, 17-10-2005.

- [6] DCMI (2004), *Dublin Core Metadata Element Set, Version 1.1: Reference Description*: URL: <http://dublincore.org/documents/dces/>, 17-10-2005.
- [7] De Robbio, A. (2002): "Open Archive per la comunicazione scientifica", in *Notiziario SIMAI Società Italiana di Matematica Applicata Industriale* 5: 2-6, URL: <http://eprints.rclis.org/archive/00000043/>, 17-10-2005.
- [8] Guédon, J.-C., 2001: "In Oldenburg's long shadow: librarians, research scientists, publishers, and the control of scientific publishing", in *Proceedings of the 138th Annual Meeting, Association of Research Libraries*, Toronto, URL: <http://www.arl.org/arl/proceedings/138/guedon.html>, 17-10-2005.
- [9] Harnad, S., 1990: "Scholarly skywriting and the prepublication continuum of scientific enquiry", in *Psychological Science* 1: 342-343, URL: <http://cogprints.org/1581/>, 17-10-2005.
- [10] Harnad, S., 1991: "Post-Gutenberg galaxy: the fourth revolution in the means of production of knowledge", in *Public-Access Computer Systems Review* 2(1): 39-53, URL: <http://cogprints.org/1580/>, 17-10-2005.
- [11] Harnad, S., et al., 1994: "A subversive proposal", in Okerson, A. and O'Donnell, J., Eds., *Scholarly journals at the crossroads: a subversive proposal for electronic publishing*, Washington DC, Association of Research Libraries, publ. 1995, URL: <http://www.arl.org/scomm/subversive/toc.html>, 17-10-2005.
- [12] Harnad, S., Brody, T., 2004: "Comparing the Impact of Open Access (OA) vs. Non-OA Articles in the Same Journals", in *D-Lib Magazine* 10(6), URL: <http://www.dlib.org/dlib/june04/harnad/06harnad.html>, 17-10-2005.
- [13] Lawrence, S., 2001: "Free online availability substantially increases a paper's impact", in *Nature* 411: 521, versione estesa: "Online or Invisible?", URL: <http://citeseer.ist.psu.edu/online-nature01/>, 17-10-2005.
- [14] Odlyzko, A.M., 1998: "The economics of electronic journals", in Ekman, R. and Quandt, R., Eds., *Technology and Scholarly Communication*, University of California Press, URL: <http://www.dtc.umn.edu/~odlyzko/doc/eword.html>, 17-10-2005.
- [15] Open Archives Initiative (OAI), URL: <http://www.openarchives.org/> 17-10-2005.
- [16] OAI Dublin Core, URL: http://www.openarchives.org/OAI/2.0/oai_dc.xsd, 17-10-2005.
- [17] OAI- PMH, URL: <http://www.openarchives.org/OAI/openarchivesprotocol>, 17-10-2005.
- [18] PLEIADI, URL: <http://www.openarchives.it/pleiadi/index.php>, 17-10-2005.
- [19] Prosser, D. C., 2003: "From here to there: a proposed mechanism for transforming journals from closed to open access", in *Learned Publishing* 16(3): 163-166, preprint URL: <http://www.arl.org/sparc/core/index.asp?page=g29>, 17-10-2005.
- [20] Suber, P., 2003: "Removing the barriers to research: an introduction to Open Access for librarians", in *College & Research Libraries News* 64: 92-94, 113, unabridged: URL: <http://www.earlham.edu/~peters/writing/acrl.htm>, 17-10-2005.
- [21] Swan, A., Brown, S., 2005: *Open Access self-archiving: an author study*, JISC Technical Report, URL: <http://cogprints.org/4385/>, 17-10-2005.
- [22] Wellcome Trust 2004: *Costs and business models in scientific research publishing*, London, ISBN 1-84129-051-3, URL: http://www.wellcome.ac.uk/doc_WTD003185.html, 17-10-2005.
- [23] XOOOPS, URL: <http://www.xoops.org/>, 17-10-2005.