

AEPIC: una piattaforma di servizi per l'*electronic publishing* e gli *open archives* degli atenei: stato dell'arte - febbraio 2004

Susanna Mornati

CILEA, Segrate

Abstract

La comunità internazionale ha sviluppato strumenti per la creazione di archivi aperti interoperabili, che aumentano la visibilità e l'impatto dei risultati della ricerca condotta nelle istituzioni che se ne avvalgono. Per supportare l'inserimento degli atenei italiani in questo contesto, il CILEA ha ideato nel 2002 AEPIC, 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 didattici e delle tesi, alla pubblicazione di riviste, atti di convegni e monografie. Il presente lavoro intende illustrare lo stato dell'arte del progetto ed in particolare le prime tre realizzazioni in ordine di tempo: E-LIS, un archivio disciplinare per la biblioteconomia e le scienze dell'informazione, DSpace@UniPr, un archivio sperimentale per la didattica umanistica, e una piattaforma di servizi per gli archivi aperti italiani. Infine si illustrano i servizi CILEA per l'implementazione di *repositories* presso le istituzioni accademiche e di ricerca.

Keywords: Electronic publishing, open access, open archives.

Il circuito tradizionale che veicola la diffusione delle pubblicazioni scientifiche mostra da qualche anno i segni di una profonda crisi. Diversi fattori impediscono o rallentano la libera diffusione dei lavori che contribuiscono al progresso scientifico e tecnologico: dalla lentezza delle pratiche tradizionali di pubblicazione alla cessione del copyright ad editori commerciali, dal prezzo crescente dei periodici ai bilanci sempre più ristretti delle biblioteche che li acquistano [1]. Nell'epoca di Internet e dell'*electronic publishing*, il risultato è paradossalmente una progressiva restrizione del pubblico che ha la possibilità di accedere ai risultati della ricerca scientifica.

Queste premesse hanno favorito la nascita di iniziative [2] che sostengono strade alternative per la disseminazione della conoscenza, attraverso la pratica di nuove modalità di comunicazione scientifica da parte degli autori, quali la pubblicazione in riviste *open access* (che non prevedono il pagamento di sottoscrizioni per la lettura degli articoli) e l'auto-archiviazione in *open archives* (*repositories* istituzionali o disciplinari di *eprints*, ossia versioni elettroniche di

articoli e rapporti). La comunità internazionale ha sviluppato inoltre strumenti *software* [3] per la creazione e gestione di archivi aperti, interoperabili mediante il protocollo OAI-PMH [4]. La libera disponibilità dei risultati della ricerca scientifica ne aumenta l'impatto [5], con ampie conseguenze sia per gli autori sia per le loro istituzioni, dall'avanzamento delle carriere individuali alla possibilità di ottenere maggiori finanziamenti per la ricerca, dall'incremento della visibilità dei singoli enti alla promozione del progresso per la società in generale [6].

Per facilitare gli atenei italiani ad inserirsi in questo stimolante contesto internazionale, il CILEA ha ideato nel 2002 il progetto AEPIC, in corso di realizzazione [7].

AEPIC 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 didattici e delle tesi, alla pubblicazione di riviste, atti di convegni e monografie. Il presente lavoro intende illustrare lo stato dell'arte del progetto [8] e in particolare le prime tre realizzazioni in ordine di tempo:

- E-LIS, l'archivio disciplinare internazionale per la biblioteconomia e le scienze dell'informazione realizzato con software GNU EPrints [9];
- DSpace@UniPr, l'archivio sperimentale per la didattica delle Facoltà umanistiche di Parma realizzato con software DSpace del MIT [10];
- una piattaforma di servizi per gli archivi aperti esistenti e futuri, in collaborazione con il CASPUR ed altre istituzioni in Italia e in Europa [11].

La ricchezza delle esperienze maturate ed in corso su ELIS, i tempi record di implementazione di DSpace@UniPr e l'aggiunta di nuove funzionalità avvalorano la tesi della validità di un ambiente collaborativo per la realizzazione di servizi tecnologicamente avanzati: le capacità progettuali dei docenti e del personale universitario unite alle competenze informatiche del personale CILEA danno origine ad iniziative originali e competitive, che ben si inquadrano non solo nell'ambito delle attività di ricerca degli atenei, ma possono contribuire a sostenerne la didattica.

Componenti *open source* per la costruzione di *services providers*

Zeno Tajoli

CILEA, Segrate

Le fondamenta dei *services providers* della piattaforma AEPIC sono costituite da diversi *software open source*, selezionati fra gli strumenti sviluppati nell'ambiente degli *Open Archives*. In particolare si è data rilevanza alle possibili integrazioni tra di loro e alla loro completezza.

Le funzioni principali alla base dei diversi *services providers* di AEPIC sono: harvester, indexer, OPAC, e tools per controlli di uniformità per UTF-8 e XML.

Come Harvester è stato scelto Celestial <<http://oai-perl.sourceforge.net>> Esso ha una buona interfaccia web. Permette di fare scarichi incrementali e permette di usare la funzione ResumptionToken del protocollo OAI-PMH per riprendere gli scarichi interrotti. Durante il test si è evidenziato che per funzionare con perl 5.8.0 e successivi, è necessario modificare il modulo perl XML::DOM su cui Celestial si appoggia per gestire il formato XML.

Come Indexer e OPAC si usa cheshire v2 <<http://cheshire.berkeley.edu/>>, che usa come fonte dei dati sia dei file XML sia delle tabelle di vari RDBMS (PostgreSQL, Illustra, MySQL, etc.). Si tratta di una raffinata applicazione di *information retrieval* che espone i dati sia sul protocollo Z39.50 sia sul protocollo http. Attualmente come fonte dati si usano file XML; se le dimensioni dovessero crescere si passerà a gestire i dati in un RDBMS, probabilmente MySQL. Si è ritenuta molto importante la caratteristica di operare sul protocollo Z39.50 perché esso permette una più facile integrazione con altri sistemi automatizzati. Inoltre cheshire è stato adottato dal progetto inglese RDN <<http://www.rdn.ac.uk>> per scopi analoghi.

E' noto che i dati forniti dai diversi *data providers* presentano spesso problemi dovuti a errate configurazioni degli applicativi usati. Gli ambiti in cui si riscontrano più errori sono la gestione dei caratteri con il set UTF-8 e la conformità ad XML. Per controllare la conformità ad UTF-8 si usa utf8conditioner <<http://www.cs.cornell.edu/people/simeon/software/utf8conditioner/>> sviluppato dalla Cornell University. Per controllare e manipolare l' XML si usa invece un modulo perl, XML::Twig <<http://www.xmltwig.com/>>, in grado di lavorare file XML di notevoli dimensioni.

Lo sviluppo della piattaforma di base è ancora in corso. Una prima versione stabile sarà disponibile verso la fine di marzo 2004, rilasciata con licenza GPL e accessibile dal sito di AEPIC.

Il cuore della piattaforma di servizi per gli archivi aperti è stato sviluppato e testato. I dettagli sono descritti in un *box* che accompagna il presente articolo. Per l'inizio di marzo 2004 è previsto il rilascio del primo prototipo, e per fine marzo 2004 la messa in funzione di un servizio sperimentale, per il quale verrà testata l'aggregazione di archivi disciplinari per la biblioteconomia e le scienze dell'informazione. 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 il terzo *work package*, ossia 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 (quarto *work package*, affidato al CASPUR), ma è comunque già attivo un sito web ricco di contenuti: <http://www.aepic.it/>.

Nell'ambito della propria attività di supporto ad alto contenuto tecnologico, il CILEA fornisce ad atenei ed enti di ricerca servizi di implementazione e *hosting* di archivi aperti per la ricerca e la didattica e consulenza per la loro costituzione. 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 altri servizi di editoria elettronica per gli atenei. 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 della direttiva ministeriale [12] riguardante lo sviluppo ed l'utilizzazione dei programmi informatici da parte delle pubbliche amministrazioni.

Per tutte queste iniziative il CILEA è aperto 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 il CILEA ha adottato come testimonia la firma della "*Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*" [13].

Bibliografia e note

- [1] Fra i tanti lavori dedicati all'argomento, si veda 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, URL: http://www.ala.org/Content/NavigationMenu/ACRL/Publications/College_and_Research_Libraries_News/Back_Issues_2003/February1/Removing_barriers_to_research.htm
- [2] Ad esempio la Budapest Open Access Initiative, nota come BOAI, URL: <http://www.soros.org/openaccess/index.shtml>
- [3] GNU EPrints, <http://www.eprints.org/>, DSpace <http://www.dspace.org/>, e altri descritti in *A guide to institutional repository software*, v. 2.0, URL: http://www.soros.org/openaccess/pdf/OSI_Guide_to_Institutional_Repository_Software_v2.pdf
- [4] Open Archives Initiative - Protocol for Metadata Harvesting, sviluppato nell'ambito dell'Open Archives Initiative, URL: <http://www.openarchives.org/OAI/openarchivesprotocol.html>
- [5] Steve Lawrence, *Free online availability substantially increases a paper's impact*, in *Nature*, 411, 6837, p. 521, 2001. Versione "open" (*Online or Invisible*), URL: <http://www.neci.nec.com/~lawrence/papers/online-nature01/>
- [6] Stevan Harnad, *Promoting Self-Archiving of Institutional Research Output*, URL: <http://www.ecs.soton.ac.uk/~harnad/Temp/self-archiving.htm>
- [7] Illustrato inizialmente da Valentina Comba, AEPIC Academic E-Publishing Infrastructures - CILEA: progetto di editoria elettronica per la ricerca e la didattica, 2002, URL: <http://eprints.rclis.org/archive/00000066/>, e Il progetto AEPIC "Academic E-Publishing

Infrastructures CILEA”: le esigenze degli autori e lo stato dell’arte dei servizi, maggio 2003, URL:

<http://eprints.rclis.org/archive/00000181/>

- [8] Il presente lavoro è basato su materiali già utilizzati per una presentazione al Convegno EPrints, Firenze, 10 febbraio 2004. I link qui citati sono stati visitati per l’ultima volta il 19 febbraio 2004
- [9] <http://eprints.rclis.org/>, cfr. Antonella De Robbio, ELIS: un Open Archive in Library and Information Science, in *Bibliotime*, VI (I), marzo 2003, URL:
<http://eprints.rclis.org/archive/00000201/>, e De Robbio A., Giuratrabocchetti F., Tentoni P., E-LIS: nuovo servizio E-Prints in Library and Information Science presso il CILEA, in *Bollettino del CILEA* n.86, 2003, URL:
<http://cdl.cilea.it/routing.asp?ISSN=11202440>
- [10] <http://dspace-unipr.cilea.it:8080/>
- [11] Per maggiori dettagli sulle caratteristiche della piattaforma cfr. il project plan, URL:
<http://eprints.rclis.org/archive/00000518/>
- [12] Direttiva del 19 dicembre 2003, pubblicata sulla Gazzetta Ufficiale n. 31 del 7 febbraio 2004.
- [13] Testo, URL:
<http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>