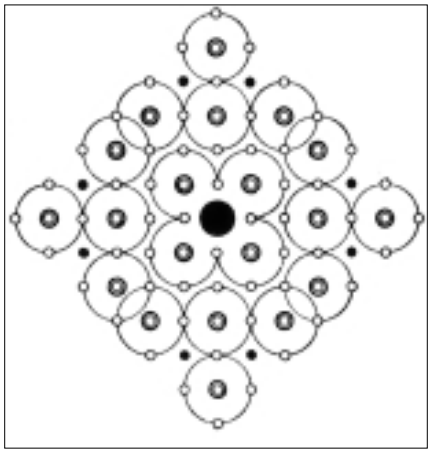
Archivi di documenti elettronici

Un modello di riferimento per la realizzazione della "biblioteca ibrida"

di Valdo Pasqui

1. Introduzione

Nel mese di aprile 2000 l'Università di Firenze ha avviato il progetto Editoria scientifica con la finalità di studiare, definire e sperimentare il modello normativo, contrattuale, economico e tecnologico riguardante le iniziative editoriali operanti nell'ambito dell'ateneo fiorentino, compresa la produzione di materiale didattico, ai fini di istituire la Firenze University Press. Il progetto, oltre ad effettuare una fase di ricognizione generale, si è raccordato con iniziative riconducibili al suo ambito, in particolare quelle del Polo scientifico di Sesto (comprendente laboratori e centri afferenti alle aree di chimica, fisica e agraria) che aveva già posto in luce l'esigenza di realizzare servizi per la pubblicazione di documenti in formato elettronico.
In questo contesto, che si sta rapidamente allargando ad altre aree (es. scienze sociali e giuridiche), è emersa la necessità di creare archivi di e-print alimentati da parte degli autori (*self-archiving*) e accessibili agli utenti finali, primariamente quelli istituzionali dell'ateneo (docenti, ricercatori e studenti).



Si è ritenuto fondamentale approcciare questa problematica attraverso la definizione di un framework di riferimento capace di: a) preservare l'autonomia, sia dal punto di vista gestionale che da quello dei contenuti, delle iniziative attuali e future di facoltà, dipartimenti e centri; b) garantire l'interoperabilità tra gli archivi in corso di costituzione; c) consentire agli utenti finali l'accesso integrato alle risorse, in particolare l'esistente catalogo in linea delle biblioteche (OPAC), le collezioni di risorse digitali primarie e i *subject information gateways*. Questi infatti sono gli assi secondo i quali assisteremo in breve tempo allo sviluppo della "biblioteca ibrida".
Il primi due paragrafi forniscono un sintetico ragguaglio dell'ambito di riferimento e descrivono il modello in corso di definizione da parte della Open Archives Initiative (OAI). Successivamente vengono analizzate le categorie di documenti individuate e viene proposto un modello di riferimento basato sulla OAI per la realizzazione di servizi e-print integrati in ambito accademico. Infine, è fornita una breve descrizione dei componenti architetturali e della piattaforma tecnologica.

2. Ambito di riferimento

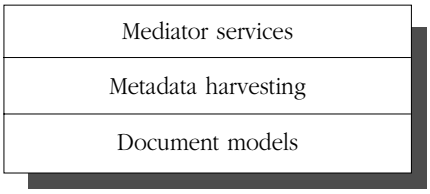
Nel tipico ambiente aziendale le collezioni di documenti elettronici vengono di norma gestite in accordo a rigidi standard interni, regole di check-in/check-out e controlli di accesso, avvalendosi di software commerciali (es. strumenti di *workflow management*) e operando nei limiti delle intranet aziendali. Il contesto accademico presenta alcune caratteristiche peculiari che lo rendono abbastanza diverso dal precedente:
– il numero elevato di utenti finali (docenti, studenti, esterni) che possono accedere sia dalla rete di ateneo (di solito geograficamente distribuita) che dall'esterno (es. da casa);
– la varietà dei tipi di documenti da gestire (cfr. par. 4.1), del loro ciclo di vita e dei loro formati;
– la necessità di definire un'architettura flessibile, composta da più archivi indipendenti e autonomi, ma facilmente integrabili per garantire l'accesso uniforme agli utenti;
– l'importanza di favorire l'interoperabilità, per ambiti tematici, con altre iniziative di livello nazionale e internazionale;
– la necessità di garantire l'integrazione con le risorse ed i servi-

zi esistenti, in particolare i cataloghi delle biblioteche ed altre risorse primarie o secondarie disponibili.
In ambito internazionale, nell'area della ricerca scientifica, sono da tempo operativi alcuni servizi che costituiscono punti di riferimento sia per l'approccio seguito che per le soluzioni tecnologiche usate. Tra questi, segnaliamo alcuni dei più significativi.
ArXiv, sviluppato e ospitato fin dal 1991 dal Los Alamos National Laboratory (LANL), è uno dei primi esempi di archivi e-print avente per tematiche la fisica, la matematica e la computer science. Il mirror italiano è gestito dalla Scuola internazionale superiore di studi avanzati di Trieste (SISSA).
Networked Computer Science Technical Reference Library (NCSTRL) è una raccolta internazionale di lavori e rapporti nell'ambito della computer science resi disponibili per uso non commerciale da un numero elevato di istituzioni. NCSTRL ha sviluppato l'architettura DIENST[1] che comprende il modello dei documenti, un insieme di servizi, il protocollo DIENST e i meccanismi per combinare i servizi ed il contenuto delle collezioni in ambito distribuito. Questa tecnologia è stata adottata in Italia dal progetto del CNR Digital Library[2] che fa parte della ERCIM Technical Reference Digital Library (ETRDL), costituita nell'ambito dell'European Research Consortium for Informatics and Mathematics (ERCIM).
CogPrints (Cognitive Sciences Eprint Archive), comprende documenti su psicologia, linguistica e neuroscienze, è ospitato dall'Electronics and Computer Science Department della Università di Southampton che è tra i promotori della Open Archives Initiative.
CERN Document Server propone un modello basato su un unico server bibliografico, contenente i dati descrittivi di tutte le collezioni sia cartacee che elettroniche, un server per i documenti full-text (pre-print e rapporti interni) e un'interfaccia integrata (WebLib) per la ricerca di tutto il materiale disponibile, compreso il catalogo della biblioteca.
Un'analisi degli aspetti tecnico-funzionali più rilevanti di questi servizi consente di individuare alcuni elementi comuni:
– ogni documento ha un identificativo univoco ed è accompagnato da un insieme di dati descrittivi (metadati) di solito immessi dall'autore al momento della sottomissione del documento all'archivio;
– a tal fine esistono delle interfacce per gestire la registrazione degli autori e per consentire loro l'immissione e il ritiro dei documenti;
– l'interfaccia di ricerca per gli utenti finali è piuttosto semplice, con un numero limitato di campi tra i quali compaiono sempre titolo, autore e abstract; spesso si può cercare anche per identificativo del documento o si può selezionare l'archivio di appartenenza (se i documenti sono organizzati in archivi diversi);
– la memorizzazione dei documenti è demandata a server dedicati (nel caso di DIENST viene usata la struttura del file system del sistema ospitante);
– i metadati sono indicizzati e gestiti mediante software dedicati (DIENST propone la freeWAIS search engine, mentre il CERN usa il sistema bibliotecario commerciale ALEPH);
– un componente autonomo supporta le funzioni di ricerca e gestisce le richieste degli utenti, attraverso interfacce di tipo www;
– al momento nessuno supporta la ricerca sul full-text dei documenti che è prevista in futuro per il CERN.
Aspetto peculiare dell'architettura DIENST è di essere stata concepita per favorire la ricerca su un insieme distribuito di archivi. A tal fine DIENST supporta la raccolta (*harvesting*) dei metadati da più server di documenti (*repository*), la loro indicizzazione in appositi server (*index-server*), l'esecuzione distribuita delle interrogazioni (*query routing*) e la fusione (*merge*) dei risultati.
Caratteristica specifica del CERN Document Server è di aver affidato ad un componente dedicato (*set-link manager*) la gestione degli "indirizzi di lunga durata" assegnati ai documenti e la loro traduzione negli URL effettivi.

3. Open Archives Initiative

Nell'ottobre del 1999 a Santa Fe ha avuto luogo un meeting (sponsorizzato da rilevanti organizzazioni a livello mondiale) avente per obiettivo la discussione e la definizione di soluzioni per l'interoperabilità tra archivi di e-print[3] ai fini della creazione di uno "scholarly communication system" globale. Dal meeting ha avuto origine la Open Archives Initiative (OAI)[4] che ha definito un modello articolato in tre livelli (cfr. fig. 1).

Fig. 1 - Modello di interoperabilità



I *document models* definiscono la struttura e i formati dei documenti, il *metadata harvesting* si occupa dell'estrazione dei dati descrittivi e i *mediator services* comprendono i servizi per gli utenti (ricerca, discovery, accesso, ecc.) costruiti sulle informazioni estratte dal livello sottostante. Ai fini dell'interoperabilità tra archivi di e-print, la specifica iniziale si è concentrata sul livello intermedio (*metadata harvesting*) definendo una chiara distinzione tra:
a) *data provider* che gestisce ➤

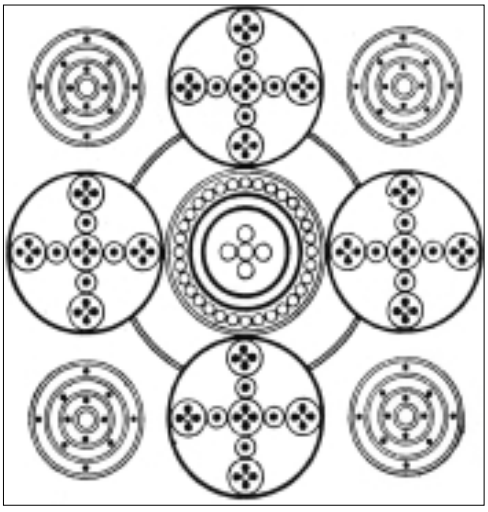
un archivio e-print per conto degli autori che sottopongono i documenti e che include:
– il meccanismo di presentazione (*submission*);
– il sistema di memorizzazione a lungo termine;
– la politica di gestione concernente la presentazione e la conservazione dei documenti;
– un'interfaccia aperta per facilitare la raccolta di dati dall'archivio da parte dei *service provider* che gestiscono i servizi di ricerca e discovery;
b) *service provider*, una terza parte che crea e gestisce i servizi per la disseminazione dei contenuti basati sui dati memorizzati negli archivi e-print. In pratica si tratta dei fornitori dei servizi di ricerca ed accesso per gli utenti finali (es. implementando una *search engine* e la relativa interfaccia utente).
L'interfaccia nativa per l'accesso da parte degli utenti finali è una componente opzionale degli archivi dei *data provider* che invece devono avere un identificatore univoco e contenere una collezione di record che soddisfino i seguenti requisiti:
– ogni record ha un identificatore univoco e persistente;
– un record contiene almeno i metadati che descrivono il contenuto del documento;
– un record può contenere anche il riferimento al contenuto (testo, software, ecc.);
– esiste un insieme base condiviso di metadati (Dublin Core Element Set), rappresentati e trasferiti in sintassi XML, anche se è possibile esportare altri insiemi specifici di metadati (es. usati in determinate comunità come quella degli e-print o dei musei).
Inoltre i *data provider* devono supportare un'interfaccia di *harvesting* che consenta ai *service provider* di recuperare informazioni sul contenuto dell'archivio e i suoi record; si tratta di un protocollo derivato dal protocollo DIENST e trasportato mediante il protocollo standard del web (HTTP).
Una terza componente del modello riguarda l'uso dei dati e prevede che i fornitori di servizi rispettino le restrizioni imposte e formalizzate dai fornitori di dati. Inoltre, la OAI distingue la raccolta dei metadati da quella del contenuto (*full-content*) che può essere soggetta a specifici controlli e protetta tramite l'uso di opportuni protocolli.
Molte delle organizzazioni che aderiscono alla OAI hanno espresso l'intenzione di rendere i loro archivi compatibili con queste specifiche (es. arXiv, CogPrints, NCSTRL, University of California Digital Library). OAI è ancora in fase di evoluzione:[5] in poco meno di un anno (Santa Fe, ottobre 1999 – Ithaca, settembre 2000) il focus si è spostato dall'obiettivo iniziale di fornire le specifiche di interoperabilità per archivi di e-print autoalimentati dagli autori a quello di definire un framework di interoperabilità per gli archivi di dati in ambito accademico, basato sul principio della raccolta (*harvesting*) di metadati e denominato Open Archives Harvesting Framework Specifications.[6] Nonostante i prevedibili assestamenti, la OAI è un'iniziativa rilevante destinata a fornire le linee guida per la realizzazione di archivi e-print e a promuovere l'interoperabilità di archivi di dati attraverso lo sviluppo di servizi di ricerca e accesso all'informazione (*search engines*, *subject information gateways*, ecc.).

4. Un modello di riferimento per gli archivi di ateneo

L'analisi di come sviluppare servizi basati su archivi e-print nell'ambito specifico del Polo scientifico di Sesto e in quello più ampio della Firenze University Press (FUP) ha compreso l'individuazione delle tipologie dei documenti da gestire e di un modello di riferimento capace di garantire sufficiente flessibilità e modularità.

4.1 I documenti e le modalità di accesso

Comunemente gli archivi e-print contengono i pre-print di articoli di riviste e i rapporti tecnici (cfr. par. 2), tuttavia l'ambito accademico propone una gamma più ampia di documenti in formato elettronico gestibili mediante questi archivi. Le categorie di documenti individuate comprendono:
– monografie e atti di convegni;
– riviste in formato elettronico;
– pre-print di articoli di riviste;
– rapporti tecnici;
– dispense e materiale didattico;
– pubblicazioni multimediali;
– collezioni di documenti digitalizzati;
– tesi e dissertazioni.
Attualmente la Firenze University Press sta sperimentando, in collaborazione con l'editore Casalini, la pubblicazione di alcune monografie, atti di convegni e riviste in formato elettronico con modalità di accesso di tipo *print on demand*. Alla fine del progetto in corso verranno prese le opportune decisioni su come proseguire e consolidare questa attività dalla quale sta



nascendo un primo archivio elettronico di pubblicazioni edite dall'ateneo fiorentino.
Gli archivi di pre-print sono di solito organizzati per aree tematiche. È in corso di valutazione la possibilità di affidare la pubblicazione dei pre-print di alcune aree d'interesse (es. fisica, matematica) ad autorevoli servizi già esistenti a livello nazionale ed europeo (SISSA, Università di Gottinga). Per altre aree si prevede la creazione di archivi locali che, oltre a contenere gli articoli prima della pubblicazione, opereranno anche come database di servizio per supportare l'attività di peer-review e potranno continuare a ospitare i testi dopo la loro pubblicazione (post-print) sulle relative riviste.
Dispense, materiale didattico e rapporti tecnici costituiscono un insieme eterogeneo sia per i formati dei documenti che per le modalità di accesso. Al momento la FUP ha previsto di distinguere tra "dispense tradizionali" e "appunti di lezione". Le prime verranno incluse nella categoria monografie e verrano rese disponibili attraverso un servizio di tipo *print on demand*. La seconda tipologia di documenti comprende il materiale di supporto di corsi e seminari (appunti, test, esercitazioni, riferimenti bibliografici) attualmente distribuito dai docenti in modo informale (su carta, dischetto, ecc.). In questo caso è previsto che gli utenti finali (studenti) accedano in linea a questi documenti in modo controllato, tramite meccanismi di autenticazione (identificativo e password) già predisposti per altri servizi esistenti (i.e. consultazione dati della carriera). Quest'ultimo insieme di oggetti, unitamente ad alcuni rapporti interni, costituirà il contenuto di un primo prototipo di archivio e-print che verrà realizzato nei prossimi mesi nell'ambito del Polo scientifico di Sesto.
Nel contesto degli archivi e-print rientrano anche le pubblicazioni in linea di tipo multimediale che per funzionalità e complessità vanno oltre gli ipertesti in formato elettronico. Di questa tipologia di materiale è previsto uno sviluppo sempre maggiore, in particolare a supporto dell'attività didattica. La creazione di archivi di metadati che descrivano questi oggetti è rilevante per renderli accessibili anche attraverso i servizi di ricerca delle risorse elettroniche in rete.
Per gli stessi motivi alcune collezioni di documenti digitalizzati (es. testi rari o di pregio), di cui è prevista la realizzazione nei prossimi anni, saranno costruite non come archivi a sé stanti, ma in modo che i relativi metadati possano essere integrati con le altre risorse informative dell'ateneo.
Le tesi e le dissertazioni sono documenti specifici dell'ambito accademico, facilmente assimilabili dal punto di vista delle modalità di ricerca e di accesso agli archivi di e-print.[7] Per l'ateneo fiorentino, in particolare nel Polo scientifico di Sesto, verrà verificata la possibilità di creare un archivio di questo materiale da integrare con l'esistente catalogo dei titoli delle tesi di laurea e diploma.

4.2 Il modello

La sintesi riportata nel paragrafo precedente mostra chiaramente che l'ambiente accademico possiede una gran varietà di documenti e che per poterli organizzare in archivi elettronici è necessario un modello che:
a) garantisca alle unità amministrative (facoltà, dipartimenti, centri) la più ampia autonomia organizzativa;
b) consenta diverse modalità di aggregazione (es. per categorie, per aree tematiche);
c) permetta lo sviluppo di servizi di ricerca ed accesso capaci di fornire agli utenti finali l'accesso integrato a tutte le risorse disponibili in ateneo.
Alla luce di quanto esposto nei paragrafi precedenti, il modello e le specifiche di interoperabilità della Open Archives Initiative sono pienamente rispondenti a queste esigenze, infatti tale approccio:
– consente di realizzare archivi distinti, autonomi e gestiti separatamente (livello *data provider*);
– garantisce l'interoperabilità tra archivi diversi anche in ambito nazionale e internazionale (es. per aree tematiche);
– permette di realizzare servizi per gli utenti finali (livello *service provider*) dedicati o integrati con altre risorse informative, consentendo quindi l'integrazione degli archivi e-print nei servizi di tipo *subject information gateway*;
– è neutro rispetto al formato dei documenti, lasciando la più ampia libertà (HTML, PDF, MS WORD, ecc.) e quindi favorendo anche l'evoluzione verso l'adozione dei nuovi standard XML (Extensible Markup Language) e DOM (Document Object Model);
– permette l'integrazione con eventuali servizi esistenti per supportare i meccanismi di controllo dell'accesso;
– consente lo sviluppo di un'ampia gamma di archivi, anche non strettamente di tipo e-print, garantendone l'interoperabilità;
– permette la creazione di un servizio (livello *service provider*) che integra i dati degli archivi e-print con l'OPAC, secondo il modello di interfaccia di ricerca unica per tutte le tipologie di risorse disponibili proposto, sebbene con approcci organizzativi e tecnologici diversi, da WebLib del CERN e dal progetto europeo EULER.[8]
Un modello di riferimento basato sulla OAI offre la possibilità di sviluppare diversi tipi di collezioni digitali e realizzare un insieme di servizi integrati destinati a costituire la futura biblioteca ibrida. ➤

Fig. 2 - Archivio ed interfaccia sullo stesso server

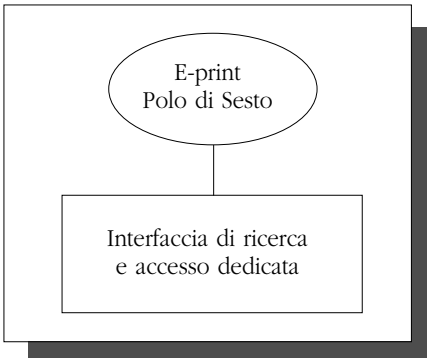
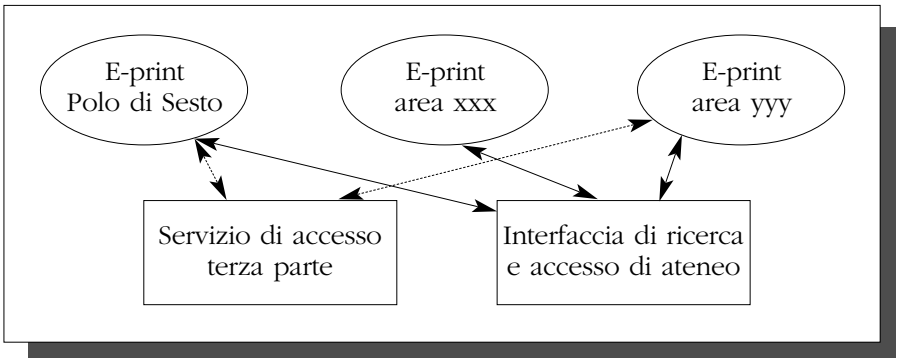


Fig. 3 - Archivi indipendenti (server separati) ed interfaccia di accesso di ateneo unica per gli e-print con possibile accesso di alcuni archivi e-print da servizi esterni



Le figure 2, 3, 4 mostrano la rappresentazione schematica di alcuni scenari possibili, tra loro compatibili e complementari, individuati nel contesto dell'Università di Firenze.

5. Componenti e piattaforme tecnologiche

Dal punto di vista architetturale i componenti base del modello individuato possono essere cosi ripartiti:

a) livello *data provider*
– sistema di memorizzazione dei documenti;
– database dei metadati (Dublin Core);
– interfaccia per immissione/ritiro documenti da parte degli autori, compresa la creazione dei metadati (di tipo www);
– meccanismo di registrazione degli autori (con interfaccia www);
– interfaccia per *harvesting* (conforme al protocollo Open Archives Initiative);
– interfaccia di ricerca per utenti finali (opzionale).
b) livello *service provider*
– interfaccia di *harvesting* (conforme al protocollo Open Archives Initiative);
– indicizzazione dei metadati;
– conversione e indicizzazione del full-text (opzionale);
– interfaccia di ricerca per utenti finali.

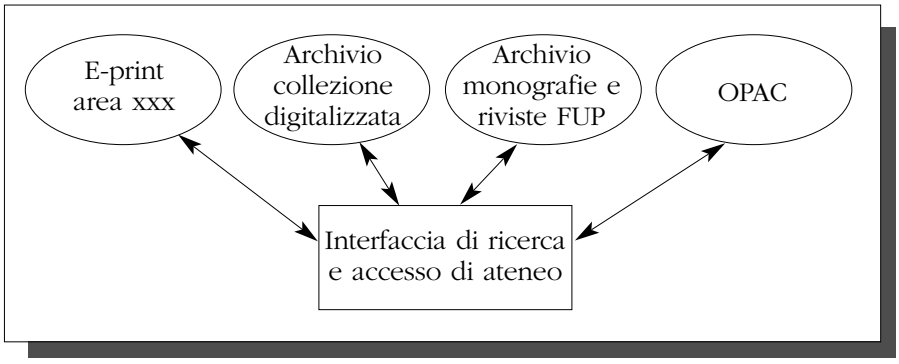
Grazie alla flessibilità del modello di riferimento, varianti ed integrazioni del precedente schema possono essere facilmente individuate.
Al livello *data provider*, nel caso in cui i documenti immessi siano soggetti ad un'attività di peer-review, il componente che gestisce l'immissione dei documenti può includere dei meccanismi per bloccarne la visibilità ed esportabilità (compresi i metadati) fin quando il documento non abbia ottenuto l'approvazione. Nei contesti che prevedono una fase di intermediazione per la revisione e la composizione dei testi e per curare l'immissione dei metadati, l'interfaccia per l'immissione/ritiro dei documenti non è disponibile agli autori. Queste funzioni sono invece riservate a gruppi selezionati di esperti: per esempio i bibliotecari possono occuparsi della descrizione dei documenti (generazione dei metadati). Inoltre, in questa ipotesi, il componente che gestisce i meccanismi di registrazione degli autori non è più necessario.
Se l'archivio contiene collezioni di materiale digitalizzato, il componente "sistema di memorizzazione dei documenti" deve gestire oggetti più complessi rispetto al tipico file di testo dei pre-print (es. gruppi di pagine scannerizzate). In questo caso si prevede anche la memorizzazione di insiemi di metadati diversi da Dublin Core e appartenenti a schemi specifici per la tipologia della collezione.
Al livello *service provider*, eventuali servizi di tipo *subject information gateway* possono essere opportunamente integrati con i servizi di livello *data provider* attraverso una raccolta selettiva dei metadati per area tematica. Questi servizi possono limitarsi a indicizzare i metadati o "catturare" anche i documenti per effettuare l'indicizzazione del full-text. Infine, può essere realizzato un servizio di ricerca integrato con l'OPAC di ateneo che fornisca agli utenti finali (in particolare gli studenti) un'interfaccia unica per l'accesso a tutte le collezioni disponibili (sia elettroniche che cartacee).
I meccanismi di controllo dell'accesso possono essere implementati sia al livello *data provider* che al livello *service provider* o soltanto a questo livello, se gli archivi sono mantenuti completamente nascosti agli utenti finali. La piena integrazione con servizi di controllo e autenticazione eventualmente già disponibili presso l'istituzione (es. controllo degli studenti e di un ateneo o degli utenti di una biblioteca) è raccomandata.
Il modello proposto non impone vincoli specifici per la scelta della piattaforma tecnologica, tuttavia, in particolare nell'ambiente accademi-

co, si ritiene fondamentale l'adozione di soluzioni di tipo aperto che consentano l'uso di software pubblico e gratuito. Per quanto concerne il sistema operativo, è largamente da preferire Unix, in particolare Linux che opera anche su server con processori di tipo Intel, perché è praticamente gratuito, soddisfa gli standard di sicurezza e supporta una vasta gamma di prodotti gratuiti e scaricabili dalla rete. Nell'ambito dei database si stanno rapidamente diffondendo strumenti come MySQL che sono gratuiti se usati per fini non commerciali. Infine, server www gratuiti come Apache e strumenti di programmazione quali Perl, TCL, Java e PHP sono comunemente usati per la costruzione dei siti web. L'Università di Southampton, nell'ambito dell'iniziativa Eprints.org, ha sviluppato un software per la creazione di archivi e-print conformi alla OAI, reso pubblico alla fine di settembre 2000, liberamente scaricabile[9] ed interamente realizzato con software di tipo gratuito.

6. Conclusioni

La diffusione di Internet e del www ha favorito lo sviluppo dei servizi che consentono l'accesso in rete a documenti elettronici di rilevante interesse per le attività didattiche e di ricerca. Un approccio di tipo minimalista, basato sull'uso di spazio dedicato all'interno di siti web (es. di un docente/autore o di un dipartimento), non può soddisfare in pieno le esigenze del mondo accademico. Le esperienze internazionali mostrano che un contesto più strutturato, basato sull'uso di metadadati, è in grado di fornire servizi migliori sia agli autori che agli utenti finali. Inoltre, lo sviluppo di servizi integrati per la ricerca e l'accesso alle risorse elettroniche necessita della massima interoperabilità tra gli archivi di e-print, sia a livello locale istituzionale (es. di ateneo) che nazionale e internazionale.
Il modello in corso di definizione nell'ambito della Open Archives Initiative include una serie di caratteristiche (differenziazione dei servizi in due livelli, adozione dello standard di metadati Dublin Core, protocollo di *harvesting* semplice e trasportato su HTTP, rilevanza delle istituzioni coinvolte) che ne fanno un punto fisso per la creazione di servizi basati su archivi di documenti elettronici. Tale modello fornisce il quadro di riferimento in cui è possibile sviluppare insiemi integrati di collezioni digitali capaci di mantenere piena autonomia gestionale e organizzativa, aspetti fondamentali nell'ambito accademico. L'integrazione in questo schema di altre risorse primarie e secondarie e dell'OPAC è il passo successivo che consentirà di realizzare la "biblioteca ibrida", capace di fornire ai suoi utenti una visione integrata e completa di tutte le risorse informative disponibili.
L'Università di Firenze sta orientando in tal senso le iniziative recentemente intraprese. In particolare, lo sviluppo di un archivio prototipale contenente il materiale didattico, prodotto da alcuni docenti che operano nel Polo scientifico di Sesto, potrà successivamente essere replicato o esteso per gestire il materiale concernente altre aree d'interesse (es. giuridico-sociale). L'adozione del modello OAI consentirà l'integrazione di questi archivi con quelli della Firenze University Press. Alcuni progetti di prossima attuazione per la creazione di collezioni digitali, derivate da specifici fondi librari delle biblioteche, si baseranno sul modello illustrato al fine di consentire l'interoperabilità di tutti gli archivi e la realizzazione di un sistema di ricerca e accesso alle risorse completamente integrato. ■

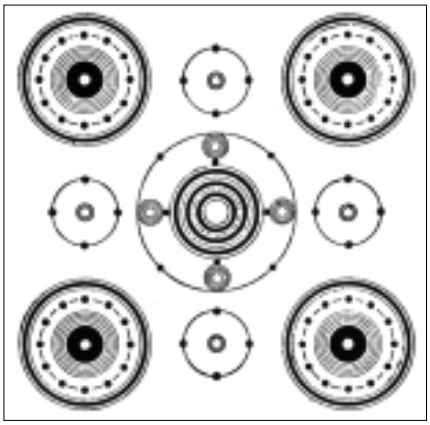
Fig. 4 - Archivi indipendenti, server FUP, OPAC ed interfaccia unificata per l'accesso alle risorse dell'ateneo



L'autore ringrazia Anna Maria Tammaro, coordinatore del progetto Editoria elettronica, per i suggerimenti e per aver promosso questo studio.

Note

[1] *Dienst architecture summary description* (ultimo aggiornamento 29/02/2000), <http://www.cs.cornell.edu/cdlrg/dienst/architecture/architecture.html]>.
[2] *The CNR Digital Library*, <http://www.area.pi.cnr.it/IT/LIB/PROJECT/info.html>.
[3] H. Van de Sompel, T. Krichel, M.L. Nelson et altri, *The UPS prototype: an experimental end-user service across e-prints archives,* "D-Lib Magazine", 6, February 2000, 2, <http://www.dlib.org/dlib/february00/vandesompel-ups/02vandesompel-ups.html>.
[4] H. Van de Sompel, C. Lagoze, *The Santa Fe Convention of the Open Archives Initiative*, "D-Lib Magazine", 6, February 2000, 2, <http://www.dlib.org/dlib/february00/vandesompel- ➤



oai/ 02vandesompel-oai.html>; *The core document of the Santa Fe Convention,* <http://www.openarchives.org/sfc/sfc.htm>.

[5] La prima specifica della OAI prevedeva un insieme base di metadati denominato Open Archives Metadata Set (OAMS) e composto da nove elementi di Dublin Core. Analogamente il protocollo inizialmente previsto, denominato Open Archives Dienst Subset, era un sottoinsieme del protocollo Dienst.

[6] *Report on Open Archives Initiative technical committee Meeting,* Ithaca NY, 7-8 September 2000, <http://www. openarchives.org/oai-tech-cornell/cornell_report.pdf>.

[7] Al riguardo sono di rilevante interesse il Networked Digital Library of Theses and Dissertation (NDLTD) e la sua estensione Networked University Digital Library (NUDL) che includono molte istituzioni accademiche (in particolare, per gli aspetti tecnologici, Virginia Tech EDT) anche europee (es. Dissertation Online Effort in Germania) che hanno manifestato l'intenzione di adottare la OAI.

[8] EULER project, <http://opac.unifi.it/euler/>.

[9] About the Eprints.org software, <http://www.eprints.org/software.html>.]

Servizi citati

ArXiv: <http://arXiv.org/>.

CERN Document Server: <http://weblib.cern.ch//Home/index.php>.

Computer Science Teaching Center (CSTC), Virginia Tech: <http://www.cstc.org/>.

CogPrints: <http://cogprints.soton.ac.uk/>.

Eprints.org: <http://www.eprints.org/>.

ERCIM: <http://www-ercim.inria.fr/>.

Firenze University Press: <http://www.unifi.it/universita/biblioteche/gruppi/g-editoria.htm>.

NCSTRL: <http://www.ncstrl.org>.

NDLTD: <http://www.ndltd.org>.

Open Archives Initiative, <http://www.openarchives.org>.

SISSA: <http://it.arxiv.org/>.

Università di Firenze, Catalogo in linea dei titoli delle tesi di laurea e diploma: <http://stud.unifi.it:8080/>.