

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

Progetto AEPIC: gli Archivi aperti italiani su una piattaforma nazionale

Susanna Mornati (mornati@cilea.it)¹

Abstract

I *repositories* istituzionali di *e-prints*, i documenti elettronici prodotti nell'ambito delle attività di atenei ed enti di ricerca, costituiscono una risorsa fondamentale per la comunicazione scientifica. Nell'architettura prevista dall'OAI-PMH essi rappresentano il lato *Data Provider*, e necessitano dello sviluppo di *Service Provider* per garantire una partecipazione estesa all'iniziativa degli Archivi aperti. La comunità internazionale si sta interrogando sui servizi necessari per la valorizzazione del contenuto degli archivi, mentre il loro numero è ancora scarso e l'implementazione richiede investimenti non trascurabili e un ambiente collaborativo avanzato.

CILEA e CASPUR, due fra i maggiori consorzi italiani che forniscono servizi ad alto valore tecnologico alle università e agli enti di ricerca, si sono associati per la realizzazione di una piattaforma nazionale di accesso agli Archivi aperti italiani, una costellazione di *Service Provider* al servizio dei *repositories* già implementati e compatibili con il protocollo standard. Inoltre è in corso un'attività di supporto per l'installazione e la gestione di nuovi *Data Provider* per conto delle istituzioni che volessero dotarsi un archivio aperto senza sostenere investimenti locali in termini di risorse umane e tecnologiche, consentendo ai propri ricercatori di aumentare l'impatto dei propri risultati di ricerca e migliorando la visibilità della produzione scientifica dell'ente nel contesto nazionale ed internazionale.

1. Introduzione²

In una giornata quasi interamente dedicata ad istanze istituzionali di archivi di *e-prints* caratterizzate dall'impiego del *software* sviluppato a Southampton, vorrei introdurre tre esperienze diverse maturate nell'ambito del CILEA e del progetto AEPIC, in corso di condivisione con il CASPUR³.

AEPIC⁴ è un progetto di editoria elettronica al servizio degli atenei, che si sviluppa lungo le tre tematiche del supporto al *self-archiving*, alla gestione dei materiali

¹ *Project Leader* di AEPIC presso il CILEA, Consorzio Interuniversitario per le Tecnologie dell'Informazione e della Comunicazione, Segrate (Milano), <http://www.cilea.it>

² I link citati in questo documento sono stati visitati per l'ultima volta il 3 febbraio 2004.

³ Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca , <http://www.caspur.it>

⁴ Academic E-Publishing Infrastructures - CILEA, <http://www.aepic.it>

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

didattici e delle tesi, alla pubblicazione di riviste, atti di convegni e produzione monografica degli atenei e degli enti di ricerca. Il presente lavoro intende illustrare lo stato dell'arte del progetto, ed in particolare le prime tre realizzazioni, che pur se sviluppatasi nell'arco di pochi mesi sono già sulla strada del successo:

- E-LIS, l'archivio disciplinare internazionale per la biblioteconomia e le scienze dell'informazione realizzato con *software* GNU EPrints;
- DSpace@UniPr, l'archivio per la didattica delle Facoltà umanistiche di Parma realizzato con *software* DSpace - e che ne costituisce la prima installazione in Italia;
- una piattaforma di servizi per gli Archivi aperti esistenti e futuri, in collaborazione con il CASPUR ed altre istituzioni in Italia e in Europa.

Queste tre esperienze dimostrano la validità della scelta di collaborare e cooperare, a livello nazionale e internazionale, per garantire ai progetti sostenibilità e continuità nel tempo.

2. E-LIS, E-Prints in Library and Information Science

All'inizio del 2003 è stato inaugurato E-LIS⁵, un archivio disciplinare internazionale specializzato per il deposito di documenti di ambito biblioteconomico e bibliografico, oltre che di scienze dell'informazione e della comunicazione⁶. Realizzato con *software* GNU EPrints⁷, presenta un'interfaccia in lingua inglese, ma accetta contributi in tutte le lingue. Installato su server CILEA, è nato e viene mantenuto per iniziativa di alcuni membri del team DOIS⁸.

La collaborazione fra queste due entità ha dato modo di aggiungere nuove funzionalità rispetto a quelle offerte dal pacchetto *software* di base. E-LIS ha infatti implementato:

- un contatore dei record;
- l'estrazione dei riferimenti bibliografici mediante ParaTools⁹.

E' inoltre in corso lo sviluppo o l'installazione di *software* aggiuntivo per arricchire l'archivio di altri servizi, per consentire in particolare:

- la ricerca nei testi pieni tramite Google;
- funzionalità di importazione ed esportazione dei dati in formato XML;
- attivazione del *browsing* per autore, attualmente non supportato da EPrints;
- installazione di un pacchetto di analisi statistica di *hit* e *download*;

⁵ <http://eprints.rclis.org/>

⁶ Per una presentazione completa dell'archivio, delle sue origini, obiettivi e funzionalità cfr. Antonella De Robbio, "E-LIS: un Open Archive in Library and Information Science", in *Bibliotime*, VI (I), marzo 2003, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000201/>.

⁷ <http://www.eprints.org/>

⁸ Documents in Information Science, <http://dois.mimas.ac.uk/team.html>

⁹ <http://paracite.eprints.org/>: ParaTools (ParaCite Toolkit) è un set di moduli Perl per la gestione delle citazioni, che comprendono il reference parsing e routines di creazione e gestione di OpenURL.

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

- dotazione di licenza Creative Commons¹⁰ per ogni testo depositato.

In relazione a quest'ultima funzionalità, va sottolineato che anche quando i testi sono depositati dall'autore che ne concede i diritti di lettura, copia e riutilizzo a titolo gratuito, questo tipo di informazione andrebbe trasmessa dall'archivio sia per la fruizione del documento sia per lo scambio con sistemi diversi¹¹. Vale la pena di menzionare che le *policies* relative al *copyright* messe a punto ed adottate da E-LIS hanno meritato un'importante citazione sulle pagine del progetto RoMEO¹².

L'archivio contiene oggi oltre 450 documenti accessibili a testo pieno e provenienti da nazioni diverse, ed il suo successo è testimoniato dall'alta percentuale di *self-archiving* e dalla regolare distribuzione nel tempo dei depositi.

3. DSpace@UniPr, un deposito per la didattica dell'Università di Parma

Nel mese di dicembre 2003 il CILEA ha attivato DSpace@Unipr¹³, la Biblioteca Digitale Umanistica dell'Università di Parma, un *repository* per i materiali didattici della Facoltà di Lettere e Filosofia che adotta il *software* DSpace¹⁴ sviluppato dal Massachusetts Institute of Technology (MIT) con la collaborazione di HP.

La scelta del *software* è motivata dalla ricchezza di funzionalità rispetto alla gestione a più livelli dei documenti, particolarmente utile per l'accesso ai materiali didattici in quanto strutturabile con modalità differenziate per le esigenze dei vari corsi di laurea e di diploma. L'interfaccia per la ricerca ed il recupero delle informazioni è stata completamente tradotta in italiano e le *policies* delle diverse comunità e collezioni sono state adattate alle necessità ed ai flussi di lavoro dei singoli docenti.

Come per E-LIS, anche per l'installazione di DSpace@Unipr il CILEA ha previsto diverse attività di sviluppo o installazione di *software* aggiuntivo, in particolare per la gestione di *set* multipli di metadati, già prevista nell'architettura del *software* ma non implementata nella versione corrente.

Sono inoltre allo studio soluzioni per la gestione ottimale di metadati in UTF-8. Il *software* di DSpace ed il *software* di base di cui si avvale (in particolare Debian GNU/Linux, Apache, Java, PostgreSQL) sono in grado di supportare UNICODE, tuttavia in particolare per quanto riguarda l'indicizzazione e la ricerca i problemi da risolvere in termini di compatibilità richiedono ancora parecchio lavoro.

¹⁰ <http://creativecommons.org/>

¹¹ Per un approfondimento delle tematiche che legano il copyleft agli archivi aperti cfr. Susanna Mornati, "Istanze 'open' nella comunicazione scientifica: open archives", in *Proceedings Contenuti Open Source*, Milano 9 dicembre 2003, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000775/>.

¹² <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/>. RoMEO è un'iniziativa britannica per lo studio delle questioni legate al diritto d'autore che interessano la comunità accademica, che ha assunto valenza internazionale grazie agli ottimi risultati prodotti.

¹³ <http://dspace-unipr.cilea.it:8080/index.jsp>

¹⁴ <http://www.dspace.org/>

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

DSpace@Unipr è la prima installazione pubblica di DSpace in Italia. Realizzata a tempo di record grazie alla preparazione del *team* di Parma¹⁵ ed alle competenze tecnologiche già maturate al CILEA, costituisce un'esperienza all'avanguardia nell'ambiente umanistico, tradizionalmente più restio all'adozione di tecnologie e modalità di comunicazione innovative.

4. Una piattaforma nazionale per gli Archivi aperti: stato dell'arte

La terza esperienza in corso di realizzazione nell'ambito del progetto AEPIC è un'iniziativa congiunta CILEA-CASPUR, e punta alla realizzazione di una piattaforma di servizi per i *repositories* italiani, che incontri le nuove esigenze di modelli comunicativi aperti ed offra valore aggiunto ai documenti prodotti dalla comunità di ricerca nazionale attraverso strumenti raffinati di ricerca, indicizzazione, analisi citazionale ed integrazione con le altre risorse a disposizione degli studiosi.

La piattaforma¹⁶ presenta un'architettura a due livelli: da un lato una costellazione di *Service Provider* al servizio dei *repositories* già implementati e compatibili con il protocollo standard, dall'altro e ad un livello superiore la creazione di un portale per l'accesso ai dati raccolti e la presentazione di risorse al servizio della comunità scientifica e della comunità di tecnici che si occupano di sviluppare, implementare e gestire archivi aperti.

Dal momento che il contesto in cui si colloca la piattaforma e le sue caratteristiche sono state descritte in precedenti occasioni¹⁷, ci limitiamo in questa sede a tracciare brevemente un quadro riguardante lo stato dell'arte dei lavori.

Il cuore dei servizi è stato sviluppato e testato. Con l'impiego di *software open source* disponibile sono state implementate le funzionalità di *harvesting* e *caching*, e messo a punto un database per i metadati dotato di strumenti per l'indicizzazione ed il recupero dell'informazione, nonché di un'interfaccia di interrogazione attraverso *web browser*. E' in corso la messa a punto del server Z39.50, la mappatura dei caratteri UTF-8 e la definizione gli indici. Per l'inizio di marzo è previsto il rilascio del primo prototipo, e per fine marzo la messa in funzione di un servizio sperimentale, per il quale verrà testata l'aggregazione di E-LIS con altri archivi dedicati alle stesse discipline o materie correlate.

Nel frattempo prosegue lo studio delle possibilità legate alla classificazione per aree di ricerca, all'OpenURL, all'analisi citazionale, alle statistiche, ai *crosswalk* e agli strumenti per un *uploading* automatizzato, anche in collaborazione con altri progetti italiani e stranieri. In particolare verranno esaminate le possibilità di contribuire alla

¹⁵ guidato da Anna Maria Tammaro.

¹⁶ Maggiori dettagli sulle caratteristiche della piattaforma possono essere ottenuti consultandone il Project Plan, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000518/>.

¹⁷ Fra le altre cfr. Susanna Mornati, "Open Archives in Italia: una piattaforma nazionale", in *Proceedings Biblioteche Digitali per la ricerca e la didattica: esperienze e prospettive*, Parma 22 novembre 2003, testo disponibile all'URL: <http://eprints.rclis.org/archive/00000519/>.

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

definizione di un'estensione del protocollo OAI-PMH¹⁸ per consentire l'*harvesting* dei testi pieni, ai fini di implementare la ricerca di Google su un *corpus* più ampio di documenti anziché a livello del singolo archivio.

Nel complesso, le attività originariamente previste nei primi due *work package* delineati nell'ambito del *project plan* originario, relativi a coordinamento, presentazioni e *deliverable* il primo ed alla messa a punto di due prototipi di *data provider* il secondo, sono state svolte ed i tempi rispettati. Per quanto riguarda l'attivazione del prototipo di *service provider* con le funzionalità centrali, il rispetto dei tempi previsti (aprile 2004) è altamente probabile. Come da programma, è allo stadio iniziale la progettazione del portale, ma è comunque già attivo un sito web ricco di contenuti.

5. I servizi offerti da CILEA e CASPUR

La tradizione sviluppata dal CILEA e dal CASPUR, due consorzi interuniversitari nati per fornire servizi ad alto contenuto tecnologico, prevede il supporto alle attività di automazione degli atenei, sia in termini di analisi delle esigenze sia per la scelta delle soluzioni più appropriate per soddisfarle. Nell'ambito di questa attività CILEA e CASPUR hanno unito i propri sforzi per fornire ai propri consorziati e ad altri enti servizi di implementazione e *hosting* di archivi aperti per la ricerca e la didattica e consulenza per la loro costituzione, privilegiando l'adozione di *software open source* e l'adesione agli standard internazionali. Il supporto alla costituzione di archivi aperti è fondamentale in una situazione quale quella italiana che dimostra ancora, nonostante alcune pregevoli ma isolate iniziative, tutta la sua arretratezza.

Il progetto AEPIC prevede inoltre l'attivazione di servizi di editoria elettronica per gli atenei facenti capo ai consorzi interessati. E' al momento in corso l'analisi dei prodotti disponibili, che condurrà quanto prima al test degli strumenti che avranno dimostrato di possedere funzionalità avanzate, di adottare standard internazionali e di rispondere all'esigenza primaria degli atenei, ossia la più ampia disseminazione dei risultati della propria ricerca ed il raggiungimento del massimo impatto possibile. Per raggiungere questi obiettivi verranno privilegiate le iniziative di pubblicazione *open access*, il modello più adeguato ai contenuti accademici, e l'impiego di *software open source* per il contenimento dei costi e l'ottemperanza delle direttive ministeriali riguardo ai programmi informatici per le pubbliche amministrazioni.

6. Conclusioni

La costruzione della piattaforma nazionale di servizi per gli archivi aperti si colloca nell'ambito delle attività istituzionali di ricerca e sviluppo nelle quali CILEA e CASPUR sono costantemente impegnati, e costituirà inoltre la base di un progetto europeo in corso di avvio. La comunità internazionale ha infatti rilevato la carenza di *service*

¹⁸ Estensione preannunciata dai partecipanti a KNAW e DARE (Olanda) come OA-X, cfr. <http://eepi.ubib.eur.nl/iliit/archives/000471.html>

L'Archivio E-Prints dell'Università di Firenze: prospettive locali e nazionali

Firenze, 10 febbraio 2004

provider e la necessità di sviluppare strumenti innovativi che incoraggino la diffusione dei *repositories* e la pratica del *self-archiving*, offrendo ai ricercatori la possibilità di aumentare l'impatto dei propri risultati di ricerca e migliorare la visibilità della produzione scientifica, in particolare di quella finanziata con denaro pubblico.

Per tutte queste iniziative CILEA e CASPUR sono aperti alla collaborazione con la comunità nazionale ed internazionale, nella direzione di una condivisione degli strumenti e delle conoscenze al servizio della comunicazione scientifica e dell'accesso aperto alla conoscenza, una missione che CILEA e CASPUR hanno adottato come testimoniano la firma della *Berlin Declaration*¹⁹ e l'adesione a SPARC Europe²⁰.

¹⁹ Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

²⁰ The Scholarly Publishing & Academic Resources Coalition, <http://www.sparceurope.org/>