

Metadati nel servizio DoIS per un accesso ibrido all'informazione professionale

di Antonella De Robbio e Vittorio Ponzani

Parole chiave: E-journal LIS - DoIS - OAI - Metadati - Risorse elettroniche - Biblioteche digitali - Biblioteconomia - Scienze dell'informazione - Servizio di accesso all'informazione biblioteconomica - Documenti di biblioteconomia in rete - Periodici elettronici di biblioteconomia

Abstract italiano

DoIS Documents on Information Science è un servizio di accesso all'informazione in linea per la ricerca e il recupero di documenti in formato elettronico a testo pieno nell'ambito della biblioteconomia e delle Scienze dell'informazione, incardinato entro il progetto ReLIS. ReLIS è un servizio di data provider che indicizza ad oggi quasi 10.000 documenti di ambito LIS, principalmente periodici elettronici e atti di convegni. È creato da un gruppo di bibliotecari e specialisti dell'informazione di varie nazionalità e ha sede sul server MIMAS a Manchester. La raccolta dei metadati avviene attraverso un service provider, o portale, noto con il nome di DoIS. Attualmente DoIS permette di recuperare (alla data del 17 marzo 2002) 4.234 documenti full text, accessibili in rete. La collezione di metadati ReLIS comprende circa 8.000 articoli di periodici elettronici e 1.792 working paper, disponibili attraverso l'infrastruttura ReLIS, che offre quindi una quantità di materiale rilevante, tale da incoraggiare la comunità internazionale di bibliotecari e professionisti dell'informazione a proseguire in questo lavoro. D'altra parte, un unico punto di accesso all'insieme delle risorse LIS sarà realmente utile a livello mondiale solo se sarà percepito come tale anche a livello locale e nazionale. ReLIS e DoIS si ispirano al fortunato progetto RePEc/WoPEC. RePEc (Research Papers in Economics) è il servizio di accesso internazionale che consente la disseminazione della letteratura scientifica di ambito economico. DoIS utilizza lo standard di metadati ReDIF, basato sul modello IAFA, e il protocollo Guildford.

Abstract inglese

DoIS Service: metadata for an hybrid access to professional information. ReLIS is a volunteer-run data provider indexing about 10.000 Library and Information Science (LIS) documents, mainly electronic journal articles and working papers. The metadata is harvested by a service provider, or portal, named DoIS. DoIS is viewed like an international database of articles and conference proceedings published in electronic format in LIS area. It was created by a group of international librarians and information professionals and is based at MIMAS (Manchester site). At the moment DoIS holds 4.234 full text documents (update on 17th March 2002), that can be accessed by all on the Internet. The existing ReLIS metadata collection for about 8.000 LIS e-journal articles and 1.792 working papers, available through ReLIS infrastructure, will provide an excellent critical mass of material with which to encourage take up by the community. It is a very useful tool to efficiency search and retrieve information on LIS fields. Equally, a single access point to a wide range of LIS resources will only benefit the LIS international community if it is perceived to be useful and usable by local and national communities too. ReLIS and DoIS are modelled on the successful JISC-funded RePEc/WoPEC project. RePEc (Research Papers in Economics) is a international collaborative effort to enhance the dissemination of research in economics. DoIS uses ReDIF metadata, based on IAFA templates, and Guildford protocol.

Nel mondo contemporaneo l'informazione ha acquistato un'importanza strategica straordinaria e l'avvento del Web ne garantisce oggi la disponibilità immediata, offrendo enormi opportunità per lo sviluppo della conoscenza. Le basi di dati utilizzabili gratuitamente in linea, gli OPAC, i periodici elettronici e tutte le altre tipologie di archivi digitali (testuali e multimediali), si integrano nelle le biblioteche digitali attraverso i loro

metadati, in un'ottica di interoperabilità. Inoltre, l'efficace utilizzo di metadati che descrivono gli oggetti digitali delle collezioni distribuite lungo la Rete permette un efficace recupero delle informazioni attraverso gli strumenti dell'Information Retrieval. Non va tuttavia sottovalutata né la complessità delle questioni relative alla gestione e allo sviluppo di tali risorse informative, né tanto meno i meccanismi della loro individuazione e localizzazione all'interno di modelli distribuiti e della creazione di servizi per l'accesso all'informazione di fasce di utenza differenziate.

L'espansione dell'ambiente digitale e la sempre maggiore disponibilità di risorse elettroniche remote pongono infatti nuovi problemi ai professionisti dell'informazione: la descrizione e l'indicizzazione dei documenti digitali con l'uso di metadati, non solo sintattici ma anche e soprattutto semantici, la conservazione e l'accesso a tali documenti, dopo opportuna organizzazione in collezioni distribuite, portano alla creazione di nuovi modelli economici di diffusione dell'informazione i quali richiedono competenze nuove e strumenti di analisi rigorosi.

La gestione dei diritti di proprietà intellettuale è per esempio uno dei temi centrali che caratterizza gli ambienti della biblioteca digitale, gestione che può trovare utile spazio di manovra entro le strutture offerte dai metadati.

L'aggiornamento professionale di bibliotecari e di documentalisti e l'attività di studio e di ricerca nell'ambito della biblioteconomia e delle scienze dell'informazione diventano perciò una priorità inderogabile, oltre che uno dei modelli di formazione continua durante tutto l'arco della vita.

A tal fine è necessario creare o sviluppare gli strumenti bibliografici in grado di segnalare e recuperare efficacemente le produzioni intellettuali nelle differenti forme in cui esse possono manifestarsi all'interno delle diverse comunità scientifiche.

Questo presupposto, valido per tutte le discipline, e a differenti livelli, italiano o internazionale, a maggior ragione deve essere applicato anche alle nostre discipline LIS (Library and Information Science).

Un esempio tradizionale di strumento repertoriale assai utile in ambito LIS è costituito dallo spoglio accurato e sistematico della letteratura italiana di ambito biblioteconomico, preziosa attività che produce la nota *Letteratura professionale italiana* (LPI) a cura di Giulia Visintin.

LPI è pubblicata dal 1975 all'interno del "Bollettino d'informazioni" dell'Associazione italiana biblioteche e oggi del "Bollettino AIB"; dal 1995 le registrazioni bibliografiche presenti nella LPI sono state riversate in CD-ROM con il titolo *BIB: Bibliografia italiana delle biblioteche, del libro e dell'informazione*, a cura della stessa Visintin e di Alberto Petrucciani.

Sul versante digitale si colloca **DoIS (Documents in Information Science)**¹, strumento fondamentalmente diverso da un repertorio cartaceo, ma diverso anche, come concezione e funzionalità, da una banca dati intesa in senso tradizionale.

DoIS, presente e indicizzato in *Google's Web Directory for Library and Information Science*, nasce dallo sforzo di un gruppo volontario di bibliotecari e professionisti dell'informazione provenienti da diversi paesi, che collaborano al fine di creare una risorsa bibliografica che possa offrire contenuti scientifici selezionati e con una copertura internazionale. Queste risorse sono gratuitamente accessibili sia per gli addetti ai lavori che per gli utenti non specialisti e sono ricercabili attraverso un motore di ricerca non commerciale, che agisce sui metadati. DoIS consente il trattamento di ogni tipo di documento, pubblicato o pre-print, ai fini dell'individuazione, localizzazione e recupero del documento originale che può essere sia su supporto cartaceo che su qualsiasi altro supporto.

DoIS è uno strumento in linea in grado di recuperare gli articoli di periodici, le comunicazioni presentate a convegni e a conferenze, i rapporti di ricerca specializzati in Library and Information Science, presenti sul web. Nella maggior parte dei casi dalla descrizione bibliografica è possibile accedere liberamente, tramite un link, al *full text* del documento.

Potenzialmente DoIS potrebbe anche gestire informazioni il cui accesso sia a pagamento; nei suoi *template* è previsto un apposito gruppo di campi integrati (*cluster*) ove inserire tali informazioni, che poi compaiono nel metadato visibile all'utenza.

DoIS contiene prevalentemente metadati con descrizioni di documenti presenti in Rete a testo pieno e liberamente accessibili. Sebbene in DoIS attualmente coesistano molte descrizioni relative a documenti su supporto cartaceo, si tratta di uno strumento molto più orientato a documenti in formato digitale; la percentuale di documenti digitali accessibili da DoIS sta in effetti aumentando progressivamente. Anche se si tratta di un servizio recente, DoIS ha già avuto uno sviluppo significativo. Come segnalato in *D-Lib Magazine*, settembre del 2000², alla data del 9 settembre 2000 DoIS conteneva appena 1000 descrizioni provenienti da sette riviste:

¹ <<http://dois.mimas.ac.uk/>>

² <<http://www.dlib.org/dlib/september00/09clips.html>>

Ariadne, Boletín de RedIris, Bulletin des Bibliothèques de France, CLIR Reports, Cuadernos de Documentación Multimedia, D-Lib Magazine, Journal of Electronic Publishing.

Attualmente, dopo soli diciotto mesi, i titoli indicizzati sono 206: 152 periodici e 54 altre collezioni (convegni, conferenze, ...). Di questi 206 titoli, 113 conducono al recupero di 4138 documenti a testo pieno su un totale di 9643 metadati disponibili, di cui 7868 relativi ai periodici e 1775 per altro materiale (dati al 24 gennaio 2001).

La presenza in DoIS di una transitoria prevalenza di descrizioni relative a documenti cartacei è dovuta alle seguenti ragioni storiche e funzionali:

- DoIS è uno strumento di ricerca ibrido, rispondente a un concetto di servizio di accesso all'informazione ibrida, cartacea e digitale;
- attualmente permane ancora un buon numero di riviste LIS disponibili solo in formato cartaceo, anche se numerose sono quelle che stanno sorgendo in formato nativo digitale o che stanno affiancando una versione elettronica a quella tradizionale;
- DoIS è prodotto dal lavoro di un gruppo misto di bibliotecari e professionisti del settore che, pur collocandosi in un contesto internazionale, si basa su una logica di selezione che rispecchia scelte nazionali. Per il momento le collaborazioni provengono in prevalenza da Spagna, Stati Uniti e Italia, ma altre collaborazioni si stanno aprendo;
- ogni Paese o progetto che collabora a DoIS con il proprio materiale di produzione nazionale è libero di scegliere quale "politica" adottare in relazione alla costruzione "virtuale" delle sue raccolte, in relazione alla produzione stessa, ma anche alle "forze" che può dedicare al progetto;
- un nucleo primario di DoIS è stato originato dal riversamento di un database di ambito spagnolo che conteneva descrizioni di documenti tradizionali; per questa ragione si trova un certo numero di descrizioni per materiale cartaceo. Tale nucleo primario sta però per essere neutralizzato dall'immissione di nuovi metadati per collezioni digitali;
- l'architettura di DoIS è flessibile e consente di utilizzare differenti "template" o "modelli" per la descrizione di differenti oggetti (digitali o cartacei).

DoIS non è propriamente né un catalogo né una bibliografia tradizionale, ma uno strumento ibrido. Infatti, se da una parte si può considerare un vero catalogo, in cui dalle descrizioni bibliografiche dei suoi metadati si accede attraverso un link ai documenti in linea, in altri casi va considerato una bibliografia, perché vi sono descritti dei documenti non presenti in Internet (quindi non "posseduti" dalla biblioteca digitale), ma che esistono solo su supporto cartaceo: l'utente quindi ottiene la notizia ma non può accedere direttamente al documento.

La catalogazione dei documenti in DoIS è effettuata secondo lo standard di metadati ReDIF (Resource Description Information Format)³ e il protocollo Guildford⁴.

ReDIF è un formato *relazionale* di metadati, nel quale ogni record rappresenta univocamente le caratteristiche di un determinato oggetto; ogni record ha un identificatore, che viene chiamato "*handle*" (maniglia) e può contenere identificatori di altri record: le relazioni fra gli elementi descritti vengono così rappresentati in forma compatta, senza duplicare le stesse informazioni. Va sottolineato che ReDIF è un formato compatibile con l'Open Archive Initiative (OAI) Protocol, e questa è una delle principali ragioni per cui è stato scelto per il DoIS. In ReDIF peraltro sono stati convertiti i formati di altri sei servers OAI all'interno del prototipo UPS (Universal Preprint Server)⁵: questo significa che esiste un mapping tra ReDIF e gli altri formati OAI compatibili tra cui Dublin Core.

Il formato dei file ReDIF, non va confuso con RDF (Resource Description Framework), che non è uno standard di metadati, ma è un modello di riferimento per l'interoperabilità tra metadati differenti. ReDIF tiene conto di RDF.

RDF è una piattaforma per il trattamento di metadati, in quanto costituisce una ipotesi di soluzione individuata dal W3C al problema della rappresentazione dei metadati soprattutto a livello semantico.

³ <ftp://netec.mcc.ac.uk/pub/NetEc/RePEc/all/root/docu/redif_1.html>

⁴ Per la documentazione sul modello RePEc e sullo standard ReDIF
<<http://ideas.uqam.ca/ideas/data/rpcrd/doc.html>>

⁵ The UPS Prototype project: exploring the obstacles in creating a cross e-print archive end-user service ", with Herbert Van de Sompel, Michael L. Nelson and others, Old Dominion University Computer Science TR 2000-01, February

<http://cs-tr.cs.cornell.edu:80/Dienst/UI/1.0/Display/ncstrl.odu_cs/TR_2000_01>

Il vantaggio di utilizzare ReDIF quale standard di metadati è che la descrizione non è relativa solo ai documenti, ma indicizza anche gli autori che quei documenti producono, i canali attraverso cui tali documenti vengono resi pubblici, i processi di controllo su questi canali, le persone e le istituzioni che gestiscono e organizzano queste attività. Per il momento DoIS non applica appieno tutte le funzionalità offerte da ReDIF come sistema relazionale perché, a differenza di altri ambiti disciplinari come l'economia, la matematica, la fisica, dove già da tempo esistono numerosi archivi elettronici contenenti documenti a testo pieno (e-print server), per il contesto LIS non si hanno, al momento, *repositories* dedicati al deposito di documenti LIS.

Per questa ragione DoIS utilizza solo una parte del modello ReDIF, pur cogliendone appieno la filosofia portante. A livello di organizzazione dell'informazione, DoIS utilizza i manuali tecnici per la creazione dei metadati che fanno riferimento ai template ispirati ai modelli IAFA⁶.

I metadati, che possono risiedere su server diversi rispetto ai documenti ai quali si riferiscono e che comunque sono "esterni" ai documenti stessi, sono ricercabili dagli utenti attraverso un motore di ricerca.

Dopo la rielaborazione dei metadati che in questo modo sono presenti in due luoghi distinti, presso il server locale e presso il server centrale, che ne copia la struttura organizzativa, avviene la rielaborazione in formato HTML. Tale elaborazione consente la creazione dell'interfaccia per l'utente, che offre la possibilità di ricercare, attraverso una semplice stringa, nel contenuto di tutti i metadati presenti in DoIS, compreso il campo abstract; è comunque possibile restringere la ricerca su un campo specifico (in particolare sui campi titolo, autore, abstract). Per ogni metadato offerto il servizio assegna un identificatore permanente attraverso il sistema noto come "handle". Ogni metadato presenta quindi un aggancio alla risorsa, utile alla sua localizzazione, che consente una certa garanzia di stabilità ai fini del recupero effettivo del documento nel tempo.

Gli scambi informativi avvengono attraverso l'utilizzo di un protocollo "comune" e così si instaura un modello comunicativo tra le parti che si scambiano/comunicano informazioni. Nello scambio e trasporto dei metadati risulta di fondamentale importanza l'adozione di un protocollo adeguato allo scopo del servizio che si vuole implementare.

Il protocollo con il quale DoIS comunica è il Guildford Protocol⁷, il quale specifica l'architettura e la gestione di un database distribuito su un insieme di siti locali. Ogni sito risiede ed è gestito da un singolo sistema informatico, in modo del tutto indipendente e autonomo, in un'area di memoria accessibile via anonymous FTP (File Transfer Protocol) e/o http (webserver).

Il protocollo Guildford prevede una serie di azioni atte a regolare due processi comunicativi distinti ma correlati. Il primo processo consente di raccogliere e rendere disponibili all'utente finale uno o più archivi, che possono contenere documenti a testo completo (gli e-print server) oppure file di metadati (come nel caso di DoIS), o software di gestione, inseriti e mantenuti localmente. Il secondo processo rende disponibili copie totali o parziali di archivi remoti, prevedendone il regolare aggiornamento.

È da sottolineare che tutte le parti componenti di DoIS e gli strumenti che esso offre alla comunità dei bibliotecari sono costruiti sul modello OpenSource. Vale a dire che tutti i software a disposizione hanno codici sorgenti aperti e possono essere liberamente prelevati, distribuiti, modificati e rielaborati, nell'ottica della filosofia che sottende le licenze GNU, a disposizione di una piena cooperazione sia entro la comunità di bibliotecari sia tra comunità diverse.

Sono stati sviluppati diversi pacchetti di software in linguaggio Perl per implementare questi processi e anche DoIS ne utilizza alcuni. Uno dei fronti di prossimo sviluppo ed evoluzione sarà appunto in relazione alla creazione e implementazione di software confezionati specificamente per DoIS, piuttosto che riadattati sul modello della rete RePEc⁸, un progetto internazionale per il recupero e l'accesso dei preprint di economia.

L'infrastruttura di tipo distribuito di DoIS riprende il modello utilizzato da RePEc, ma se ne differenzia per quanto riguarda la configurazione degli archivi. RePEc, le cui origini risalgono al 1993 con il progetto WoPEc (Working Papers in Economics), dal 1999 è costituito da una rete di oltre 200 archivi⁹ che mettono a

⁶ IAFA Templates in use as Internet metadata
<<http://www.w3.org/Conferences/WWW4/Papers/52/>>

⁷ Thomas Krichel "Guildford Protocol"
<<http://ideas.uqam.ca/ideas/data/Papers/rpcrdocGuildP.html>>

⁸ RePEc fa parte del gruppo NetEc finanziato dal Joint Information Systems Committee (JISC) Higher Education Funding Councils, come parte del del programma eLib - Electronic Libraries Programme, per gli anni 1996-1999

<<http://netec.mcc.ac.uk/>>

⁹ Per la lista degli archivi vedere:

disposizione più di 160.000 articoli e rapporti tecnici di ambito economico, dei quali oltre 60.000 disponibili online. Si tratta di un sistema decentralizzato, basato cioè su una rete di archivi che raccolgono e conservano metadati relativi a documenti di ambito economico (e talvolta anche i documenti stessi). Ogni istituzione che partecipa al progetto mette a disposizione i propri dati ma, contemporaneamente, diventa un *mirror* che garantisce l'accesso anche ai dati di tutti gli altri archivi. Esistono quindi molti accessi a RePEc, che poi si differenziano tra loro per i diversi servizi a valore aggiunto offerti ai propri utenti. Alcuni, per esempio, oltre ai dati presenti negli archivi RePEc, forniscono le registrazioni bibliografiche di articoli e rapporti tecnici disponibili solo in versione cartacea, altri forniscono indici navigabili sui dati relativi alle istituzioni che partecipano al progetto, altri infine permettono di svolgere ricerche non solo sugli indici relativi ai documenti indicizzati in RePEc, ma anche su altri siti web che possono essere rilevanti in ambito economico.

DoIS, pur nascendo dall'esperienza di RePEc e riprendendone la filosofia portante, se ne differenzia sostanzialmente dal punto di vista dell'organizzazione degli archivi. Infatti mentre RePEc è un sistema totalmente decentralizzato, DoIS ha un server centrale (a Manchester) dove vengono raccolti tutti i metadati prodotti dalle varie istituzioni.

All'interno della struttura a modello distribuito chiunque può descrivere risorse (proprie o di altri enti) attraverso metadati descrittivi standard, collocarli su un server FTP anonymous o web e comunicare al server centrale dove reperire le informazioni (cioè i metadati). Il server centrale conserva la lista con gli indirizzi di tutti i server distribuiti e attraverso questa lista il motore è in grado di raccogliere ogni notte i file (contenenti uno o più metadati), copiandone i contenuti, per poi trasportare le copie dei file al server centrale. I metadati vengono organizzati in base al campo "*handle*", che svolge la funzione di identificatore URN (Uniform Resource Name), e viene poi creato un indice collettivo, mantenendo al contempo anche una "copia della struttura locale". Dopo aver elaborato le informazioni, DoIS è in grado, la mattina dopo, di presentarle in modo aggiornato all'utente finale.

Il trasporto dei metadati dai server distribuiti al server centrale avviene grazie ad una struttura gerarchica che organizza le informazioni in metadati di archivio e di serie (metadati amministrativi gestionali MAG) e in metadati descrittivi, relativi alle risorse da descrivere. In questo modo è possibile lavorare in modo cooperativo, ma autonomo, nella creazione "partecipata" di metadati, coinvolgendo direttamente gli enti produttori di informazione (biblioteche, università, associazioni), distribuendo il lavoro all'interno di un sistema che consente la massima visibilità dei documenti, la loro disseminazione, senza costi aggiuntivi se non quelli del personale coinvolto nella descrizione delle risorse, cioè nella creazione dei metadati.

La disponibilità sul web dei documenti a testo pieno, liberamente accessibili da un unico punto d'aggregazione, quale può essere il servizio DoIS, è un'ottima opportunità per il mondo delle biblioteche e dei centri di documentazione italiani, che in questo modo hanno la possibilità di recuperare e raccogliere più facilmente i lavori italiani e stranieri di ambito LIS disponibili in rete. Le istituzioni pubbliche e private, gli editori di periodici elettronici, gli enti che organizzano congressi o conferenze e tutti i produttori di letteratura professionale biblioteconomica hanno così l'opportunità di autodocumentare le proprie attività e di diffondere i risultati delle ricerche svolte, partecipando al progetto DoIS e creando i metadati relativi alle proprie pubblicazioni.

La partecipazione di un gruppo di lavoro italiano al progetto DoIS rappresenta, per i bibliotecari e per i documentalisti italiani, una duplice opportunità: da una parte infatti permette di offrire ai propri utenti un servizio di documentazione e informazione rapido ed efficace, in grado di arrivare subito al *full text* del documento; dall'altra parte, DoIS rende visibile e diffonde la produzione intellettuale dei professionisti dell'informazione italiani a livello internazionale, favorendo così la crescita della professione nel nostro Paese.