

Una Biblioteca Digitale CNR per la scienza e la Tecnica: dal prototipo al servizio (2008 – 2009)

Rapporto N°1 sull'attività svolta per il Servizio Civile Nazionale presso l'ISTI- CNR

di Marina Rovetti

1 Settembre 2008 – 31 Agosto 2009

INDICE

1. INTRODUZIONE.....	pag. 3
1.1 Il progetto e i suoi obiettivi.....	pag. 3
1.2 Ambito di attività e strumenti utili.....	pag. 4
2. LA BIBLIOTECA DIGITALE.....	pag. 5
3. PUMA	
3.1 Definizione e obiettivi di PUMA.....	pag. 5
3.2 Analisi del materiale	pag. 6
3.3 Modalità di acquisizione.....	pag. 6
3.4 Modalità di inserimento.....	pag. 6
3.5 Modalità di amministrazione.....	pag. 7
3.6 Monitoraggio e test delle procedure del sistema	
PUMA	pag. 8
4. MONITORAGGIO ED ALLINEAMENTO DEI	
DATI IN PUMA E LIBERO.....	pag. 9
5. COLLEZIONE CSCE E REALIZZAZIONE	
DEL SITO WEB.....	pag. 10

1. INTRODUZIONE

Il seguente rapporto ha lo scopo di descrivere l'attività da me svolta in qualità di Volontario del Servizio Civile Nazionale durante l'anno 2008/2009 presso la biblioteca dell'Istituto di Scienze e Tecnologie dell' Informazione “Alessandro Faedo”.

1.1 Il progetto e i suoi obiettivi

“Una Biblioteca Digitale CNR per la scienza e la Tecnica: dal prototipo al servizio” è il progetto approvato dall'Ufficio del Servizio Civile Nazionale, ideato e realizzato dalla biblioteca dell'Istituto di Scienze e Tecnologie dell'Informazione “Alessandro Faedo” e dal Servizio di Documentazione Scientifica dell'Area di Ricerca del CNR di Genova, che ha previsto la costituzione di una biblioteca virtuale, di un sistema in grado di gestire il patrimonio e diffondere informazioni su parte della documentazione scientifica del CNR, ovvero periodici cartacei ed elettronici (riviste scientifiche) e pubblicazioni dei ricercatori del CNR.

L'obiettivo è dunque quello di estendere la fruibilità di tali informazioni ad un numero maggiore di persone tra cui il personale di ricerca delle strutture del CNR, la comunità scientifica nazionale ed internazionale, i gestori della Biblioteca e di un Servizio di Documentazione Scientifica offrendo loro di:

- accedere direttamente ai formati elettronici;
- ricercare libri per un prestito tra le strutture che lo permettono;
- ricercare le pubblicazioni del personale del CNR;
- utilizzare i sistemi di archiviazione, aggiornamento e conservazione del materiale bibliografico;
- garantire un servizio di fornitura di articoli, con un sistema automatizzato che registri tutte le richieste evase e metta a disposizione le statistiche del servizio offerto;
- partecipare al catalogo collettivo dei periodici;
- consentire il deposito delle pubblicazioni del personale della propria struttura.

1.2 Ambito di attività e strumenti utilizzati

Il mio contributo alla realizzazione di tale sistema informativo ha riguardato :

- l'inserimento di articoli scientifici nella Biblioteca Digitale PUMA;
- il supporto informatico e lo sviluppo delle applicazioni della biblioteca digitale del CNR.
- la gestione e aggiornamento dati del catalogo LIBERO¹;
- la verifica della consistenza delle risorse bibliografiche fruibili sui siti degli editori;
- l'ordinamento del materiale documentario cartaceo;
- le procedure di prestito del materiale documentario (prestito interbibliotecario);
- l'orientamento degli utenti all'uso della biblioteca;
- l'assistenza all'uso dei cataloghi cartacei e soprattutto on-line;
- l'assistenza all'uso delle postazione informatiche;
- la ricerche bibliografiche tematiche per aiutare gli utenti meno esperti;

2. LA BIBLIOTECA DIGITALE

Vivendo in un'era tecnologica, dominata dallo sviluppo del Web, la Società dell'informazione (la conoscenza) ha subito una rivoluzione e in essa ha trovato la sua collocazione la Biblioteca Digitale, “uno spazio virtuale avulso da confini temporali e fisici che riunisce collezioni (costituite da banche dati testuali, dati numerici, immagini, documenti sonori e video), architettura di reti, servizi e persone a supporto dell'intero ciclo di vita dell'informazione”².

Si tratta di una biblioteca telematica, che non vuole sminuire l'importanza della biblioteca tradizionale o dimostrare la superiorità del testo elettronico su quello a stampa, ma che vuole essere un supporto nella diffusione della cultura, distribuire le risorse informative in formato digitale sfruttando le funzioni di acquisizione, archiviazione, preservazione, recupero e accesso che vengono realizzate attraverso l'uso di tecnologie digitali e permettere di:

¹ Infine **LIBERO** è il sistema utilizzato per la catalogazione, gestione dei periodici, gestione opac e web-opac all'interno della biblioteca dell'ISTI. Il catalogo della biblioteca può essere infatti consultato attraverso il modulo WebOpac: l'interfaccia di questo strumento accede al database rendendo possibili anche operazioni di prestito e prenotazioni. Il web Opac permette anche operazioni di ricerca sui vari indici (quindi per autore, titolo, note, descrittore e altri) e prevede inoltre l'uso di operatori logici (and, or, not), la visualizzazione dei record è in formato Marc e ISBD.

² http://digitalia.sbn.it/upload/documenti/digit00_tammaro.pdf

- condividere l'informazione e renderla accessibile da ogni luogo (cosicché non sia l'utente a muoversi verso l'informazione, ma l'informazione a muoversi verso l'utente);
- consentire agli utenti l'accesso a un'informazione aggiornata costantemente;
- migliorare l'accesso all'informazione grazie a funzionalità nuove o diverse;
- consentire un uso migliore dell'informazione attraverso l'eliminazione delle barriere di tempo e spazio e rendere disponibile una migliore infrastruttura;
- consentire una migliore collaborazione tra ricercatori;
- ridurre il *digital divide* scavalcando o diminuendo le distanze dei popoli in Internet.

Componenti principali di una biblioteca digitale sono quindi: **la collezione** che è formata da metadati ovvero documenti, video e audio; **i servizi di accesso** che danno all'utente la possibilità di reperire velocemente e facilmente le informazioni ed estendere la ricerca a documenti collegati tra loro e per ultimo, ma non ultimo, **l'utente** che da solo, senza barriere spaziali e temporali, interagisce con altre persone, diventa un agente attivo che può compiere più operazioni.

3. PUMA

3.1 Definizione e obiettivi di PUMA

Tra i servizi che la Biblioteca dell'Area della Ricerca di Pisa mette a disposizione e su cui si è basata la maggior parte della mia attività è il sistema di PUblication Management **PUMA**.

Si tratta di uno strumento per la gestione di Archivi Istituzionali e/o tematici, un “servizio dedicato raccolta, catalogazione, conservazione e divulgazione di informazioni bibliografiche e documenti digitali (letteratura pubblicata e/o grigia)”³ che offre l'accesso a più collezioni degli Istituti del CNR partecipanti alla biblioteca digitale.

Tramite il sistema PUMA l'utenza, costituita non solo da autori, ma anche da bibliotecari, amministratori e dalla comunità scientifica ha a disposizione quelle funzionalità descritte precedentemente, quindi creazione, gestione, interrogazione e conservazione di documenti di varia natura come, ad esempio, tutto ciò che rientra nella **letteratura pubblicata** dagli editori commerciali (articoli ed editoriali pubblicati su riviste, articoli e abstract pubblicati su atti di congresso e workshop, libri, brevetti, schede del software, materiale cartografico) e ciò che rientra nella **letteratura grigia**, ad accesso libero, (abstract e poster a congressi, rapporti tecnici, note interne, tesi, rapporti di progetti ed altri ancora).

³ <http://puma.isti.cnr.it/>

3.2 Analisi del materiale

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

- **Publicati** (Riviste Nazionale e Internazionali, Contributi a Monografie/Libri Nazionali e Internazionali, Libri/Monografie Nazionali e Internazionali, Abstract/Poster a Conferenze Internazionali e Nazionali);
- **Non Publicati** (presentazioni/comunicazioni, abstract/poster a congressi, workshop, recensioni);
- **Brevetti**.

3.3 Modalità di acquisizione

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

La digitalizzazione è avvenuta quindi grazie alla macchina RICOH – AFICIO 3224c dotata di scanner ad alta risoluzione; dove possibile, per ogni documento digitalizzato e salvato in formato *.pdf*, è stato estratto l'abstract tramite il sistema di riconoscimento ottico di caratteri (OCR – *Optical Character Recognition*) del software *Adobe Professional*; quindi ogni abstract è stato copiato e salvato in un nuovo documento *.txt*, rinominato con lo stesso codice del documento pdf e collocato in una cartella virtuale a fianco ad esso, in modo che entrambi fossero facilmente individuati e selezionati nella fase successiva, quella cioè di inserimento.

3.4 Modalità di inserimento

La fase di **inserimento** dei documenti acquisiti in digitale in PUMA ha previsto, oltre l'upload dei documenti e degli abstract digitalizzati, la compilazione della loro scheda bibliografica in cui sono definiti i metadati.

Per definire i metadati PUMA ha adottato il formato Dublin Core esteso per le esigenze del CNR, un sistema che garantisce l'identificazione ed il recupero degli oggetti digitali grazie ai suoi elementi descrittivi e si caratterizza per:

- **l'interoperabilità semantica**, che stabilisce una comune rete di dati concordati nel loro

significato e valore;

- **la flessibilità**, che permette di integrare e sviluppare la struttura dei dati con significati semantici diversi ed appropriati al contesto di applicazione.

L'ultimo elemento che in PUMA chiude la scheda d'inserimento di un documento è il campo **Diritti di accesso**: esso segnala all'utente il grado di accessibilità alla visualizzazione, al download e alla stampa di un documento e può essere:

- **Libero**: significa che il documento è accessibile a tutti (per quanto riguarda gli Unpublished Document);
- **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 informazioni sul documento.

3.5 Modalità di amministrazione

L'**amministrazione** del documento, che chiude il processo di archiviazione, permette all'amministratore, in questo caso io stessa, di assegnare al record bibliografico e al suo documento allegato un codice univoco, proposto automaticamente dal sistema dal sistema e formato da quattro elementi: nome della collezione, sigla della tipologia, anno e numero progressivo interno all'anno stesso.

All'interno della sezione di amministrazione è anche possibile **modificare** un record immesso o realizzarne una **nuova versione**.

La differenza tra la scelta di modificare un record bibliografico e di realizzarne una nuova versione sta nel fatto che il sistema non permette di effettuare modifiche sul campo relativo al **tipo** e all'**anno** di pubblicazione (quindi solo se l'amministratore decide di fare una nuova versione di un documento potrà ipoteticamente convertirlo da tipo A0 a tipo B0 o sostituire la data di pubblicazione da 2007 a 2008; in tutti gli altri casi, egli potrà selezionare il pulsante di modifica e agire sugli altri campi).

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:

ISTITUTO/COLLEZIONE	ANNO
CNUCE	1989-1990-1991-1997
IEI	1996-1997
IBF - PI	1995

3.6 Monitoraggio e test delle procedure del sistema PUMA

L'esperienza all'interno della biblioteca non ha solo coinvolto la sfera umanistico-bibliotecaria, ma anche quella che potremmo definire tecnico-informatica. Infatti dopo una fase di analisi della infrastruttura PUMA, del suo DB, dello stile di programmazione e delle procedure che vi sono alla base mi è stata data la possibilità di dare supporto tecnico con l'obiettivo di:

- modificare ed ottimizzare dal lato server le principali procedure attive sulla macchina Orione (piattaforma di sviluppo e di testing di PUMA) che riguardano l'input, la visualizzazione, il salvataggio, la modifica dei dati e la loro amministrazione;
- testare le funzionalità lato client delle procedure sopraelencate attraverso la compilazione delle schede bibliografiche di ogni tipo di documento e la segnalazione di eventuali problemi sorti in seguito alle costanti modifiche apportate;
- testare le funzionalità attive sulla macchina Orione che riguardano l'interfaccia di amministrazione;
- aggiungere nel record bibliografico un campo "ACRONIMO" visualizzabile nelle tipologie di documenti A2, B2, A6, A3, B6;
- aggiornare l'Help online di Puma in italiano ed inglese in seguito all'introduzione di nuovi elementi del record bibliografico;
- modificare le procedure per allineare i DataBase di GEO e PUMA;
- modificare alcuni elementi grafici delle tabelle in PUMA agendo sui fogli di stile;
- introdurre nella pagina di ricerca dei documenti e in quella di consultazione delle liste i simboli grafici che rappresentano il tipo di accesso (libero, protetto, riservato) ai documenti;
- inserire nel DataBase i nuovi Istituti/Collezioni che nel tempo hanno aderito al progetto PUMA (ISE, ILIESI, CSCE ed altri);
- creare una serie di procedure di check dei dati inseriti su PUMA;
- aggiungere la procedura che permette durante la nuova versione di un documento la modifica dell'anno di pubblicazione e del tipo.

4. MONITORAGGIO ED ALLINEAMENTO DEI DATI IN PUMA E LIBERO

Una nota a parte merita il lavoro svolto per la Collezione **IMATI** di Pavia e per la creazione di un DB di prova chiamato ZTEST.

1. Per la Collezione **IMATI** di Pavia:

- sono state recuperate dal sistema ETRDL ed allegate alle schede bibliografiche già esistenti le pubblicazioni che risultavano mancanti nel precedente;
- sono stati estratti gli abstract dai documenti open-access reperiti dalla rete e sono stati inseriti nelle relative schede bibliografiche su PUMA;
- sono stati controllati i pdf già presenti, allegati ai record bibliografici;
- sono state modificate le keyword e corrette le affiliazioni per uniformità al sistema;

Per quanto riguarda invece la creazione della collezione **ZTEST** sono state testate le uscite bibliografiche delle varie tipologie di documenti. In questo spazio virtuale sono stati creati e compilati 300 documenti di varia tipologia per verificarne la corretta uscita bibliografica e sono state segnalate eventuali anomalie.

Tra le attività di supporto alla biblioteca ho partecipato alla revisione dei documenti presenti nella Collezione ISTI dal 2001 al 2007 tramite controlli incrociati nel catalogo Libero e nella biblioteca digitale PUMA; la revisione ha previsto la segnalazione dei documenti mancanti in uno dei due sistemi, quindi il loro inserimento, la riassegnazione dei codici corretti, l'inserimento dei pdf e degli abstract relativi ai record del CNUCE 2005, la correzione dei laboratori e delle affiliazioni.

All'interno del catalogo LIBERO ho aggiornato parte del database denominato "SER" contenente i record delle riviste elettroniche: essendo compito mio quello di revisionare solo i dati dell'editore Springer, ho limitato la selezione attraverso il filtro *Publisher*; in seguito sono intervenuta verificando la disponibilità/consistenza online delle riviste in esame (sia nel sito del Portale periodici in consorzio, sia nel sito dell'editore stesso) e apportando modifiche necessarie all'aggiornamento dei campi relativi al volume, all'anno di fruibilità e al link ipertestuale che conduce al full-text del documento.

5. COLLEZIONE CSCE E REALIZZAZIONE DEL SITO WEB

In occasione del Convegno *Pisa, culla dell'Informatica: mezzo secolo dopo la CEP e l'Olivetti di Barbaricina*, tenutosi l'11-12 giugno 2009 presso il CNR di Pisa, la biblioteca ha deciso di creare la Collezione CSCE relativa alla documentazione storica/scientifica prodotta dal personale e dai collaboratori del Centro Studi Calcolatrici Elettroniche, dal 1954 al 1968 e conservata presso l'Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo". Il materiale, composto da pubblicazioni, note interne e altre tipologie è suddiviso in due serie:

- prima serie (1957-1965)
- seconda serie (1962-1968).

Per tale attività ho partecipato alla digitalizzazione del materiale cartaceo e alla realizzazione del sito web accessibile all'indirizzo <http://csce.isti.cnr.it>.

Il primo passo compiuto in collaborazione con i colleghi del Servizio Civile è stato la digitalizzazione dell'intera collezione, costituita da circa 400 documenti così divisi:

PRIMA SERIE	SECONDA SERIE
Pubblicazioni: 43	Pubblicazioni: 128
Note Interne: 147	Note Interne: 71
Totale documenti I serie: 190	Totale documenti II serie: 199

Il lavoro è stato molto complesso e lungo: perché, trattandosi di documenti “storici”, le pagine si sono presentate lacerate, gli inchiostri consumati dal tempo e dalla polvere, le righe inclinate dalla battitura a macchina. Come conseguenza di ciò l'acquisizione è avvenuta più volte utilizzando tutti gli strumenti messi a disposizione dal software Acrobat Professional, tra cui lo strumento di rotazione (per raddrizzare il testo inclinato) e di ridimensionamento delle pagine (per tagliare i margini anneriti e consumati) per garantire un livello per lo meno sufficiente di leggibilità sullo schermo a discapito però della rappresentazione del formato originale.

Una volta digitalizzati i documenti sono stati inseriti in PUMA con le rispettive schede bibliografiche e di seguito amministrati assegnando loro il codice identificativo proposto dal

sistema (ad esempio: cnr.csce/1956-B4-001).

A questo punto, vista l'importanza storica della Collezione si è pensato di creare una pagina web specifica per dare accesso ai dati tramite liste e ricerche più dettagliate rispetto a quelle proposte da PUMA. Per questo motivo è nato il sito del CSCE, realizzato con il CMS (Content Management System) Joomla, strumento ideale per costruire pagine web veloci e dinamiche, portali di dimensioni medio-grandi facilmente gestibili garantendo di:

- mantenere un sito sempre aggiornato a livello di contenuti e di servizi;
- avere piena libertà grafica rispecchiando l'esatta immagine aziendale;
- intuire navigare con estrema facilità nell'interfaccia e di intuirne i comandi d'uso che sono basati sui più comuni applicativi (ad esempio MS Word);
- erogare servizi online come newsletter, forum, e-Commerce, news, e altri;
- avere un valido supporto nella configurazione iniziale ed assistenza all'utilizzo.

In seguito alla scelta del template più adatto si è proceduto alla definizione di una navigazione che potesse essere funzionale e intuitiva e permettesse di:

- trovare le informazioni della collezione originale e le modalità di acquisizione per la digitalizzazione;
- accedere all'indice dei disegni delle pubblicazioni e delle note interne della prima serie e quindi di effettuarne il download;
- offrire strumenti di ricerca per titolo, soggetto, autore, anno;
- accedere alla lista bibliografica delle note interne e delle pubblicazioni della prima e seconda serie del CSCE;
- accedere a link utili e servizi correlati.

Il codice XHTML, CSS e PHP prodotto nelle pagine web e validato secondo gli standard del W3C è stato sottoposto a modifiche per posizionare gli elementi della pagina (tabelle, margini, testo, immagini) e per definire stili come *border-color*, *text-decoration*, *font-weight*, *background-color*. Le immagini inserite nel sito sono state create ed elaborate con lo strumento Adobe Professional utilizzato nella grafica professionale.