

**Progetto “Una biblioteca digitale CNR: dal prototipo al servizio”: rapporto attività -  
2008/2009, n 2, UNSC**

Rapporto attività Settembre 2008 – Agosto 2009

di Maria De Vizia Guerriero

CNR.ISTI/2009-PR-009

## Sommario

1. Introduzione
2. Le collezioni
3. Modalità di acquisizione
4. Conclusioni

## **Abstract**

In questo Project Report viene illustrato il lavoro svolto sulla biblioteca digitale PUMA durante il periodo che va dal Settembre 2008 ad Agosto 2009. Vengono descritte le collezioni, il loro incremento e lo stato della situazione attuale.

## 1. Introduzione

Il presente lavoro ha come obiettivo la descrizione dell'attività, da me svolta, nell'anno 2008-2009 all'interno del progetto "Una biblioteca digitale CNR: dal prototipo al servizio".

L'ambito di interesse del progetto riguarda il settore "Documentazione scientifica" che si applica alla costituzione di una biblioteca digitale che si occupi della gestione e della fruibilità della documentazione scientifica sia acquistata che prodotta dal CNR, come:

- periodici cartacei e/o elettronici
- monografie
- pubblicazioni del personale del CNR

coinvolgendo anche altre strutture del CNR sparse sul territorio nazionale.

La costituzione di una biblioteca digitale prevede tre fasi di lavoro.

La prima prevede diverse passi: analisi del materiale, contestualizzazione, analisi delle necessità che la documentazione comporta e requisiti per gestire i documenti.

La seconda: reperimento i documenti di cui si ha bisogno dalla rete.

La terza: acquisizione documenti da formato cartaceo a formato digitale, convertirli prima di essere inseriti nel sistema di gestione.

I documenti digitalizzati vengono inseriti nel sistema PUMA- Publication Management - che permette la raccolta, la catalogazione, la conservazione e la fruizione di informazioni bibliografiche prodotte dal personale del CNR, aumentandone la consistenza e divenendo l'archivio istituzionale delle pubblicazioni degli Istituti del CNR.

### 1.1 Introduzione a PUMA

LPUMA è costituita da diverse parti:

- **la collezione:** formata da metadati e documenti digitali,
- **i servizi di accesso:** riservati all'amministratore del sistema, ai bibliotecari e agli utenti;
- **l'utente finale** che interagisce con altre persone, diventa un agente attivo e può compiere diverse operazioni.

Tramite il sistema PUMA l'utenza, costituita non solo da autori, ma anche da bibliotecari, amministratori e dalla comunità scientifica può intervenire direttamente nella gestione dei dati,

quindi creazione, gestione, interrogazione e conservazione di documenti di varia natura come, ad esempio, tutto ciò che rientra nella **letteratura pubblicata** dagli editori (articoli ed editoriali pubblicati su riviste, articoli e abstract pubblicati su atti di congresso e workshop, libri, brevetti, schede del software, materiale cartografico) messa a disposizione in modalità OPEN ACCESS e ciò che rientra nella **letteratura grigia e ad accesso libero** (abstract e poster a congressi, rapporti tecnici, note interne, tesi, rapporti di progetti ed altri ancora).

Le tipologie di documenti che l'archivio di PUMA contiene si possono suddividere in:

- **Publicati** (National e International Journal, Contribution to International Books/Monograph, Contribution to National Books/Monograph, Books/Monograph International, Books/Monograph National, International Conference Abstract/Poster, National Conference Abstract/Poster, International Journal Abstract/Poster, National Journal Abstract/Poster);
- **Non Publicati** (presentazioni/comunicazioni, abstract/poster a congressi, workshop, recensioni);
- **Brevetti.**

PUMA consente, inoltre, la conservazione e la diffusione della letteratura pubblicata National journal, International journal, International e National conference, Proceedings, Contributions, più altre tipologie relegabili nella cosiddetta "letteratura grigia", come, rapporti tecnici, note interne, project report, tesi e altro ancora.

Il codice identificativo di ogni lavoro è quello attribuito automaticamente dal sistema e formato dal nome della collezione, dalla sigla corrispondente alla tipologia, dall'anno e dal numero progressivo all'interno dell'anno stesso. Per ogni documento è stata mantenuta, comunque, l'informazione sul codice identificativo originale apposto sulle copie cartacee, archiviato nel campo Numero Documento/Codice Originale nel quale è stata mantenuta anche l'informazione sull'appartenenza alle collezioni precedenti.

### 3. Modalità di acquisizione

La fase di acquisizione del materiale inserito in PUMA è avvenuta in due modi: laddove i documenti non risultavano già presenti in formato elettronico all'interno del sito web dell'editore, o nelle pagine web degli autori, è stato necessario trasformarli da cartacei a digitali con gli strumenti hardware messi a disposizione dalla biblioteca.

La digitalizzazione dei documenti è avvenuta mediante l'utilizzo di uno scanner ad alta risoluzione in dotazione alla Biblioteca dell'Area della Ricerca del CNR.; dove è possibile, per ogni documento digitalizzato salvarne una copia con estensione *.pdf*, inserirlo, in una cartella predefinita nominata Aficio.

Dopo aver compilato la scheda bibliografica del documento, è possibile allegare il file alla fine del processo. Il file è nella sua cartella, lo si apre e si estraggono alcune informazioni, ad esempio l'abstract che viene copiato e salvato in un file *.txt* e poi inserito nella scheda catalogafica.

#### 3.1 Catalogazione e Archiviazione

- La fase di **inserimento** dei documenti digitalizzati in PUMA ha previsto la compilazione della scheda bibliografica in cui sono definiti i metadati. Per descrivere i metadati PUMA ha adottato Dublin Core.

L'ultimo elemento che in PUMA chiude la scheda d'inserimento di un documento è il campo **Diritti di accesso**. Quest' ultimo elemento segnala all'utente il livello di accessibilità alla visualizzazione, al download e alla stampa del documento. Ci sono tre diversi livelli di accesso:

- **Libero**: significa che il documento è accessibile a tutti (riguarda il materiale *open access*);
- **Riservato**: significa che il documento è accessibile solo dagli indirizzi IP abilitati all'interno della struttura del CNR;
- **Protetto**: significa che l'utente deve contattare l'autore o il bibliotecario per ottenere la visualizzazione del documento.

#### 3.2 Amministrazione

La fase che chiude il processo di inserimento del documento è l'**amministrazione**.

In questa parte si ha la possibilità di controllare che il record bibliografico contenga tutte le informazioni corrette, richiedere il codice del documento ed eventualmente modificarlo, di

visualizzare il tipo di accesso al documento (Libero, Riservato, Protetto) potendolo modificare a seconda delle richieste dell'autore.

Il codice che viene assegnato è univoco, è proposto automaticamente dal sistema ed è formato da una serie di elementi: nome della collezione, sigla della tipologia, anno e numero progressivo interno all'anno stesso.

Il totale dei documenti che in questo modo sono stati da me inseriti ed amministrati è di circa 300.

Nella tabella sottostante sono segnalati gli Istituti/Collezioni per cui ho lavorato e le relative annate:

<b>ISTITUTO/COLLEZIONE</b>	<b>ANNO</b>
CNUCE	1989-1995-1997-1998
IEI	1996-1997
IBF - PI	1995

Per le collezioni su cui ho lavorato è stato valido il suddetto metodo, per la collezione CSCE, invece, il discorso è un po' diverso.

Le difficoltà maggiori riscontrate in questo processo sono dovute, soprattutto, allo stato dei documenti "storici", in particolare a quello delle note interne, dato che le pubblicazioni erano quasi tutte costituite da estratti ben conservati. Spesso le note interne si sono presentate stampate con il testo inclinato, con inchiostri che con il tempo si sono sbiaditi e con numerosi fogli lacerati. Il lavoro di acquisizione, dunque, è stato lungo e complesso ed ha richiesto, talvolta, diversi passaggi allo scanner e successive rielaborazioni con un software specifico che consente di editare i file ottenuti con l'acquisizione, fino a che il risultato non è stato ritenuto soddisfacente. Il criterio che è stato seguito ha favorito, soprattutto, la leggibilità e la facile consultazione delle note interne ma ha prodotto, spesso, la perdita del formato originale delle stesse. Sono stati acquisiti anche i disegni ed altre particolarità per le quali è eventualmente consigliata la visione delle copie cartacee, comunque conservate presso la biblioteca dell'ISTI.

Per alcuni lavori, anonimi o di attribuzione incerta, data l'obbligatorietà del campo autori prevista dal sistema, è stato inserito l'autore collettivo CSCE. Le descrizioni sono state corredate dei rispettivi Sommari o Abstract quando presenti nel lavoro o quando dal lavoro stesso era possibile,

per il catalogatore, estrapolare informazioni utili alla compilazione del campo. Eventuali notizie riportate sui documenti e non associabili a nessuno dei campi specifici previsti dal sistema, ma ritenute importanti ai fini della descrizione, sono state inserite nel campo Note. Sia ai documenti pubblicati che ai documenti non pubblicati, inseriti nella collezione, è stato assegnato il diritto di accesso Libero, di solito usato soltanto per i documenti non pubblicati, perché ogni vincolo contrattuale, derivante da eventuali accordi con gli editori è da considerarsi ormai estinto per il superamento dei limiti temporali previsti dalla normativa italiana sul D.A .

#### 4. Conclusioni

Dopo una veloce panoramica sui criteri metodologici che si sono applicati alle collezioni, è importante quantificare l'incremento documentario che si è avuto durante questo anno.

La collezione del CNUCE è stata tutta digitalizzata e inserita, per questo istituto la tipologia documentaria a disposizione, per lo più, B4- note interne e Technical Report.

I dati di seguito possono dare un'idea della quantità documentaria:

Anno	Numero documenti
1979-2002	604
CNUCE- Manuali	
1968-1978	220

Il totale generale del fondo del CNUCE consta di 824 documenti che vanno dal 1968 al 2002.

La collezione dell'Istituto I.E.I. è molto più variegata e corposa rispetto a quella precedente.

Oltre al reperimento materiale in archivio, è stato possibile acquisire i documenti digitali anche dalla rete.

Il materiale acquisito riguarda:

Anno	Numero documenti
1968-1997	422

I documenti sopra-elencati riguardano la prima raccolta documentaria effettuata sui TR e B4 dell'I.E.I.

La terza collezione dell'ISTI riguarda i documenti del CSCE.

I documenti erano divisi tra Note interne e pubblicazioni, l'arco temporale coperto va dal 1954 al 1968, per ulteriori chiarimenti <http://csce.isti.cnr.it/>.

I documenti acquisiti e catalogati sono stati in tutto 385.

La collezione dell'Istituto di Linguistica Computazionale "A. Zampolli" ha riguardato l'acquisizione in pdf, il trattamento dei dati acquisiti e l'inserimento su PUMA di 300 documenti dal 1979 al 2000. I documenti trattati sono tutti TR, l'obiettivo era quello di inserire tutta la Letteratura Grigia dell'Istituto per fare un punto della situazione. Gli argomenti di cui i TR trattavano erano i più diversi: linguistica, didattica delle lingue, filosofia del linguaggio, storia delle lingue, grammatica trasformazionale - generativa, fonetica, stilistica, dizionari comparati.

Sono inoltre compresi alcuni riferimenti a strumenti per l'elaborazione di dati linguistici e opere relative alle applicazioni del calcolatore alle ricerche umanistiche e letterarie e all'uso di modelli matematici nella Linguistica.

L'incremento registrato quest'anno in Puma è stato notevole. Sono state arricchite collezioni già esistenti e create delle nuove.

Inoltre è stata fatta anche un'attività di *testing* sulle uscite bibliografiche e di continui studi sul sistema volti a renderlo sempre più ricco.