

Sistemi di biblioteche digitali per la ricerca e la didattica: l'esperienza del CNR-ISTI

Stefania Biagioni e Donatella Castelli
Istituto di Scienza e Tecnologie dell'Informazione
"A Faedo" – CNR
stefania.biagioni@isti.cnr.it
donatella.castelli@isti.cnr.it

1. Biblioteche digitali

La sempre maggiore disponibilità di documentazione in formato digitale ha sollecitato negli ultimi anni la costruzione e diffusione di biblioteche digitali. Esse presentano numerosi vantaggi rispetto alle biblioteche tradizionali. Innanzitutto sono molto più economiche: consentono, infatti, di ridurre i costi di produzione dei documenti, di distribuire i costi della catalogazione e di eliminare i costi legati alla gestione manuale e degli spazi fisici di una biblioteca. Consentono inoltre una disseminazione molto più ampia e veloce: chiunque abbia accesso ad Internet può, previa autorizzazione, ricercare ed accedere la documentazione in formato digitale; una volta "pubblicata" dalla biblioteca digitale, il documento diventa immediatamente accessibile a tutti gli utenti, ovunque essi siano fisicamente dislocati.

Le biblioteche digitali possono essere geograficamente distribuite, perciò servizi quali archiviazione, catalogazione, memorizzazione e mantenimento dei documenti, possono essere gestiti da diverse istituzioni distribuite sul territorio; il concetto di biblioteca unica centrale è in questo caso solo una nozione virtuale percepita dall'utente finale.

Infine, le biblioteche digitali sono capaci di gestire e rendere accessibili in modo integrato non solo descrizioni catalografiche e documenti testuali, ma anche immagini, video, registrazioni sonore, ecc. La natura digitale dei documenti e la disponibilità di una infrastruttura di rete rendono possibile la realizzazione di numerosi servizi di supporto alle attività dei fruitori della biblioteca, quali, ad esempio, la disseminazione personalizzata guidata dal profilo dell'utente.

Nei documenti programmatici della Commissione Europea, le biblioteche digitali sono viste come una nuova infrastruttura che, con l'integrazione e l'uso dei mezzi di calcolo, della comunicazione e dei contenuti digitali in scala globale, è destinata a diventare parte essenziale delle infrastrutture informative del ventunesimo secolo: le biblioteche digitali metteranno a disposizione di tutti i cittadini europei i contenuti culturali e scientifici e conserveranno per le generazioni future il sapere e i risultati della ricerca.

2. L'attività dell'ISTI

L'Istituto CNR-ISTI è l'istituto informatico che per primo in Europa ha iniziato a lavorare nel settore delle biblioteche digitali soprattutto con le attività di ricerca e sviluppo del laboratorio "Sistemi informativi multimediali in rete - Multimedia Networked Information Systems" e della Biblioteca di Informatica. Negli ultimi anni esso ha coordinato diversi progetti finanziati dalla Comunità Europea, dal Ministero italiano per l'Educazione, l'Università e la Ricerca (MIUR) e dallo stesso CNR.

Le attività dell'ISTI nell'ambito delle biblioteche digitali iniziano nel 1995 con la creazione del Gruppo di Lavoro sulle Biblioteche Digitali dell'European Research Consortium for Informatics and Mathematics–ERCIM. Nel 1995, a Budapest, durante il primo incontro fra bibliotecari ed informatici, vengono gettate le basi per la realizzazione di una Biblioteca Digitale finalizzata alla diffusione della letteratura scientifica non pubblicata del settore Information Technology, in collaborazione con la Cornell University che metteva a disposizione il software con cui era stata realizzata la Networked Computer Science Technical Reference Library –NCSTRL. Lo stesso ERCIM, nell'anno successivo, incarica l'ISTI di coordinare la Ercim Digital Library Initiative, iniziativa che ha il compito di creare una comunità di ricerca nel campo delle biblioteche digitali in Europa. E' nell'ambito del Delos Working Group on Digital Library¹ che viene realizzata, nel 1997, la Ercim Technical Reference Digital Library –ETRDL (<http://dienst.iei.pi.cnr.it/>) alla cui nascita hanno partecipato 8 Istituti ERCIM.

ETRDL è una collezione specializzata di NCSTRL con funzionalità estese per soddisfare le necessità dell'ambiente di ricerca europeo, ha per es. interfacce multilingua e permette ricerche avanzate su più campi. Ha una architettura distribuita e multiformato, realizza un modello alternativo di editoria scientifica, offre i servizi tradizionali di una biblioteca su documenti tradizionali in formato digitale, il che consente una disseminazione più rapida, più ampia e meno costosa dei documenti. E' inoltre uno strumento per la creazione di collezioni di qualità controllata di documenti immediatamente accessibili in rete e gestiti dagli autori stessi, e dalle loro istituzioni nella veste di produttori ed editori delle informazioni.

Attualmente è un servizio attivo nell'area informatica e matematica. Sono state realizzate due istanze personalizzate di ETRDL nel campo della Biologia marina: Biblioteca Digitale della Laguna di Venezia (<http://archimede.isti.cnr.it:81/intro.htm>) e nel campo della fisica: biblioteca digitale tematica finalizzata a raccogliere la documentazione dei partecipanti al Progetto europeo "Infrastructure Cooperation Network SENTINEL" (<http://archimede.isti.cnr.it>).

Dal 1997 l'ISTI diventa il punto di riferimento delle attività di progetto e di diffusione delle conoscenze finanziate dalla Comunità Europea nel campo delle biblioteche digitali. Alla direzione scientifica del DELOS WG on Digital Library, segue infatti il coordinamento della DELOS Network of Excellence on Digital Library² che ha per obiettivo un forum dove scambiare le nuove idee ed esperienze, addestrare giovani ricercatori, cooperare con le attività di standardizzazione, facilitare l'utilizzo delle tecnologie e, infine, dove identificare aree di cooperazione fra iniziative europee e nazionali e di cooperazione internazionale nel dominio della ricerca nelle Digital Library (DL). La responsabilità della Rete di Eccellenza DELOS è stata affidata all'ISTI anche per il triennio 2004-2007³.

Nel periodo 2001-2003 l'ISTI ha inoltre coordinato i seguenti progetti:

D-Lib Competence Center⁴ – ha scopo di fornire a specifiche comunità (bibliotecari, studenti, archivisti, editori) l'accesso a sistemi avanzati per DL, tecnologie, servizi, esperienza e conoscenza per facilitare il passaggio ad un modello di auto-pubblicazione/archiviazione; per dimostrare come le DL potrebbero facilitare la

¹ DELOS WG, EC 4th FP (1997-1999)

² DELOS NoE EC 5th FP (2000-2003)

³ DELOS NoE EC 6th (2004-2007)

⁴ Dlib Competence Center Project (IST-2001-32587). <http://dlibcenter.iei.pi.cnr.it>

collaborazione; per facilitare l'accesso aperto alle DL; per conservare l'accesso all'editoria elettronica; per promuovere la produzione e l'uso di archivi audio-video digitali.

Open Archive Forum⁵: un forum per la disseminazione dell'informazione sulle attività europee collegate alla Open Archive Initiative.

ECHO⁶: European Chronicles On-Line – un sistema di DL per la gestione di documenti storici audio-video basato su sistemi di indicizzazione automatica.

CYCLADES⁷: An Open Collaborative Virtual Archive Service Environment - un sistema per lo scambio di conoscenza, che supporta ricerche avanzate in grossi archivi digitali, eterogenei, multidisciplinari, la profilazione degli interessi e la raccomandazione. Cyclades usa il protocollo OAI per recuperare e indicizzare record da ogni archivio che supporta il protocollo OAI PHM (<http://www.openarchives.org>).

SCHOLNET⁸: A Digital Library Testbed to Support Networked Scholarly Communities – un sistema per supportare la collaborazione e la comunicazione fra comunità accademiche e di ricerca, per la gestione di tipi di dati non testuali, dell'annotazione ipermediale, la ricerca cross-language, la disseminazione personalizzata dell'informazione. L'architettura è "aperta" e permette l'estensibilità dei servizi per incontrare le esigenze delle differenti comunità accademiche.

Nel 6th Programma Quadro della Commissione Europea, l'ISTI partecipa ai seguenti progetti:

BRICKS Integrated Project⁹ – Building Resources for Integrated Cultural Knowledge Services ha lo scopo di stabilire i fondamenti delle DL a livello della Memoria Digitale Europea. BRICKS creerà una Biblioteca Digitale realizzata come sistema di servizi in rete di collezioni di documenti multimediali, che darà una varietà di informazioni ad una vasta gamma di utenti con modalità di accesso differenti

DILIGENT – An Experimental Digital Library Infrastructure on Grid Enabled Technology¹⁰ ha lo scopo di creare una infrastruttura di biblioteca digitale che permetta ai membri di organizzazioni virtuali di accedere a sorgenti informative condivise e di collaborare in modo sicuro, coordinato ed economico.

L'attività del D-Lib Competence Center continuerà all'interno del Rete di Eccellenza DELOS.

3. OpenDLib - Un sistema software per la gestione di Biblioteche Digitali

Nell'ambito del progetto SCHOLNET ISTI ha realizzato *OpenDLib*, un sistema software che offre servizi avanzati di biblioteca digitale.

OpenDLib consente di costruire una biblioteca digitale installando il software, configurandolo appropriatamente secondo le esigenze della comunità di utenti e caricando

⁵OAF Project (IST-2001-32015). <http://www.oaforum.org>

⁶ECHO Project (IST-1999-11994). <http://pc-erato2.iei.pi.cnr.it/echo>

⁷CYCLADES Project (IST-2000-25456). <http://www.ercim.org/cyclades>

⁸SCHOLNET Project (IST-1999-20664). <http://www.ercim.org/scholnet>

⁹BRICKS Project (507457)

¹⁰DILIGENT Project (IST 004260)

documenti e descrizioni catalografiche. I documenti gestiti possono essere sia immessi direttamente dall'autore o dall'amministratore, sia possono essere ottenuti importandoli automaticamente da fonti pre-esistenti. Questa doppia modalità consente di cambiare le procedure di pubblicazione senza perdita del pregresso.

E' possibile, inoltre, creare un comitato editoriale preposto alla pubblicazione dei documenti utilizzando il "meccanismo dei gruppi

OpenDLib è capace di gestire un'ampia gamma di documenti di diversa tipologia, struttura e "media". Una biblioteca OpenDLib, ad esempio, può mantenere e gestire in modo integrato descrizioni catalografiche e documenti digitali; può contenere riviste e atti di convegni composti da articoli; tesi, strutturate in capitoli e sezioni; filmati, strutturati in sequenze e shot. Inoltre, essa può anche gestire nuove tipologie di documenti che non hanno una controparte fisica, quali documenti composti di lucidi, filmati e registrazioni sonore relative ad un seminario o ad un corso di aggiornamento. Può mantenere edizioni, versioni e manifestazioni multiple di uno stesso documento e più descrizioni catalografiche in formati diversi possono essere associate ad uno stesso documento.

OpenDLib consente di organizzare i documenti in collezioni virtuali caratterizzate da politiche di accesso diverse. Queste collezioni possono essere create dinamicamente specificando un criterio di definizione. Ad esempio, si possono definire la collezione delle pubblicazioni scientifiche prodotte all'interno di istituti del CNR/Università, accessibile a tutti, le collezioni di rapporti di progetto, ciascuna delle quali è accessibile solo ai partecipanti al relativo progetto, la collezione della documentazione amministrativa, riservata al solo personale amministrativo, le tesi di dottorato coperte dalle normative sui diritti di autore ecc. Ad ogni collezione possono essere associate regole di accesso e servizi perché di fatto la collezione è una sorta di biblioteca virtuale.

OpenDLib, inoltre, è stato progettato per essere interoperabile con sistemi adottati in altri circuiti di pubblicazione digitale. Esso è in grado di raccogliere e mantenere descrizioni catalografiche di vari formati prodotte da altre organizzazioni e può rendere accessibili dall'esterno porzioni del proprio catalogo tramite un protocollo molto diffuso chiamato OAI-PMH.

Le descrizioni bibliografiche mantenute possono essere migrate in altri formati al fine della loro preservazione. Lo stesso vale per i formati dei documenti.

Dal punto di vista architeturale, OpenDLib è costituito da una federazione aperta di servizi che possono essere distribuiti e replicati su computer diversi. Questo tipo di architettura consente una grande flessibilità. E' possibile, infatti, costruire una biblioteca digitale i cui servizi sono gestiti da un unico centro che possiede le risorse di calcolo e umane necessarie, ma è anche possibile avere architetture distribuite sui computer di varie istituzioni. Ogni istituzione può decidere di mantenere solo un certo insieme di servizi. Ad esempio, essa può decidere di mantenere solo il servizio di "memorizzazione" per avere il controllo dei propri documenti, o di mantenere assieme a questo anche un servizio di indicizzazione. Un'istituzione può anche decidere di far gestire propri documenti ad una istituzione consorziata o ad una società di consulenza esterna. La scelta della topologia architeturale deve essere fatta al momento della creazione della biblioteca digitale, ma può anche cambiare dinamicamente, per soddisfare nuove esigenze, senza che il funzionamento della biblioteca digitale ne risenta. Ad esempio, un nuovo servizio di supporto alla ricerca può essere creato per diminuire il carico su quelli esistenti e migliorare i tempi di risposta del sistema; nuovi servizi di memorizzazione possono essere aggiunti per soddisfare la richiesta di altre istituzioni che vogliono rendere accessibili i loro documenti tramite la biblioteca digitale; nuovi servizi di ricerca possono essere aggiunti per consentire ricerche con linguaggi d'interrogazione e vocabolari controllati diversi.

Ulteriori informazioni su OpenDLib e una Biblioteca digitale dimostrativa sono accessibili all'indirizzo <http://www.OpenDLib.com>.

3.1 Conclusioni: una possibile applicazione di OpenDLib

Per le sue caratteristiche OpenDLib si presta molto bene a servire come sistema di supporto per una biblioteca digitale di letteratura scientifica e tecnica che mantiene e rende accessibile la documentazione scientifica prodotta da istituti diversi distribuiti sul territorio nazionale.

A questo scopo noi suggeriamo per una tale biblioteca un'architettura distribuita su vari computer mantenuti in poli diversi, quali, ad esempio, Istituti e/o Sezioni territoriali. Ogni polo che ha le necessarie risorse umane e di calcolo può occuparsi della gestione e preservazione dei documenti prodotti localmente ed, eventualmente, mantenere sui propri computer anche parte dei servizi condivisi. I poli che non hanno tali risorse possono appoggiarsi a poli "amici". La gestione dei servizi della biblioteca digitale è affidata ad un polo che registra le istituzioni afferenti alla biblioteca, mantiene aggiornata la configurazione dei servizi e garantisce la qualità del servizio nel suo complesso.

L'accesso alla documentazione avviene attraverso Internet.