

Enrico Seta

Biblioteca digitale: modelli, linee di tendenza e progetti per la Biblioteca della Camera dei deputati

1 - Introduzione: definizione del termine biblioteca digitale e ambito della ricerca; 2 - Tecnologie e costi; 3 - Il project planning e il project management; 4 - I progetti in corso; 5 - Conversione retrospettiva del catalogo; 6 - Conversione delle registrazioni analitiche; 7 - Digitalizzazione degli Atti parlamentari; 8 - Futuri sviluppi; 9 - Elaborare un'interfaccia; 10 - Individuare le raccolte e disegnare un'architettura dell'accesso.

1 - Introduzione: definizione del termine biblioteca digitale e ambito della ricerca

Preliminarmente occorre chiarire che l'espressione *digital library*, che qui si traduce con biblioteca digitale, ha significati diversi a seconda del contesto in cui viene adoperata. Fra ingegneri informatici e produttori di software il termine sta ad indicare semplicemente un tipo di prodotto: database che gestiscono documenti multimediali. Nell'ambito, invece, della *library and information science* per biblioteca digitale si intende — più ampiamente — l'evoluzione delle biblioteche fisiche nell'era dell'informazione elettronica, evoluzione che si realizza fondamentalmente attraverso l'editoria elettronica o attraverso la conversione di documenti dal formato analogico al formato digitale. Qui si parla di biblioteca digitale in questo secondo senso.

Chiarito questo primo punto, sarà utile ricordare che l'espressione indica più un processo che una cosa: la biblioteca digitale non è una biblioteca in formato digitale, cioè una raccolta — statica — di documenti elettronici, aggregazione di opere in formato digitale. Il termine è venuto infatti assumendo una valenza più ricca perché rinvia alle tematiche dell'evoluzione dei processi informativi. Il vero problema della biblioteca digitale non è infatti di carattere tecnologico, ma è costituito dal fatto che la rappresentazione digitale, oltre a viaggiare sulle reti, permette delle potenzialità molto maggiori nell'aggiungere funzionalità e utilità ad un *corpus* di informazioni. Tutto questo apre — come dimostra l'universale

dibattito sul Web — una sorta di reazione a catena su distinti piani: tecnologico, linguistico-conoscitivo, organizzativo, amministrativo, legale.

Un modo più esatto di definire la biblioteca digitale sarebbe quello di considerare in un unico insieme sia le raccolte in formato digitale, sia gli strumenti di rappresentazione dei dati e di gestione delle raccolte che rendono possibile il recupero delle informazioni, o meglio la conversione dei dati in informazione e conoscenza. Tuttavia anche questa definizione — che già accenna a una maggiore complessità dei problemi coinvolti — è stata ritenuta insufficiente: in un recente *workshop* promosso dalla National Science Foundation del concetto di *digital library* si è data la seguente definizione:

«is not merely equivalent to a digitised collection with information management tools. It is rather an environment to bring together collections, services, and people in support of the full cycle of creation, dissemination, use, and preservation of data, information, and knowledge» ⁽¹⁾.

Vi è molto interesse, quindi, anche attorno ad una teoria della biblioteca digitale, probabilmente perché vi è molta incertezza su ciò che essa significherà per il mondo della cultura, dell'educazione, per il commercio, il lavoro e la società in generale, secondo alcuni per la stessa democrazia ⁽²⁾. Considerando la complessità a cui una tale definizione rinvia, risulterà chiaro che la presente breve ricerca non può rappresentare che una sommaria introduzione a queste problematiche, e non certo un loro svolgimento. A queste note più generali si sono aggiunte alcune considerazioni circa l'impatto che l'avvento della biblioteca digitale — intesa appunto come nuovo ambiente in cui si svolgono i processi della comunicazione e dell'informazione — potrà avere, e sta già avendo, sulla concreta realtà della Biblioteca della Camera dei deputati.

A questo proposito è forse utile qui ricordare un altro fenomeno di carattere generale: le nuove potenzialità della biblioteca digitale stanno modificando profondamente le istituzioni tradizionali, le cosiddette biblioteche fisiche. Di fronte alla crisi di un'immagine ereditata dal passato, si fronteggiano due previsioni opposte.

Secondo la prima di queste previsioni l'informazione digitale sta crescendo ad un ritmo talmente vertiginoso — si calcola una decuplicazione ogni anno — da rendere obsoleto rapidamente, anzi sotto i nostri stessi occhi, praticamente tutto ciò che rappresenta storicamente il quadro di riferimento dell'istituzione biblioteca, a partire naturalmente dall'oggetto libro. Già oggi, nel campo delle biblioteche tecnico-scienti-

fiche negli USA circa il 50% in media del budget delle acquisizioni è destinato ai formati elettronici; fra l'altro in gran parte si tratta di abbonamenti a servizi in linea, quindi a rigore non si può neanche parlare di acquisizione di opere, non essendovi supporto materiale. Questo trend, per certi aspetti rapidissimo, sta rivoluzionando pratiche e modelli gestionali, ponendo anche ardui problemi legali e amministrativi in relazione alla gestione dei bilanci. Modifica inoltre i paradigmi attraverso i quali avviene l'interazione fra persone e informazioni e in particolar modo i processi della comunicazione scientifica, processi a loro volta strettamente intrecciati con la produzione, gli scambi, l'economia. Da questi nuovi scenari l'intermediazione effettuata dalle biblioteche sarebbe destinata rapidamente a sparire.

Secondo un'altra prospettiva assisteremo invece — in modo meno traumatico — ad una lunga fase di assestamento attraverso un processo di progressiva integrazione. Biblioteca digitale e biblioteca tradizionale sono destinate a convivere: da un lato la comunicazione stampata, dall'altra la comunicazione elettronica; da un lato testi lineari, dall'altro ipertesti; da un lato un accesso mediato, dall'altro un accesso diretto; da un lato collezioni fisiche, dall'altro collezioni digitali; da un lato edifici, dall'altro interfacce ⁽³⁾.

Quale di queste due ipotesi si realizzerà non è facile dire.

Certamente si può affermare che alcune tendenze in atto — di velocissima trasformazione — non possono essere ignorate o sottovalutate allorché si decidono investimenti o strategie. La sottovalutazione di queste trasformazioni può condurre, infatti, a errori nelle decisioni di investimento che, per essere in parte mascherati non sono certo meno gravi. Nelle pagine che seguono si riportano, a questo fine, alcuni dati quantitativi di orientamento. D'altra parte, l'ipotesi dell'integrazione, «*And, not Or*», offre dei motivi di interesse, e non solo perché essa assegna ancora un futuro ad un mondo al quale moltissimi si sentono legati, ma anche perché da essa risulta valorizzato maggiormente — rispetto all'ipotesi più apocalittica — il momento dell'analisi, dell'esame critico di vantaggi e svantaggi di ogni soluzione tecnica, amministrativa, culturale, e quindi la necessità di elaborare programmi, priorità e scadenze. Si vuole qui affermare che la risposta alle sfide della biblioteca digitale non è un atteggiamento di generica e globale adesione alle novità, ma piuttosto un impegno di accurata analisi di processi, di ponderazione di alternative e infine di progettazione di soluzioni nuove e originali.

Ricordare questo dibattito sul futuro delle biblioteche nell'era della biblioteca digitale ⁽⁴⁾ non serve, dunque, ad indicare frettolosamente una

previsione o una preferenza ma a sottolineare lo sfondo su cui collocare una riflessione (e le future scelte) sui concreti progetti di digitalizzazione di una biblioteca fra le principali del nostro paese.

Se le prospettive più generali sono oggetto di dibattito, ciò che invece appare certo è che nel prossimo futuro gli attuali bibliotecari e gli altri professionisti dell'informazione dovranno collaborare, interagire, sperimentare nuovi modelli di integrazione con i detentori di un numero crescente di tecnologie. Essi dovranno anche dialogare maggiormente con i fruitori delle istituzioni, perché anche i servizi forniti saranno profondamente diversi dal passato.

Un altro elemento facilmente prevedibile è che anche la Biblioteca della Camera dei deputati dovrà confrontarsi in modo crescente con nuove sollecitazioni nella direzione della digitalizzazione; e tutto ciò andrà presto disposto in un quadro progettuale più ampio che definisca obiettivi, priorità e percorsi. Tuttavia in questa ricerca non si è tentato neanche di abbozzare un programma generale di digitalizzazione per la Biblioteca della Camera dei deputati: per la sua complessità esso richiederà — nei prossimi anni — un impegno collettivo e istituzionale e probabilmente non avrà l'aspetto organico del documento scritto onnicomprensivo (anche se prestigiose istituzioni hanno invece scelto proprio questo metodo più sistematico).

In questa ricerca si offre solo del materiale di carattere informativo generale che potrà servire alla definizione di quel programma, e si fa il punto su alcuni dei progetti già in corso di realizzazione: conversione del catalogo a schede, digitalizzazione degli Atti parlamentari, conversione degli spogli di riviste.

Il tema delle biblioteche digitali — proprio per i motivi a cui si accennava sopra — comprende alcune problematiche di particolare complessità, quali: nuovi metodi di indicizzazione e di accesso all'informazione; editoria elettronica; gestione del copyright; conservazione delle raccolte digitali; tariffe. In questa ricerca — che è ancorata a obiettivi concreti o già in fase di realizzazione — si sono fatti solamente brevi riferimenti a queste tematiche più generali, se non semplici rinvii: una loro effettiva esposizione non potrebbe essere contenuta in poche pagine. L'attenzione è stata, invece, maggiormente rivolta alla pianificazione e gestione di progetti di digitalizzazione, cioè di conversione retrospettiva al formato digitale. Sempre al fine di mantenere un collegamento con il concreto contesto di riferimento, si è parlato di digitalizzazione pensando prevalentemente a documenti di carattere testuale, più che ad archivi di immagini, o di documenti sonori o di filmati. Il capitolo della

multimedialità costituisce infatti un argomento a sé, di sicuro interesse per una biblioteca come quella parlamentare, ma in una scala di priorità esso ha una posizione subordinata rispetto alla digitalizzazione di quelle raccolte documentarie testuali che svolgono un ruolo chiave nello svolgimento dei prioritari compiti istituzionali. Anche in questo senso, dunque, il tema della biblioteca digitale, nelle pagine che seguono, è appena introdotto; l'approfondimento di tutti i vari aspetti che lo compongono sarà nel prossimo futuro, anche per la Biblioteca della Camera dei deputati, oggetto di un'attività collettiva e articolata.

Infine, si vuole concludere questa premessa con una notazione. È stato osservato e indagato ⁽⁵⁾ un fenomeno negativo, ma molto diffuso: l'affacciarsi dell'innovazione tecnologica in molti casi produce una vera e propria crisi gestionale, manageriale e addirittura professionale nelle biblioteche. La spiegazione di come un'opportunità di crescita possa trasformarsi in un regresso la si trova nel punto nodale degli investimenti. L'evoluzione tecnologica ha, infatti, un suo ritmo ideale, ma la struttura amministrativa responsabile dell'erogazione delle risorse finanziarie spesso non consente alla biblioteca di procedere all'innovazione alla velocità che sarebbe richiesta. Ciò ha degli effetti negativi — intuibili, ma che comunque si trovano ben descritti in letteratura — sia sul piano dell'organizzazione generale della struttura, sia sul piano di quella preziosa risorsa professionale che è la capacità di entrare in sintonia con il cambiamento. Nell'articolo citato si indica — a titolo di esempio, ma emblematicamente — l'ipotesi di una biblioteca a cui si richieda, oggi, di svolgere i propri compiti istituzionali, cioè fornire documentazione pertinente e aggiornata, disponendo di PC della serie 486, di infrastrutture di rete deboli e di pochi punti di consultazione Internet. È curioso osservare come l'esempio astratto scelto dall'autore dell'articolo corrisponda esattamente alla situazione in cui — a tutt'oggi (ottobre 1998) — si trova la Biblioteca della Camera dei deputati.

I motivi per cui questo fenomeno si verifica così spesso e così universalmente sono molteplici, ma comunque ben radicati se la situazione classica è quella per cui gli investimenti tecnologici per la biblioteca sono posti in coda alla lista di priorità: l'autore appena citato dice, testualmente, «at the back of the list». È possibile che vi sia una macroscopica sottovalutazione del valore (economico, anche se non sempre commerciale) dell'informazione. È possibile che vi sia a volte anche una semplice carenza di comunicazione fra i due distinti soggetti (biblioteca — erogatore delle risorse finanziarie). I dati e gli elementi presentati nei paragrafi successivi mirano anche a favorire questa comunicazione.

2 - *Tecnologie e costi*

Oggi quasi ogni testo che viene scritto proviene da un word processor e non presenta, quindi, problemi di conversione. I problemi per una biblioteca in questo caso sono di natura diversa: archiviare e rendere accessibili questi documenti. Ma, rispetto ai documenti che nascono in formato digitale la vera sfida è: attrezzarsi per riuscire a tenere il passo con la crescita del mercato dell'editoria elettronica, sia di CD-ROM — e presto anche dei DVD — sia on line. Anche in un mercato periferico come quello italiano, a partire dal 1996, si è avuta, ad esempio, una vera esplosione dei servizi editoriali su *www*: secondo i dati forniti dall'Associazione Nazionale dell'Editoria Elettronica (ANEE), fra le sole testate giornalistiche ormai circa il 50% è coinvolto direttamente in attività di editoria on line ⁽⁶⁾. Questi servizi sempre più spesso non hanno un equivalente in formato tradizionale: in queste condizioni l'obiettivo della completezza delle raccolte diventa per le biblioteche sempre più arduo e sfuggente. Se si guarda al problema in una scala globale, la tendenza si afferma in modo molto più marcato. Ad esempio, si calcola che già oggi — per quanto riguarda testi in lingua inglese — ciò che è stato scritto nel periodo più recente (ultimi 18 mesi) si trovi sul Web in una percentuale maggiore rispetto alla più grande biblioteca fisica del mondo, la Library of Congress. Il controllo bibliografico universale, secondo la nota definizione dell'IFLA, si pone, nell'ambiente della biblioteca digitale, in termini del tutto nuovi: fondamentalmente come problema di interconnessione fra reti. Ancora un dato quantitativo ⁽⁷⁾ può aiutare a descrivere meglio come si propone questa competizione fra digitale e analogico. L'informazione presente sul Web nel 1997, limitatamente a documenti in formato testo, era di 2 TB ⁽⁸⁾. È vero che digitando in caratteri ASCII l'intera LC si realizzerebbe una banca dati di 20 TB, ma Internet — è stato calcolato — sta crescendo ad un ritmo di 10 volte l'anno. Questo ritmo è destinato a rallentare nei prossimi anni, ma intanto possiamo concludere che già oggi (fine 1998) risiede sulla rete Internet — e non più presso la LC — il più grande archivio di documenti testuali.

Molti di questi documenti, si diceva, nascono già in forma digitale, ma in gran parte finisce su Internet il risultato di un'attività vastissima di digitalizzazione: iniziative di varia natura e proporzione — come vedremo in alcuni esempi riportati — gestite per lo più proprio da biblioteche, autonomamente o in consorzio con altri soggetti. Insieme alla sfida dell'editoria elettronica, le biblioteche devono affrontare anche quella

della digitalizzazione, cioè della conversione ⁽⁹⁾, in formato digitale di documenti di alto contenuto informativo che, nella misura in cui rimangono nei formati tradizionali, non possono essere immessi nel circuito delle reti.

Digitalizzazione significa letteralmente trasformazione di un segnale continuamente variabile (segnale analogico) in un altro costituito da unità discrete rappresentate da numeri binari (segnale digitale). Teoricamente tutti i tipi di documenti possono dunque essere digitalizzati, sia pure con tecniche differenti. In questa sede, come si è accennato, si fa riferimento solo a documenti testuali. Per questi, a parte la digitazione su una tastiera collegata ad un computer per la produzione di testi (prevalentemente) in caratteri ASCII, oggi — grazie ai progressi nella tecnologia degli scanner — una valida alternativa all'oneroso procedimento della digitazione del testo è offerta dalla scansione: digitalizzazione di immagini, cioè riproduzione della pagina in forma di file grafico di tipo bitmap. Il procedimento consiste nella ricezione dei segnali elettrici da parte di una fotocamera (CCD *charge-coupled device*) e nella trasformazione del segnale elettrico in bit (quindi in numeri) da parte di un convertitore. Anche se il documento contiene testo, il software dello scanner lo interpreta, tuttavia, come figura, e questo comporta l'impossibilità di modificare il testo e di effettuare tutte le funzioni di ricerca sul testo. Il procedimento della scansione può, allora, accompagnarsi, per i documenti testuali, ad un altro procedimento, quello del riconoscimento ottico dei caratteri da parte di un software OCR.

Questa tecnologia, al contrario di quanto accade per gli scanner, non ha prodotto ancora risultati veramente soddisfacenti. Pur migliorando ogni anno, si tratta ancora di progressi lenti ⁽¹⁰⁾. I fattori che entrano in gioco sono non solo la qualità e regolarità dei caratteri, ma anche la qualità e le condizioni di conservazione della carta. Ad esempio, anche i migliori programmi di OCR presentano una percentuale di errore molto alta per i testi di giornali quotidiani, e ciò dipende dalla qualità della carta e non dalla qualità della stampa. Nella realizzazione di progetti di conversione di grandi proporzioni, quali quelli che in genere interessano le biblioteche, gli interventi (manuali) di correzione degli errori dell'OCR, quasi sempre necessari, hanno costi troppo alti.

La scelta tra digitazione e scansione dipende quindi da molte variabili che — come vedremo — incidono notevolmente sui costi. In primo luogo va detto che il risultato finale è molto diverso in termini di memoria necessaria. Mentre una pagina stampata occupa 2 KB in ASCII, la stessa pagina occupa 30 KB in formato-immagine. Quindi, per un volu-

me di medie dimensioni (200-300 pagine), digitando i caratteri su una tastiera si produce un file ASCII delle dimensioni di circa 500 KB (un terzo di un floppy disk ⁽¹¹⁾) mentre sottoponendo tutte le pagine a scansione si produce un file-immagine delle dimensioni di circa 8-10 MB (Megabyte). Il rapporto, in termini di memoria occupata, è di almeno 1/15. Ma la disponibilità di grandi memorie di archivio non rappresenta, oggi, un problema, neanche in termini di costi, come si vedrà. Lo diventa, invece, quando si digitalizzano file di immagini in movimento: un film in formato digitale occupa, ad esempio, 10 GB (cioè 10.000 MB).

Venendo invece alla conversione vera e propria, il costo medio ⁽¹²⁾ per la digitazione, il controllo e l'eventuale correzione degli errori di una pagina può raggiungere oggi le 4000-5000 lire. Il costo medio per la scansione è di circa 500 lire. Ma costi molto più bassi possono ottenersi attraverso la scansione da microfilm o utilizzando scanner veloci ad alimentazione automatica, fino alle 100 lire/pagina.

In termini di costo il rapporto oscilla dunque tra 1/10 e 1/50. È un rapporto troppo alto che fa propendere in molti casi per la soluzione della scansione, rinunciando alla possibilità di effettuare ricerche sul testo.

I costi indicati possono, però variare molto. In primo luogo grandi variazioni possono essere determinate dalla tipologia del materiale che si passa allo scanner: solo testo, testo e immagini, immagini in bianco e nero o a colori ecc. Ad esempio, la British Library ha calcolato un costo medio di £ (sterline) 3,56 per articolo di rivista (costi relativi agli anni 1993-94), ma è interessante osservare che l'Image Demonstrator Project, a cui il dato si riferisce, pur riguardando la digitalizzazione di materiali abbastanza omogenei, articoli di riviste, ha registrato grandi variazioni dei costi ⁽¹³⁾. Ma occorre rinunciare — in questa sede — a una effettiva e compiuta analisi di questi aspetti poiché le variabili sono troppe ed esse si chiariscono solo in sede di esame dettagliato dello specifico progetto. Limitandosi quindi ad enunciare i fattori che possono incidere maggiormente sui costi della digitalizzazione, e in particolare della conversione in formato immagine, si deve considerare, accanto al costo di scansione e di elaborazione dell'immagine, il costo di estrazione dei dati, poiché è evidente che un numero indefinito di file-immagine non sarebbe, in quanto tale, ricercabile in alcun modo.

Il problema è di importanza strategica nella progettazione: anche qui non esiste una soluzione ottimale, ma quasi tante soluzioni ottimali quante sono le raccolte da digitalizzare. Inoltre si tratta di un campo in cui è in atto una forte ricerca e si realizzano esperienze pilota che rendono il

panorama quanto mai mutevole. Qui ci si limiterà a fornire alcuni riferimenti orientativi. In primo luogo a proposito dell'uso o meno di un OCR.

Quello che costa non è il procedimento di riconoscimento dei caratteri, che la maggior parte dei programmi OCR esegue contemporaneamente alla scansione, né il software di OCR vero e proprio, pur esistendone di prezzo e qualità molto differenti, quanto le correzioni necessarie. Dover riprendere ogni singola pagina e individuare gli errori presenti ha un costo che in genere scoraggia l'adozione degli OCR per la conversione di archivi di grandi dimensioni. Dunque, dall'applicazione del programma di OCR c'è da aspettarsi un testo contenente molti errori, soprattutto se parliamo di documenti non recentissimi, e quindi costi di correzione molto alti.

Tuttavia il discorso non può essere chiuso qui: gli errori di trascrizione dei caratteri, infatti, non vanno necessariamente visti come inficianti la validità di un progetto che preveda l'applicazione dell'OCR. È proprio su questo aspetto che negli ultimi anni si sono realizzati i risultati e le ricerche più interessanti. Normalmente, infatti, l'applicazione OCR ha lo scopo di rendere il testo ricercabile ⁽¹⁴⁾ attraverso un programma di information retrieval e il dato interessante è che oggi si vanno diffondendo sistemi sempre più sofisticati che usano algoritmi resistenti agli errori dell'OCR. Quello che sta mutando completamente è la tecnologia del recupero dell'informazione. Mentre negli IR tradizionali gli strumenti erano, infatti, prevalentemente basati su elementi linguistici (dalle liste di stop words al troncamento alle derivazione dalle radici) i sistemi moderni — potendo disporre di memorie molto superiori e di una ricerca che è molto progredita, anche perché ha interessato l'industria privata — usano tecniche sofisticate, derivate dalle ricerche sulle reti neurali, che aboliscono le barriere alle quali siamo abituati. In questa ricerca, come si è detto, il tema del recupero dell'informazione non viene affrontato poiché richiederebbe un suo spazio. Qui può accennarsi al metodo di fondo: molti dei moderni motori di ricerca non confrontano stringhe di parole, ma «patterns digitali», cioè non devono strutturare pesanti indici per poter effettuare la ricerca, ma — secondo una metodologia più vicina al modo in cui il cervello umano effettua il riconoscimento degli oggetti di esperienza — si rivolgono direttamente a cogliere la fisionomia di fondo delle parole e delle frasi, trascurando la non perfetta corrispondenza fra termine ricercato e termine recuperato (è la cosiddetta *fuzzy searching*). Uno dei risultati di questa tecnologia è — ad esempio — che oggi per molti IR, anche con una percentuale di errore

del 50% rispetto al totale di parole trascritte non si producono effetti significativi sul recupero delle informazioni. In uno studio del 1996 ⁽¹⁵⁾ è stato presentato in via sperimentale un sistema ipertestuale di IR basato su algoritmi in grado di gestire sia errori di trascrizione e di trasmissione che testi in lingue diverse dall'inglese. Per quanto riguarda gli errori di trascrizione, il sistema è risultato tollerare fino al 30% di caratteri errati, cioè testi in cui il programma di OCR aveva trascritto praticamente ogni parola in maniera errata: si tratta di un risultato molto significativo. Confidando in questa tendenza si può ipotizzare l'applicazione dell'OCR senza correzione manuale ⁽¹⁶⁾, anche senza ricercare livelli troppo alti di esattezza. Da questi dati sembra, infatti, che la soluzione al problema della ricerca full-text su testi in formato-immagine arriverà, nei prossimi anni, dai progressi molto promettenti dei sistemi di IR più che da software infallibili di riconoscimento ottico dei caratteri.

I complessi problemi accennati — e si è volutamente tralasciato il copyright — dovrebbero dimostrare come la digitalizzazione rappresenti una sfida che attraversa trasversalmente tutto lo spettro delle competenze dei bibliotecari e dei professionisti dell'informazione. La sfida è complessa e non può certo essere banalizzata, tuttavia l'obiettivo è talmente stimolante da rendere forse interessante una semplificazione che qui si propone.

Provando a servirsi dei dati appena riportati per dare contorni più definiti al problema della digitalizzazione, risulta che la scansione (ed eventuale applicazione OCR) di un libro di 300 pagine, utilizzando uno scanner ad alimentazione automatica, oggi ha un costo di circa 30.000 lire. Già questo dato ci dice che, rispetto ai vantaggi conseguibili, il problema dei costi appare ormai superato. Già due anni fa si calcolava ⁽¹⁷⁾ che con una spesa di 3 miliardi di dollari si sarebbe potuto finanziare un programma di digitalizzazione di una copia per ciascun libro esistente al mondo (100 milioni di volumi). Come si vede, non esiste più un ostacolo di natura finanziaria di fronte all'obiettivo di superare definitivamente le biblioteche tradizionali.

Ma il problema dei costi deve essere valutato più globalmente, cioè comparativamente e in particolare considerando i costi di immagazzinamento che rappresentano una delle componenti maggiori dei costi gestionali della biblioteca tradizionale. Ai costi ormai irrisori, e fra l'altro decrescenti, delle memorie di massa, sia magnetiche che ottiche ⁽¹⁸⁾ si devono invece contrapporre i costi di immagazzinamento del libro tradizionale che, legati come sono ai valori immobiliari, presentano per la maggior parte delle biblioteche una tendenza espansiva. Il costo di im-

magazzinamento — esclusi i costi della scaffalatura — di un libro alla Biblioteca centrale dell'Università di Berkeley è di 30 dollari. Forti risparmi sono conseguibili con il ricorso a magazzini decentrati: l'Università di Harvard che dispone di un deposito distante 35 miglia dal *campus* ha abbassato questi costi fino a 2 dollari. Per la British Library si calcola rispettivamente in 69 e 14 pence il costo annuo/volume per l'immagazzinamento nella sede centrale di St. Pancras e nel magazzino decentrato. Tuttavia i magazzini decentrati rappresentano una valida soluzione solo per quei volumi che non verranno mai (o quasi) richiesti in lettura. La biblioteca dell'Università di Harvard ha calcolato in 4 dollari il costo medio della movimentazione di un volume conservato nei magazzini decentrati: si tratta di cifre alte che non possono essere ignorate in un calcolo effettivo dei costi.

Un calcolo comparativo tra costi di digitalizzazione e costi di immagazzinamento in una biblioteca con depositi nel centro storico di Roma può offrire dei dati interessanti. A seconda del tipo di scaffalatura — intensiva o meno — si ottiene un costo di immagazzinamento minimo di 20.000 lire/volume ⁽¹⁹⁾. Ma è un calcolo puramente teorico poiché è difficilmente ipotizzabile (per motivi statici) una scaffalatura intensiva distribuita sui vari piani di un edificio storico. Considerando l'effettiva densità di immagazzinamento nei depositi librari presso la sede centrale della Biblioteca della Camera dei deputati (145 volumi/mq) il costo medio sale invece a 69.000 lire/volume ⁽²⁰⁾. In sintesi, la digitalizzazione di una raccolta di 100.000 volumi attualmente ubicati in depositi nel centro storico di Roma avrebbe un costo di poco inferiore ai 5 miliardi, ma renderebbe disponibili spazi che attualmente hanno un valore commerciale di poco inferiore ai 7 miliardi. Si conseguirebbero poi molti altri vantaggi. Per limitarsi a quelli con immediato risvolto economico si possono citare almeno i maggiori: le spese per la movimentazione dei volumi, per la spolveratura o per periodiche disinfestazioni, rilegatura, restauro di volumi danneggiati. Sull'altro versante, è vero che la digitalizzazione comporta dei costi per attrezzatura tecnologiche (schermi ad alta risoluzione, distribuzione in rete per mezzo di reti a fibre ottiche, attrezzature adeguate per la stampa) ma, trattandosi di tecnologie di impiego generale e non circoscritto al progetto di digitalizzazione, rappresentano investimenti, peraltro almeno in molte realtà necessari se non già programmati, che rientrano nel processo ordinario di adeguamento tecnologico.

Pur trattandosi di un calcolo molto approssimativo, queste cifre danno conto di un aspetto che non deve essere trascurato: l'interesse delle biblioteche verso la digitalizzazione è anche radicato in quello che da alcuni anni è un motivo dominante della riflessione sul futuro delle

biblioteche, il tema delle risorse finanziarie, delle quantificazioni dei costi reali, dei risparmi conseguibili.

Ma un volume digitalizzato, al contrario di un volume di carta, è disponibile presso un numero indefinito di biblioteche. Infatti il soggetto privilegiato per la gestione di questi progetti è il consorzio più che la singola biblioteca. In questo caso i risparmi conseguibili sono, quindi, molto maggiori di quelli appena indicati. Sempre negli Stati Uniti, paese in cui le biblioteche seguono con maggiore cura i problemi di bilancio, è stato calcolato che digitalizzando i 22 milioni di opere, corrispondenti ai 400 milioni di volumi conservati dall'insieme delle biblioteche americane si conseguirebbe un risparmio di 44 miliardi di dollari. Questo calcolo risale al 1995 e ormai non risponde alla realtà poiché le tendenze dei costi, sia della gestione tradizionale sia soprattutto della digitalizzazione, sono in rapido mutamento. I costi complessivi della biblioteca tradizionale seguono un trend di aumento del 4% all'anno, fondamentalmente: costo della carta e dei magazzini librari. I costi della biblioteca digitale diminuiscono del 50% ogni 5 anni ⁽²¹⁾. Se poi si fanno raffronti più settoriali, vi sono ambiti in cui il dato è ancora più marcato. Ad esempio, nel campo delle pubblicazioni periodiche, ad un aumento, a volte una moltiplicazione, del costo degli abbonamenti corrisponde una sempre maggiore diffusione in rete di riviste elettroniche, a prezzi decrescenti: dietro questa tendenza vi è un vero e proprio riassetto strutturale di tutto il settore dell'editoria scientifica. Ma questo tema non potrebbe essere svolto adeguatamente senza fare riferimento a quello che rappresenterà, nei prossimi anni, uno dei problemi decisivi per l'evoluzione della biblioteca digitale, il trattamento del copyright, problema che, come si è già detto, rimane fuori dall'ambito di questa ricerca.

Per concludere sui costi, entro 2/3 anni — anche per la singola biblioteca — sarebbe sicuramente più conveniente — almeno sul piano semplicemente finanziario e non considerando i problemi di copyright — dotarsi dell'attrezzatura necessaria e passare allo scanner tutti i volumi acquistati piuttosto che immagazzinare libri di carta, anche nel magazzino decentrato più a buon mercato.

3 - Il project planning e il project management

In questo paragrafo non si tenterà di descrivere l'effettiva pianificazione e gestione di un progetto settoriale di digitalizzazione, sia per motivi di spazio, sia perché l'organizzazione più appropriata — dallo studio

preliminare, alla scansione delle fasi operative, al controllo di qualità — dipende dalle dimensioni del progetto, che possono variare moltissimo, e dalle tipologie, che possono essere anch'esse molto eterogenee. Non può esistere, quindi, una modalità unica di gestione dei progetti, anche se esistono delle metodologie generali a cui potrà essere utile riferirsi in fase di realizzazione: ad esempio, la British Library, per alcuni suoi progetti, ha utilizzato la tecnica di project management PRINCE ⁽²²⁾.

Per la Biblioteca della Camera la tipologia dei progetti in corso, come si vedrà meglio più avanti, è molto varia. Per la conversione retrospettiva del catalogo l'obiettivo principale è quello che i record convertiti siano conformi allo standard UNIMARC: tutti i passaggi del progetto — studio della raccolta da digitalizzare, scelta fra le varie opzioni disponibili e degli interlocutori esterni, stesura delle specifiche tecniche, attività di formazione legate al progetto — sono informate a questo obiettivo, cioè ad ottenere una certa struttura dei dati da cui discenderanno poi automaticamente tutte le successive applicazioni informatizzate. Nel caso invece della conversione della raccolta di Atti parlamentari, trattandosi di dati prevalentemente non strutturati, o meglio semi-strutturati, vengono in primo piano altri problemi e il progetto si configura essenzialmente come scelta della migliore procedura e come coordinamento di diverse tecnologie informatiche: processo di scansione, estrazione dei dati ed indicizzazione, eventuale codificazione, manipolazione dei dati in fase di visualizzazione, sottosistema di storage, sottosistema di output e di stampa, piattaforma software e struttura del database. I due progetti verranno pianificati e gestiti in modo molto differente.

In questo breve paragrafo si vuole, invece, mettere a fuoco la dimensione organizzativa e progettuale guardando di più all'insieme che ai progetti settoriali. Si tratta in primo luogo di evidenziare come il passaggio dalla biblioteca tradizionale alla biblioteca digitale, l'evoluzione verso questo nuovo ambiente, per le implicazioni che si sono indicate, non rappresenta semplicemente l'applicazione ai contenuti già noti di una sovrastruttura tecnologica nuova, ma spinge verso un ripensamento più globale, dei servizi offerti, dell'organizzazione e delle priorità della struttura-biblioteca. A monte del project planning e del project management troviamo, nelle esperienze già realizzate di biblioteca digitale, quella che potrebbe definirsi come un' idea-guida e che è destinata ad orientare tutte le fasi della progettazione e dell'esecuzione dei diversi segmenti. Per quanto riguarda i grandi progetti di creazione di biblioteche digitali, possiamo distinguere due differenti tipi di idee-guida, a cui corrispondono due diversi modi di procedere.

Secondo la prima di queste modalità, la decisione sui materiali da digitalizzare è sostanzialmente opportunistica: selezionare oggetti dalla raccolta posseduta, scegliendo quelli per i quali non potranno sorgere complicazioni sul terreno del copyright. Un esempio di questo genere è offerto dal Library of Congress Digital Library Program, secondo l'originaria denominazione. Il piano si è evoluto, successivamente, nel LC National Digital Library Program condotto in collaborazione con la Ameritech Foundation che garantisce un contributo finanziario di 2 milioni di dollari, assumendo una dimensione cooperativa con l'obiettivo di creare una collezione distribuita di risorse digitali provenienti da varie biblioteche. La parte del progetto gestita dalla Library of Congress si chiama ora American Memory. Il piano — pur ambizioso e costoso — prevede tuttavia una selezione di materiale fuori dal copyright: collezioni storiche per le quali i diritti siano scaduti o opere prodotte da agenzie governative che non reclamano i diritti. Qui si mette in primo piano la procedura operativa che si tende a liberare da complicazioni di carattere legale e da eccessivi oneri finanziari. Ma la filosofia ispiratrice di questo progetto, nonostante l'approccio apparentemente riduttivo, è invece di carattere universale: si tratta di far transitare dal supporto cartaceo verso un supporto più stabile, e quindi verso il futuro, il patrimonio culturale che abbiamo ereditato dal passato. Le biblioteche digitali sono qui lo strumento attraverso cui si realizza questo sforzo collettivo che ha un obiettivo finale ben definito. Il più famoso dei progetti ispirati a questo principio è il Progetto Gutenberg, la cui idea originaria risale addirittura al 1971 ⁽²³⁾.

Nella seconda modalità, al contrario, è dato massimo risalto alla finalità immediata del progetto, alla qualità e al significato culturale del risultato che si intende conseguire, facendosi carico e cercando soluzioni adeguate, in gran parte contrattate con gli editori, al problema del copyright. È l'approccio scelto — fra gli altri — dalla British Library. Ma come caso più rappresentativo di questo secondo orientamento si può citare uno degli esempi di *digital library* più impegnativi oggi in via di realizzazione: l'Alexandria Project dell'Università di California Santa Barbara ⁽²⁴⁾, finalizzato alla creazione di una raccolta in formato digitale di materiale specialistico, in gran parte cartografico, sulle scienze geografiche e dello spazio. Attorno all'ente promotore, la Davidson Library, si sono raccolti una serie di soggetti (università, enti governativi, società private) con l'obiettivo di: sviluppare un ambiente unico per la consultazione via Internet di un vasto patrimonio di dati e di metadati già disponibile in formato digitale; digitalizzare documenti attualmente in

formato analogico; associare ogni oggetto presente nella biblioteca digitale ad una o più localizzazioni sulla superficie terrestre. Qui lo scopo è quello di creare una raccolta distribuita specializzata e molto mirata finora non esistente come *corpus* documentario unitario.

Come si vede si tratta di due modi di affrontare il problema della digitalizzazione molto differenti, probabilmente derivanti dalle caratteristiche e dalle missioni differenti delle biblioteche promotrici.

Nel momento in cui per la Biblioteca della Camera dei deputati verrà configurandosi un più organico programma di biblioteca digitale sarà, dunque, sempre più necessario che una finalità fortemente riconoscibile venga individuata. Il passaggio dalla biblioteca tradizionale alla biblioteca digitale, infatti, non deve realizzarsi come una sommatoria di progetti di conversione separati l'uno dall'altro; semplici passaggi tecnologici, dipendenti — in ultima analisi — dalla disponibilità di risorse finanziarie. Questa caratteristica unitaria è, infatti, riscontrabile nelle esperienze più significative ⁽²⁵⁾: per le biblioteche questa transizione deve procedere per linee progettuali marcate, cogliendo le opportunità offerte dalle tecnologie per ridefinire le priorità del servizio, legando i successivi passi verso la creazione della biblioteca digitale a progetti ben riconoscibili e innovativi, nei contenuti e non semplicemente nella loro veste tecnologica. Questo è un dato generale, che non riguarda solo i grandi progetti. L'evoluzione verso la digitalizzazione deve, anche nelle realtà minori, connettersi all'innovazione dei servizi, allo stesso ripensamento delle missioni.

Un esempio significativo può ricavarsi dal progetto, varato per ora in via sperimentale, di un più puntuale raccordo fra il Servizio Biblioteca e le 14 commissioni permanenti durante tutto il corso dell'esame delle proposte di legge. Dall'esperienza realizzata in questi ultimi mesi è forse prematuro far discendere già un disegno dai contorni definiti ed infatti all'interno del Servizio tale nuovo impegno viene seguito con grande cura, ma anche con atteggiamento aperto: il contributo che una struttura come la Biblioteca può offrire alla documentazione di corredo all'iter di formazione delle leggi, e in particolare nella fase dell'esame in commissione, presenta, infatti, una grande varietà di tipologie e quindi una ricchezza di possibili ipotesi di lavoro. Ciò rende prematura una definizione troppo stringente di prodotti o strutture. Tuttavia vi è un elemento caratteristico che emerge con nettezza dall'esperienza in corso: la tendenza ad indirizzare un'attività di documentazione e ricerca direttamente verso le commissioni parlamentari e quindi verso la sede in cui si forma la maggior parte dei testi legislativi, e nel corso stesso della loro

formazione. Alla valorizzazione di tale elemento in termini progettuali può dare un contributo significativo l'ambiente della biblioteca digitale.

L'idea-guida di una raccolta digitale di documentazione di supporto al procedimento legislativo potrebbe nascere dall'elemento che si è appena detto. La struttura sarebbe modellata sul profilo delle competenze e dei calendari delle commissioni permanenti. I suoi tratti caratteristici: l'immissione all'interno della rete locale della documentazione bibliografica e di legislazione comparata, la progettazione di un accesso unico e immediato per la documentazione più recente e di supporto diretto alle proposte in esame, un disegno di percorsi di approfondimento attraverso legami a raccolte digitali ulteriori, residenti sulla rete geografica o in archivi locali. Infine un forte impegno nella manutenzione e nel costante aggiornamento della biblioteca digitale.

Il progetto — ipotetico — non sarà semplicemente un progetto di digitalizzazione o di conversione, di acquisizione di opere in formato elettronico e di collegamento a basi dati remote. Esso dovrà contestualmente affrontare i seguenti problemi:

- architettura dell'accesso all'informazione digitale, cioè soluzione dei problemi intellettuali (indicizzazione) e tecnici (interconnessione, livelli di autorizzazione)

- gestione del copyright

- cooperazione interistituzionale.

Quest'ultimo aspetto deve essere valutato con grande cura sin dalle fasi iniziali di ogni progetto. Infatti è possibile individuare qui un'altra specificità propria dell'ambiente della biblioteca digitale: empiricamente è facile osservare come moltissimi dei progetti siano il risultato di una collaborazione fra una pluralità di istituzioni. In primo luogo per motivi economici: fronteggiare gli alti costi. In secondo luogo perché biblioteca digitale significa spesso biblioteca distribuita. Ma probabilmente anche per un motivo più profondo: le sfide proposte da questa transizione sono di dimensioni troppo grandi per essere fronteggiate dai singoli istituti. Si tratta di competenze in tecnologie informatiche complesse e varie, ma non solo. È necessario conoscere standard e tecniche di indicizzazione, seguire importanti ricerche in corso sulla struttura dei testi e sulla loro codifica ⁽²⁶⁾, oppure — su un altro versante — sui metodi di rappresentazione della conoscenza, presupposto alle problematiche degli information retrieval. Ma poi non si possono certo ignorare i numerosi problemi legali: non solo il copyright, ma anche, ad esempio, nuove

questioni che si presentano sul piano dell'autenticità e della responsabilità dei documenti digitali. Infine vi sarebbe anche quello che, secondo alcuni, è addirittura la principale sfida odierna per le biblioteche: acquisire competenze ed elaborare strategie efficaci per programmare ed ottenere adeguati finanziamenti. Qualunque progetto significativo di biblioteca digitale non può trascurare nessuno di questi aspetti: ecco probabilmente il motivo per cui la collaborazione fra più soggetti diventa un passaggio obbligato.

Come esempio molto avanzato di cooperazione si può citare l'American Heritage Project ⁽²⁷⁾. Il progetto ha la finalità di creare un archivio virtuale che unifichi — in un database distribuito e con un'interfaccia unica — gli archivi di fonti primarie relative al XIX e agli inizi del XX secolo di quattro grandi biblioteche accademiche statunitensi. L'obiettivo è quello di riunire in un unico database sia i documenti stessi, sia tutti gli strumenti di ricerca descrittivi: indici, guide, inventari, note ecc. Il progetto ha una sua complessità poiché non si tratta solo di digitalizzare pagine dei manoscritti, ma di far confluire in un unico database documenti che presentano formati nativi diversi (manoscritti, dattiloscritti, testi stampati, registrazioni sonore, immagini in movimento, grafici, oggetti tridimensionali ecc.) e inoltre di creare standard per l'indicizzazione. I quattro enti promotori del progetto hanno deciso di creare dei record descrittivi a livello delle raccolte utilizzando il formato USMARC; di convertire le guide, gli indici e tutti gli strumenti descrittivi delle raccolte stesse in SGML; di collegare le guide, il cui testo integrale è reso così ricercabile, ai documenti digitalizzati. Il risultato finale sarà un database unico costituito da record USMARC relativi alle raccolte, o sezioni di raccolta. Quando di una raccolta è disponibile un indice o una guida descrittiva, un semplice bottone consentirà di raggiungere attraverso un localizzatore uniforme il documento SGML in full-text. Da qui, attraverso legami ipertestuali si accederà ai documenti primari in formato digitale. In questo progetto le istituzioni interessate non si limitano ciascuna a riversare i propri documenti, ma sviluppano anche gli standard e la ricerca che precede l'effettiva formazione della raccolta. Così l'Università di Berkeley, attraverso il Berkeley Finding Aids Project sviluppa la parte che riguarda la traduzione in SGML degli strumenti di ricerca, mentre l'Università di California, attraverso il California Heritage Digital Image Access Project, cura la parte che riguarda la rappresentazione digitale dei documenti primari, e quindi i problemi prevalentemente tecnologici di trattamento delle immagini. Si tratta, quindi, di un progetto di cooperazione molto articolato, in cui le biblioteche coinvolte, attraverso un

intreccio di ruoli e di competenze che riguarda tutte le fasi progettuali ed esecutive, creano di fatto un'unica biblioteca digitale.

Si può concludere questo paragrafo, dedicato alla dimensione organizzativa della digitalizzazione, facendo riferimento alla storia recente delle biblioteche e soprattutto all'avvento dell'automazione. Già a partire dagli anni '70, l'automazione dei cataloghi aveva proposto alle biblioteche il tema della cooperazione in termini nuovi. Con l'espressione rete di biblioteche, infatti, non si intende più — almeno da venti anni — un'istituzione, ma una serie di funzioni condivise da più centri, imperniata sull'esistenza di un catalogo automatizzato: l'elemento fisico e territoriale, così come l'elemento giuridico e istituzionale sono del tutto in secondo piano. Nella biblioteca digitale questa tendenza viene ad accentuarsi. Per identificare questa dimensione si usa quella che è già una nuova locuzione: *large-scale cooperation* che significa necessità di cooperazione anche per le più grandi biblioteche del pianeta (finora autosufficienti), necessità di cooperazione internazionale, necessità di cooperazione con il settore privato.

Questa, naturalmente, la tendenza più globale, ma anche sulla più piccola scala di una realtà quale la Biblioteca della Camera, occorre cogliere bene il senso di questa evoluzione. Ciò che si realizza, non solo nel catalogo, ma anche nelle raccolte e nei servizi, è proprio il superamento dei confini della singola istituzione verso il nuovo ambiente dell'accesso integrato alla raccolta distribuita. Ai fini del progetto non ha rilievo la collocazione fisica dell'archivio digitale. Inoltre, pur avendo gli istituti che promuovono il progetto fini istituzionali, dimensioni, strutture finanziarie, caratteri giuridici completamente diversi ed eterogenei, questi soggetti non hanno bisogno di operare alcuna forzatura su tale realtà, ma solo di chiarire i rispettivi ruoli nella realizzazione del progetto dotandosi di protocolli inter-istituzionali per la costruzione del database cooperativo, per lo sviluppo di specifici ambiti di ricerca, per l'elaborazione di standard.

La pesante struttura logistica e giuridica — di così vitale importanza nella biblioteca tradizionale — si svuota qui di significato, perché ciò che realmente conta, ciò che determina il successo o meno del progetto, è la leggera sovrastruttura dell'accesso all'informazione.

4 - I progetti in corso

I più impegnativi progetti di digitalizzazione attualmente in via di realizzazione presso la Biblioteca della Camera dei deputati vengono qui presentati solo nelle linee generali. Una dettagliata esposizione delle

modalità e delle fasi operative, e quindi delle procedure e delle tecniche di esecuzione, è oggetto di specifici studi preparatori. Per la conversione retrospettiva del catalogo il documento tecnico è stato completato nel giugno '97, mentre per la digitalizzazione degli Atti parlamentari sarà ultimato entro l'estate '99. Quanto alla conversione delle registrazioni analitiche, non si tratta ancora di un progetto concreto, ma di un'ipotesi di lavoro che si è voluta inserire al solo fine di collocare la conversione del catalogo in un quadro di riferimento compiuto. Sul piano più strettamente operativo e descrittivo qui è sufficiente sottolineare come i due progetti principali si basino su due procedimenti e su tecnologie diverse. Nel caso della conversione retrospettiva del catalogo e delle registrazioni analitiche, infatti, oggetto dell'intervento di digitalizzazione sono record bibliografici, cioè testi strutturati in «aree»: intestazione, titoli, elementi descrittivi vari, segnature di collocazione, notazioni semantiche ecc. Data l'importanza di questa struttura ai fini del trattamento informatico, poiché ad essa corrispondono i campi del record magnetico, la tecnica più efficace e in grado di garantire il miglior risultato qualitativo è quella della digitazione e formattazione manuale delle schede. I costi sono, ovviamente, maggiori rispetto alla scansione, come si è visto nelle pagine precedenti, ma trasformare schede catalografiche in file-immagine avrebbe come risultato un catalogo inutilizzabile. Nel caso degli Atti parlamentari, invece, i testi da convertire non sono strutturati, il termine tecnico più appropriato è quello di dati semi-strutturati. La tecnica da impiegare sarà quella della immissione e codifica dei testi o quella della scansione e successiva indicizzazione, cioè apposizione di chiavi per il recupero dei file-immagine. Il passaggio più importante, la chiave di volta del progetto, non è in questo caso la formattazione del testo quanto i problemi di indicizzazione dei documenti e, in primo luogo, la possibilità o meno di effettuare il riconoscimento ottico dei caratteri.

5 - Conversione retrospettiva del catalogo

Già da molti anni — praticamente dall'introduzione nel 1984 della catalogazione on-line — l'ipotesi di trasferire su supporto magnetico il ricco patrimonio di informazioni contenuto nel catalogo a schede, in parte manoscritte, è stata più volte valutata. Negli ultimi due anni — a partire dal gennaio '97 — questa ipotesi è diventata un progetto defini-

to, sviluppato e sostenuto dal Servizio, approvato il 25 febbraio 1998 dal Collegio dei questori e attualmente in via di esecuzione per la parte preliminare relativa all'attivazione delle procedure di gara.

I motivi per cui proprio in questi anni si è arrivati a tale definizione sono sostanzialmente due.

Il primo è costituito dalla evoluzione degli standard internazionali per il formato elettronico delle registrazioni bibliografiche e dall'evoluzione delle tecnologie. Si pensi al fatto che il formato UNIMARC si è venuto affermando e consolidando in Europa, con un impegno ufficiale anche degli organi comunitari, solo agli inizi degli anni '90. Le stesse tecnologie — intese in questo caso come società di servizi capaci di gestire complessi catalografici di una certa importanza — hanno segnato, sempre parlando dell'Europa, un punto di svolta con la conclusione di tre grandi progetti di conversione retrospettiva: British Library (1993), Deutsche Bibliothek (1994) e Bibliothèque Nationale di Parigi (1995). Finora in Italia si era parlato di conversione retrospettiva solo per piccole biblioteche che, libere da problemi di uniformità a standard internazionali, hanno proceduto a sommarie rischedature o anche ad esperienze pilota con tecniche di OCR rivelatesi poco efficienti e molto costose. È solo con i due progetti della Biblioteca Apostolica Vaticana e della Biblioteca Universitaria di Venezia, conclusi nel '96-'97, a cui si sta aggiungendo ora una terza esperienza presso la Biblioteca Marciana di Venezia, che anche grandi biblioteche italiane si sono potute avvicinare a soluzioni valide del problema. Può, quindi, dirsi che solo negli ultimi due anni sono maturate le condizioni oggettive per cui una fra le maggiori biblioteche italiane potesse porsi l'obiettivo di trasferire — a costi non proibitivi — il proprio catalogo storico su un supporto magnetico, con una piena garanzia qualitativa, e in primo luogo con la garanzia di un risultato conforme agli standard internazionali.

Il secondo motivo è, invece, legato alla circostanza particolare — anch'essa maturata negli ultimi due anni — del cambio del sistema integrato di gestione della Biblioteca. È nel corso del 1999 che — concluso il lungo processo di selezione — un nuovo sistema con architettura client-server sostituirà il pacchetto DOBIS-LIBIS che segnò l'inizio dell'automazione nella Biblioteca della Camera. Con l'avvento del nuovo sistema verranno introdotte procedure automatizzate anche in nuovi segmenti gestionali (trattamento dei seriali), il catalogo (OPAC) sarà ricercabile anche in modalità *www*, diverse altre funzionalità saranno disponibili. In particolare, per quanto riguarda i record bibliografici, il passaggio al nuovo sistema permetterà di abbandonare il vecchio formato

— che era un formato proprietario, sia pure derivato da USMARC, il D-MARC — e di introdurre lo standard UNIMARC. La conversione retrospettiva del catalogo cartaceo non soffrirà quindi di questa particolare rigidità, ma sarà invece parte integrante, insieme alla migrazione dei record DOBIS (circa 100.000 registrazioni bibliografiche che andranno a loro volta convertite nel nuovo formato), della formazione di un'unica base di dati nel formato internazionale UNIMARC.

Secondo le quantificazioni effettuate in sede di studio preparatorio, il totale delle registrazioni bibliografiche da convertire sarà superiore alle 450.000. Se a questo insieme si sommano gli oltre 100.000 record DOBIS, si ottiene un complesso di circa 600.000 record bibliografici. Si tratterà, quindi, del più ricco catalogo italiano (non collettivo) on-line. Ma un apprezzamento dell'obiettivo non può prescindere da altri fattori, in primo luogo di carattere qualitativo. Le 450.000 schede cartacee che saranno convertite sono, infatti, tutte dotate di accessi semantici. Questa circostanza, del tutto peculiare del catalogo della Biblioteca della Camera dei deputati, si deve all'adozione — sin dalla nascita stessa della Biblioteca — di un sistema di classificazione molto articolato e aderente alle esigenze di supporto documentario, informativo e culturale all'attività parlamentare. Il risultato della applicazione di metodi automatici di information retrieval a un patrimonio catalografico così strutturato sarà di sicuro interesse. Un altro elemento caratteristico è l'alto numero di spogli da opere poligrafe, oltre 100.000: si tratta di registrazioni bibliografiche in gran parte uniche, non reperibili in nessun'altra biblioteca né in formato cartaceo, né tanto meno in formato magnetico, la cui presenza è dovuta alla particolare cura riservata storicamente dalla Biblioteca della Camera dei deputati alla schedatura analitica. Infine saranno le caratteristiche stesse del patrimonio della Biblioteca — ricchezza di materiale di pregio e di collezioni rare, specializzazione per alcune materie, forte presenza di pubblicazioni straniere — a dare alla banca dati che risulterà dalla conversione una originalità e un alto livello qualitativo.

6 - Conversione delle registrazioni analitiche

Ma la conversione retrospettiva del catalogo — pur considerandone i costi, l'impegno organizzativo, gli effetti sulla qualità dei servizi — va vista a sua volta come parte di un progetto più ad ampio raggio. Il processo di standardizzazione, sia sul piano delle regole di catalogazione, sia

sul piano dei formati per l'archiviazione e lo scambio di record leggibili dal computer — processo che si riassume nelle due sigle di ISBD e UNIMARC — permette oggi di porre il problema degli archivi di registrazioni bibliografici in termini globali e unitari. Inoltre l'evoluzione dei data-base relazionali consente di potenziare e semplificare enormemente le funzioni di ricerca. Si può concludere che ogni biblioteca può darsi oggi l'obiettivo di unificare tutti i propri cataloghi in un unico data-base dotato di indici unificati e ricco di tutte le funzionalità ricavabili da un data-base relazionale.

Per la Biblioteca della Camera dei deputati questo obiettivo significa principalmente unificare il catalogo in tutte le sue stratificazioni cronologiche da un lato e gli spogli da riviste e da opere poligrafe dall'altro. Come alcuni sanno, la Biblioteca della Camera possiede un ricchissimo patrimonio di spogli di riviste: secondo una prima quantificazione effettuata, non meno di 500.000 registrazioni, tutte dotate di accessi semantici. A queste registrazioni vanno poi aggiunte quelle, forse più note perché in parte pubblicate, degli spogli biografici. Questo patrimonio, che oggi rappresenterebbe una fonte di ricerca storica di grande interesse, si deve ad una ininterrotta tradizione risalente ad una illuminata delibera del Commissione di vigilanza sulla Biblioteca del giugno del 1878 e proseguita per un intero secolo fino alla nascita della banca dati RIVI. Forse è meno noto che, per l'epoca a cui risale, per l'ampiezza e sistematicità, oltre che per la qualità degli spogli, si tratta di un'impresa che non ha eguali in Europa e un unico esempio contemporaneo negli Stati Uniti.

Se la digitalizzazione delle registrazioni degli spogli per materia, attualmente manoscritte in parte su volumi e in parte su schede, e degli spogli biografici appare dunque un obiettivo non eludibile, un valore ben maggiore assumerebbe un'ipotesi di cooperazione fra un gruppo di enti per la conversione anche dei documenti primari. Con la creazione di un unico *corpus* documentario di queste dimensioni, già dotato di tutti gli accessi semantici, si darebbe vita alla più ricca raccolta digitale per la ricerca sulla storia italiana contemporanea.

7 - Digitalizzazione degli Atti parlamentari

Il progetto di conversione in formato digitale della raccolta retrospettiva degli Atti parlamentari italiani è stato già annunciato nella Relazione sullo stato dell'Amministrazione per l'anno 1997 ⁽²⁸⁾. Ai fini

della ricerca l'importanza della sua realizzazione, fra l'altro (quasi) in coincidenza con la apertura del sito Internet e la messa in rete degli Atti della legislatura corrente, non ha bisogno di essere sottolineata ed è tale da richiedere, prevedibilmente, un impegno diretto di diversi settori dell'Amministrazione. Inoltre non va trascurato anche l'obiettivo della conservazione: si tratta, infatti, di una raccolta che nella sua versione integrale non è certo disponibile in molte copie. Sulla conservazione delle risorse digitali, tema che qui non si è sviluppato, si può accennare brevemente al suo principio-base: il formato digitale, pur disponendo ancora di media meno stabili rispetto al formato analogico ⁽²⁹⁾, è considerato oggi quello meglio rispondente alle esigenze di conservazione poiché, attraverso periodiche migrazioni, permette di trasferire i documenti da una configurazione hardware/software ad un'altra e di seguire l'evoluzione tecnologica a costi molto bassi ⁽³⁰⁾.

Trattandosi di un progetto caratterizzato, come si è detto, dalla ricerca della migliore tecnologia da applicare, o meglio della migliore combinazione fra diverse tecnologie, può essere utile una rapida elencazione dei principali problemi operativi. La raccolta è attualmente disponibile su due supporti: quello cartaceo, delle dimensioni di circa 6 milioni di pagine, e il microfilm ⁽³¹⁾. Una valutazione comparativa dei costi suggerirebbe decisamente la digitalizzazione dell'edizione su microfilm. Ma, come spesso accade nella digitalizzazione in formato immagine, potrebbe esistere un tradeoff tra qualità dei risultati, tempi di conversione, costi ⁽³²⁾. La comparazione andrà accuratamente effettuata, valutando tutte le variabili: costi di controllo dell'integrità della raccolta, costi di cattura delle immagini, costi degli interventi migliorativi sulla qualità delle immagini ottenute, quali deskewing e pulizia dei margini, costi di correzione nel caso di applicazione OCR, costi dell'indicizzazione e dell'estrazione di metadati. Specifici software offrono, poi, la possibilità di effettuare operazioni più complesse di estrazione di immagini o rimpaginazione di testi stampati su colonne ⁽³³⁾.

I problemi di tecnologie e di variabili nell'acquisizione delle immagini — risoluzione, compressione ecc. — non dovrebbero presentare, in questo progetto, grandi difficoltà. Si tratta, infatti, di una raccolta omogenea — con uniformità di caratteri e priva di immagini — in cui si tratterà, dopo una fase di accurati test, di scegliere il supporto da cui effettuare la scansione.

Il vero nodo del progetto riguarda invece, come si è già detto, i problemi dell'indicizzazione, e in particolare dell'indicizzazione full-text. Per l'importanza della raccolta a cui si fa riferimento, è evidente come

non possa essere mancato l'obiettivo di un risultato qualitativo massimo, cioè all'avanguardia rispetto a ciò che la tecnologia oggi è in grado di offrire. È altrettanto importante sottolineare come — a questo fine — vada ricercato un valido modello di cooperazione e di partnership tra varie competenze. Il campo della ricerca full-text su collezioni di alto valore culturale è, infatti, un terreno d'elezione per la collaborazione fra le biblioteche e i settori più innovativi e dinamici dell'industria dell'information technology ⁽³⁴⁾.

Infine il progetto dovrà comprendere l'intero ciclo dell'intervento, fino alla scelta dei supporti per l'immagazzinamento dei dati. Da altre esperienze realizzate risulta — ad esempio — come nei progetti di scansione massiccia e veloce vengano spesso sottovalutati i problemi di immagazzinamento dei dati e come ciò possa creare imprevisti colli di bottiglia. Il progetto dovrà, dunque, articolarsi in:

- acquisizione delle immagini ⁽³⁵⁾
- estrazione dei dati e indicizzazione
- rilascio
- problemi di immagazzinamento, networking, output.

Una considerazione conclusiva sui progetti che brevemente si sono presentati: la conversione retrospettiva di un catalogo e il trasferimento su supporto digitale del testo integrale di una raccolta di migliaia di volumi. I due obiettivi sembrano disomogenei, non solo per le differenze dei supporti d'origine e finali e non solo per le differenti procedure operative di realizzazione. A ben vedere, la principale differenza risiede nella natura diversa dei due oggetti. Nel caso della conversione retrospettiva del catalogo ci troviamo infatti di fronte a un grande contenitore di metadati, mentre nel secondo caso è proprio la sede in cui risiedono i dati — i volumi o le bobine di microfilm contenenti il testo degli Atti del Parlamento italiano — che si andrà a trasferire su supporto digitale. Secondo una visione molto naturale — perché ereditata dal passato — si tratta di due oggetti differenti e separati. È sul presupposto di questa separazione, anche fisica — cataloghi e magazzini librari — che le biblioteche hanno storicamente organizzato e diviso gli spazi, hanno modellato le procedure, hanno pensato le strutture operative e direttive, hanno proposto e fornito i servizi.

Ma nell'ambiente della biblioteca digitale questa separazione non esiste più: i metadati compaiono sullo schermo sempre più integrati e frammisti ai dati dei documenti full-text. Inoltre gli stessi oggetti a cui si

riferiscono i metadati progressivamente si «frantumano» ⁽³⁶⁾: i metadati si riferiscono sempre più al capitolo piuttosto che al libro, all'articolo piuttosto che al periodico ⁽³⁷⁾.

Trattare insieme di questi progetti solo apparentemente differenti ci permette, quindi, di mettere meglio in luce uno degli elementi caratteristici delle biblioteche digitali: un catalogo e il testo di un libro, una foto, un documento sonoro, un abstract o un link ad un sito Web — che a sua volta ci porta ad un ulteriore catalogo o ad una rivista elettronica — sono ormai elementi solo apparentemente eterogenei. Tutto viene posto sullo stesso piano, potenzialmente all'interno dello stesso spazio fisico cioè sullo stesso schermo. Nella progettazione della biblioteca digitale questa potenzialità non può essere mancata ed anzi va perseguita: i grandi archivi digitali non si sovrapporranno, ciascuno impermeabile agli altri. L'architettura d'insieme dovrà prevedere sessioni di lavoro uniche in cui gli archivi si intrecciano con facilità attraverso una rete di legami, e in quest'architettura andranno scomparendo molte divisioni — divenute ormai fittizie — fra cataloghi e documenti primari.

8 - *Futuri sviluppi*

Affinchè una biblioteca che dispone di raccolte digitali — di dati o di metadati — si evolva in una biblioteca digitale sono necessari alcuni passaggi ulteriori. In primo luogo deve trovare spazio la volontà e la capacità di affrontare in modo globale le questioni poste dalle raccolte digitali. Un esempio ben documentato di questo approccio lo fornisce la British Library con il suo Digital Library Programme in cui i vari progetti di digitalizzazione non vengono trattati separatamente, ma posti in un quadro d'insieme di cui fanno parte anche il trattamento delle opere in formato digitale ricevute per acquisto o per deposito legale, nonchè i servizi — commerciali, gratuiti o forniti dalla stessa British Library — diffusi via network. Inoltre il programma mira a coprire l'intero spettro dei problemi — tecnologici, economici, gestionali, legali, di accesso, formativi, di partnership con soggetti esterni ⁽³⁸⁾.

Ma accanto a questo obiettivo generale, affrontare la biblioteca digitale con un approccio globale, vi sono dei passaggi più operativi da realizzare. Per descriverli con maggiore chiarezza può essere utile fare ancora una volta riferimento al recente passato dell'automazione bibliotecaria.

L'automazione delle biblioteche si è indirizzata — a partire dagli anni '70 — fondamentalmente sui cataloghi (e di conseguenza sulle procedure amministrative) per consentirne l'accesso remoto. Questa direttrice di sviluppo, orientata ad un salto qualitativo dei servizi offerti all'utenza, ha portato nell'arco di un ventennio all'attuale evoluzione degli OPAC. Questi rappresentano oggi il massimo risultato raggiunto nella applicazione di strumenti informatici ai servizi bibliotecari: nell'OPAC si cerca oggi di unificare tutti i servizi forniti all'utenza e questo ha cambiato il modo di funzionare dell'intera struttura. La sfida della biblioteca digitale sarà quella di portare agli utenti — sugli stessi canali dell'OPAC e in ambiente Internet/Intranet — anche i documenti primari. Quello che può prefigurarsi è un'architettura basata su un'interfaccia unica di accesso a tutti i dati in formato digitale ⁽³⁹⁾.

Per realizzare questa evoluzione è necessario:

- elaborare un' interfaccia
- individuare le raccolte e disegnare un'architettura dell'accesso.

9 - *Elaborare un'interfaccia*

L'elaborazione di un'interfaccia è in genere un'obiettivo da perseguire attraverso il confronto e la cooperazione fra più competenze. Le competenze informatiche sono, infatti, necessarie ma non sufficienti, poiché il successo della progettazione dipende da una conoscenza delle raccolte e del loro uso. In alcuni casi, gli stessi utenti sono stati (indirettamente) coinvolti nella fase di studio che prevede, comunque, la definizione di «scenari» di utenza, cioè la prefigurazione di percorsi differenziati di ricerca ⁽⁴⁰⁾.

I problemi da affrontare non riguardano genericamente l'arricchimento dei servizi da mettere a disposizione sfruttando le potenzialità offerte dalla rappresentazione digitale. Quest'ultima infatti modifica le strategie di ricerca e impone in primo luogo di trovare una soluzione valida ai problemi della navigazione. Nonostante il successo del Web, in realtà molti studi hanno evidenziato le carenze e gli effetti dispersivi della navigazione ipertestuale, gli svantaggi, sia per chi scrive che per chi legge, consistenti nel dover ricorrere a continui salti per la comprensione dell'insieme del testo e dei suoi riferimenti. Dalla comparazione, effettuata attraverso studi sperimentali, tra la comprensione di un ipertesto e dello stesso testo nella versione lineare, risulta invariabilmente un livello

di comprensione migliore e una maggiore ergonomia di quest'ultimo. Di fronte a siti che propongono un insieme di raccolte, il navigatore finisce quasi sempre con l'essere distratto e con l'abbandonare le raccolte e i loro contenuti — verso cui si era in un primo momento indirizzato — per disperdersi in una navigazione ormai fine a se stessa. La risposta a questi problemi risiede in una serie di accorgimenti che mirano a migliorare continuamente le interfacce. Tali accorgimenti andranno attentamente valutati in relazione alle raccolte, agli «scenari» e ai profili di utente che si sono prefigurati. In generale può dirsi che nelle biblioteche digitali è particolarmente utile appiattire la gerarchia dei legami ed ancorare la navigazione al contesto e comunque ancorare l'utente alle raccolte rendendolo consapevole dei suoi contenuti, anche con l'uso appropriato di esempi, overviews, previews.

Un altro sviluppo interessante nelle interfacce di biblioteche digitali è quello di consentire la creazione di cartelle multiple di ricerca personalizzate in cui l'utente può conservare memoria delle ricerche e dei link con i siti utili di volta in volta visitati ⁽⁴¹⁾.

Infine una grande attenzione va rivolta alla elaborazione e presentazione di indici: tutti i siti di più recente progettazione offrono la possibilità di puntare a pagine di indice da qualunque pagina del documento. Tuttavia anche gli indici tradizionali non sono un ausilio sufficiente quando da un sito si diparte una rete di accessi a raccolte multiple. Una ricerca che promette sviluppi innovativi in questo campo è quella del WebTOC ⁽⁴¹⁾: nuovo strumento per la navigazione ipertestuale messo a punto dall'Human-Computer Interaction Laboratory (HCIL) e una cui applicazione è visibile sul sito della biblioteca digitale della Library of Congress, *American Memory*. Il programma funziona attraverso un software per l'analisi delle pagine Web ed un altro per la visualizzazione. Le dimensioni di ogni pagina Web vengono quantificate, in termini di testo, immagini, suoni, immagini in movimento. I documenti vengono quindi rappresentati gerarchicamente e l'intero contenuto del sito è visualizzabile in forma di Table of contents ad istogrammi colorati di lunghezza corrispondente alle dimensioni dei documenti.

10 - Individuare le raccolte e disegnare un'architettura dell'accesso

Come si è già premesso in apertura, è solo da un impegno collettivo e istituzionale che potrà derivare un progetto generale di evoluzione della Biblioteca della Camera verso il modello di biblioteca digitale. Inoltre,

per la crescente integrazione fra le varie componenti della documentazione per il Parlamento, tale progetto rientrerà a sua volta in un processo più ampio riguardante più soggetti. In questa sede non si è fatto che un fugace accenno a questa complessa tematica (peraltro ancora del tutto aperta), quindi l'indicazione delle raccolte e il riferimento al disegno complessivo degli accessi non è ancora uno schizzo di un concreto disegno generale, ma semplicemente una prima traccia di biblioteca digitale possibile, in quanto tale — inevitabilmente — ancora molto lontana da un piano di lavoro.

In una fase iniziale si possono comprendere — sotto la denominazione comune di Biblioteca digitale i seguenti archivi:

In primo luogo il catalogo unificato in formato UNIMARC: qui potrebbe confluire, attraverso un'apposita conversione, o collegarsi attraverso un'integrazione dei sistemi di interrogazione, oltre all'attuale base di dati RIVI, anche la versione digitale delle registrazioni analitiche precedenti al 1970.

Raccolta digitale retrospettiva degli Atti parlamentari italiani.

Collezioni speciali: le raccolte speciali che maggiormente si prestano a interventi di digitalizzazione potrebbero essere individuate, nella Biblioteca della Camera dei deputati, attraverso uno studio specifico. Sicuramente le collezioni antiche, fra cui la collezione Kissner, alcuni periodici rari di interesse storico, oppure le preziose raccolte legislative degli stati preunitari. Un altro terreno di ricerca stimolante potrebbero rivelarsi le serie storiche di documenti ufficiali, spesso dotate di documentazione tecnica di corredo, o alcune sezioni della raccolta (retrospettiva e corrente) di letteratura grigia.

Ricerca bibliografica: si tratterebbe di unificare l'accesso alle sempre più numerose banche dati bibliografiche su CD-ROM gestite attraverso drive ottici integrati con juke boxe, o direttamente attraverso l'infrastruttura di rete.

Multimedialità: rappresenterebbe in un primo momento il link ad archivi di immagini o di file sonori o di immagini in movimento, quali collezione delle carte geografiche, ma anche nuove iniziative di acquisizione e sviluppo di raccolte (ad esempio, raccolta della versione digitale di filmati di interesse storico politico). Le opportunità per un istituto come la Biblioteca della Camera dei deputati sul terreno della multimedialità — soprattutto in termini di collaborazione con altri servizi informativi — potrebbero formare oggetto di uno specifico programma di ricerca e di intervento ⁽⁴³⁾.

Infine nella biblioteca digitale dovrà trovare posto quell'insieme di materiali che vengono prodotti nella intensa ed articolata attività di

ricerca che caratterizza la Biblioteca della Camera. In primo luogo i materiali di documentazione bibliografica e di legislazione comparata prodotti dalla Biblioteca nella forma del dossier, ma successivamente il progetto potrebbe evolversi secondo una direttrice generale che è caratteristica dell'ambiente della biblioteca digitale. Nelle biblioteche fisiche non si conserva memoria, e tanto meno si indicizzano, i risultati dell'attività di reference, se non nella forma di agili archivi di lavoro sui temi più ricorrenti, costruiti su base empirica (i cosiddetti *tickler files*). In una realtà diversa, in cui si sia affermata la gestione elettronica dei documenti (GED), con unità diffuse di input — scansione, OCR, indicizzazione — e collegamento in rete, può essere posto il problema della creazione di una memoria di questa attività di ricerca. Nella Biblioteca della Camera dei deputati il reference è un'attività ricca e spesso dotata di un forte carattere di specificità. Pur necessitando di un intervento selettivo, sicuramente la memoria di questa attività potrebbe costituire la base per un archivio digitale di interesse non solo interno.

BIBLIOTECA DIGITALE

Catalogo UNIMARC Spogli bibliografici

(Web TOC)

Atti parlamentari

Collezioni speciali

Ricerca bibliografica

Multimedialità

Ricerche e dossier

Il processo formativo di una biblioteca digitale avverrà (sta avvenendo) in una fase iniziale anche attraverso sovrapposizioni di prodotti non compatibili e non integrati, ma è facile prevedere che le interfacce di accesso saranno sempre più amichevoli e uniche, come è avvenuto per l'OPAC.

Tuttavia è bene ricordare che dietro le interfacce vi sono i problemi di struttura degli archivi, cioè di:

1. Formati dei dati (che saranno diversi, ma integrati ⁽⁴⁴⁾)
2. Organizzazione del contenuto dei documenti (indicizzazione)
3. Immagazzinamento dei dati.

È da una strutturazione coerente degli archivi (e non semplicemente da un dispositivo tecnologico) che può derivare un risultato qualitativamente alto. Questo problema rinvia ad un altro che merita una breve —

e conclusiva — riflessione: la centralità del problema dell'accesso all'informazione.

È stato notato uno stretto nesso fra evoluzione delle biblioteche verso il modello della biblioteca digitale e crescita dell'importanza relativa dei servizi per l'utenza. Questa tendenza è stata più volte verificata anche nel contesto della Biblioteca della Camera dei deputati: le funzioni tradizionali più strettamente legate alla gestione amministrativa si semplificano, perdendo parte del loro contenuto professionale, soprattutto stanno perdendo la loro posizione di centralità, anche nel processo formativo della professionalità dei bibliotecari. Le principali sfide nella biblioteca del prossimo futuro saranno invece sul terreno dei servizi, e in particolare dei servizi informativi.

Fra l'altro, partendo da questo dato, sarebbe utile anche riconsiderare le funzioni manageriali che nella biblioteca fisica erano equamente distribuite secondo uno schema che partiva dalla selezione e acquisto delle opere, per poi proseguire con la catalogazione e l'indicizzazione, la conservazione, l'informazione al pubblico ⁽⁴⁵⁾. Le competenze professionali, così come le strutture organizzative e direttive erano, cioè, modellate sullo schema dell'iter del libro, schema in cui — fra l'altro — non trovano posto proprio quelle in cui molti vedono le nuove funzioni di punta: nuove tecnologie, aspetti legali, gestione economica. Questo schema derivato dall'iter del libro oggi appare a tutti obsoleto.

Molti dei motivi di questa inadeguatezza sono riconducibili all'avvento della biblioteca digitale: qui il nucleo strategico è infatti meno distribuito e più compatto, e soprattutto non può essere imperniato sulla gestione amministrativa da un lato e sui servizi di reference dall'altro, intese quali realtà giustapposte e dotate ciascuna di una propria autonomia. Nella biblioteca digitale il nuovo nucleo decisionale e progettuale è rappresentato unitariamente dall'accesso all'informazione, in termini gestionali (autenticazione, autorizzazioni, servizi) e in termini di struttura concettuale e di rappresentazione dei dati: l'accesso all'informazione pertinente si può realizzare solo attraverso una forte coerenza fra mappa dei servizi e delle autorizzazioni e mappa dei percorsi e degli strumenti concettuali. Le capacità di innovazione dovranno quindi essere concentrate sull'obiettivo di dar vita a un disegno, un'architettura complessiva di accesso all'informazione coerente, innovativa, rispondente ad esigenze reali. Di fronte a questa sfida passeranno in secondo piano la gestione degli spazi fisici, ma anche le raccolte e le stesse attrezzature tecnologiche (se male utilizzate) che hanno rappresentato in passato i caratteri identificativi di una biblioteca. Sarà invece quest'architettura,

l'architettura degli accessi all'informazione, che costituirà la vera fisio-nomia della biblioteca, ciò per cui essa sarà riconosciuta e valutata.

Dietro l'interfaccia unica dei documenti in formato digitale vi è, quindi, il problema della struttura degli archivi. E dietro alla struttura degli archivi vi è il principale nodo: quello di un'architettura dell'accesso all'informazione. Un problema che a prima vista si presenta con le caratteristiche del problema tecnologico, da delegare magari a tecnici informatici, ne comprende in realtà altri, che non possono essere delegati, ma su cui è utile invece investire le risorse più qualificate di cui dispongono le biblioteche e formarne di nuove attraverso un impegno di acquisizione di competenze e di esperienze sempre nuove.

Note

(¹) GRIFFIN S.M., p. 24.

(²) COLLIER M.

(³) CRAWFORD W. - GORMAN M., pp. 180-181. «*And*» *Not* «*Or*» si intitola significativamente il capitolo conclusivo di quest'opera (molto conosciuta) scritta da due fra i massimi studiosi dell'automazione bibliotecaria, opera che rappresenta un po' il manifesto della linea di pensiero citata.

(⁴) Alcuni interventi in questo dibattito, a partire dal documento forse più famoso, il Rapporto Benton (1996), sono anche riportati in un numero monografico della rivista *Library Trends*: Goldhor H.

(⁵) BEASTALL G., pp. 31-32.

(⁶) LISCIA R.

(⁷) I dati quantitativi riportati in questo paragrafo sono ricavati per lo più da: LESK M. [1] e LESK M. [2].

(⁸) Il TB (1 Terabyte corrisponde a 1000 GB) è l'unità di misura usata per calcolare i grandi archivi di memoria digitale. Ma è una misura ancora insufficiente per chi ha voluto tentare calcoli globali: per scansire l'intera Library of

Congress e ottenere tanti file-immagine quante sono le sue pagine, e digitalizzare tutti i file sonori e di immagini in movimento posseduti, occorrerebbero, infatti almeno 3 PB di memoria (1 Petabyte corrisponde a 1000 TB).

(⁹) Il termine conversione non va confuso con il termine migrazione «the process of moving data from one computer system to another without converting the data» (*IBM Dictionary of Computing*, N.Y., McGraw-Hill, 1993, p. 433).

(¹⁰) Inoltre è bene chiarire che i miglioramenti riguardano per lo più la velocità di riconoscimento dei caratteri (da una media di 25 caratteri/secondo si è passati in pochi anni ad una media di oltre 200 caratteri/secondo). Sul piano dell'accuratezza e dei margini di errore invece i progressi non sono così rapidi.

(¹¹) Ma fra non molto verranno immessi sul mercato, a poche migliaia di lire (14 dollari), i nuovi dischetti fat floppy (tecnologia HiFD) con una capacità 140 volte superiore (200 MB), su cui potranno trovare posto ben 100.000 pagine di testo in caratteri ASCII.

(¹²) Si tratta di costi puramente orientativi, vista la grande variabilità di elementi coinvolti.

(¹³) CARPENTER L. - SHAW S. - PRESCOTT A., pp. 66-67

(¹⁴) Per quanto riguarda i formati più diffusi, tema molto ampio a cui in questa sede si sono potuti fare solo rapidi accenni, vedi, ad esempio: KASDORF B. che analizza comparativamente i due formati SGML e PDF. Ma al tema del linguaggio SGML, per le sue grandi implicazioni metodologiche oltre che operative, non può, in questa sede, che farsi un semplice rinvio.

(¹⁵) PEARCE C. - NICHOLAS C. (1996).

(¹⁶) Questa affermazione non può naturalmente avere valore assoluto: le scelte concrete deriveranno da un'attenta ponderazione delle caratteristiche del singolo progetto.

(¹⁷) LESK M. [2], p. 3.

(¹⁸) Mentre le memorie ottiche sono state prodotte sin dall'inizio con costi relativamente bassi, i costi delle memorie magnetiche hanno mostrato un trend decrescente marcantissimo: nel 1956, quando fu costruito il primo disco magnetico il costo era di circa dollari 10.000/MB; oggi è di circa 40.000 volte inferiore.

(¹⁹) Si è considerato un valore commerciale medio dei locali che ospitano le scaffalature di 10 milioni/mq.

(²⁰) La cifra è alta, ma non deve impressionare: per la Biblioteca Nazionale di Parigi, la grande opera realizzata a Tolbiac, il costo di immagazzinamento è stato calcolato in 150.000 lire/volume!

(²¹) GARRETT J. - WATERS D.

(²²) CARPENTER L. - SHAW S. - PRESCOTT A., p. 19.

(²³) <http://www.promo.net/pg/>.

(²⁴) <http://www.alexandria.sdc.usbc.edu>.

(²⁵) Una pagina con tutti i links ai principali progetti di digital library è tenuta costantemente aggiornata all'URL: <http://www.glue.umd.edu/~dlrg/otherdl.html>.

(²⁶) Su questo argomento si rinvia ad una sigla: Text Encoding Initiative (TEI) e ad un'opera: IDE N. - VERONIS J.

(²⁷) <http://sunsite.berkeley.edu/amher>.

(²⁸) Camera dei deputati, 1998, pp. 51-52.

(²⁹) Il microfilm può arrivare ad una stabilità media valutata in 500 anni, mentre per i dischi ottici si calcolano 100 anni e per i nastri magnetici addirittura 30.

(³⁰) BELLINGER M., pp. 179-180.

(³¹) I microfilm sono dotati di un sistema CAR (Computer Aided Retrieval) per la localizzazione veloce della pagina desiderata.

(³²) Per una descrizione dettagliata di un progetto di digitalizzazione da microfilm realizzato presso la British Library, vedi: CARPENTER L. - SHAW S. - PRESCOTT A., pp. 68-72.

(³³) Vedi: LESK M. [2], pp. 64-67.

(³⁴) Come esempio fra i più significativi può citarsi la collaborazione fra British Library e Excalibur. Vedi: CARPENTER L. - SHAW S. - PRESCOTT A., pp. 178-190.

(³⁵) L'ipotesi dell'acquisizione di un formato-testo e di una eventuale codificazione SGML/XML, almeno per le parti più significative della raccolta, non viene qui illustrata poiché oggetto di uno specifico studio preparatorio attualmente in corso.

(³⁶) *Granularity* è il termine tecnico che definisce la tipologia (e la qualità) dei metadati. Per un uso e una spiegazione del termine, si veda, ad esempio, G. MARCHIONINI.

(³⁷) Anche la Biblioteca della Camera ha realizzato un progetto (che presto sarà a disposizione del pubblico) per la diffusione in rete locale dei TOC (Table of Contents) dei fascicoli dei periodici.

(³⁸) CARPENTER L. - SHAW S. - PRESCOTT A.: in particolare il saggio conclusivo di Brian Lang.

(³⁹) Per un'introduzione alle tecnologie delle interfacce e delle architetture di sistema per le biblioteche digitali, vedi: MUKAIYAMA H.

(⁴⁰) L'esperienza condotta dallo staff che ha elaborato l'interfaccia per la National Digital Library della Library of Congress è descritta in PLAISANT C. - MARCHIONINI G. - KOMLODI A.

(⁴¹) Una descrizione dettagliata del prototipo (Artemis) di un'interfaccia di ricerca ricca di funzionalità innovative è sul sito della Digital Library dell'Università del Michigan: [<http://www.si.umich.edu/UMDL/>].

⁽⁴²⁾ Il documento che offre una illustrazione completa del progetto è: NATION D.A. - PLAISANT C. - MARCHIONINI G. - KOMLODI A.

⁽⁴³⁾ È proprio nell'integrazione con altri fornitori pubblici e privati di informazione — radio, televisioni, network locali — che il citato Rapporto Benton individua la più promettente linea strategica di sviluppo per le biblioteche nell'era digitale.

⁽⁴⁴⁾ Vedi l'esempio sopra riportato dell'American Heritage Project.

⁽⁴⁵⁾ Anche la attuale struttura quadripartita della Biblioteca della Camera deriva da questa tradizione.

Opere citate

BEASTALL G., (1998) «The View From the Supply Side», *The Electronic Library*, v. 16, n. 1, Febr. 1998, pp. 31-32.

BELLINGER M., (1998) «The Transformation from Microfilm to Digital Storage and Access», *Journal of Library Administration*, v. 25, 1998, n. 4, pp. 177-185.

CARPENTER L. - SHAW S. - PRESCOTT A., (1998) *Towards the Digital Library: the British Library's Initiatives for Access Programme*, London, The British Library, 1998.

COLLIER M., (1997) *Toward a General Theory of the Digital Library* [<http://www.dl.ulis.ac.jp/ISDL97/proceedings/collier.html>].

CRAWFORD W. - GORMAN M., (1995) *Future Libraries: Dreams, Madness, and Reality*, Chicago, ALA, 1995.

CRAWFORD W., (1998) «The Danger of Digital Library», *The Electronic Library*, vol. 16, n.1, Febr. 1998, pp. 28-30.

GARRETT J. - WATERS D., (1996) *Preserving Digital Information*, Washington DC, Commission on Preservation and Access and Research Libraries Group, 1996.

GOLDHOR H., (1997) (issue editor) «Buildings, Books, and Bytes: Perspectives on the Benton Foundation Report on Libraries in the Digital Age», *Library Trends*, 46 (1), Summer 1997, pp. 1-228.

- GRIFFIN S.M., (1998) «Taking the Initiative for Digital Libraries», *The Electronic Library*, v. 16, n. 1, Febr. 1998, pp. 24-27.
- IDE N. - VÉRONIS J., (1995) (edited by) *Text Encoding Initiative: Background and Context*, Dordrecht ecc., Kluwer, 1995.
- KASDORF B., (1998) «SGML and PDF. Why We Need Both», *The journal of electronic publishing*, v. 3, n. 4, June 1998 [<http://www.press.umich.edu/jep/03-04/kasdorf.html>].
- KENNEY A.R. - CHAPMAN S., (1997) «An Introduction to Digital Imaging for Archives» in *Imaging Technologies for Archives*, Roma, Ministero per i Beni Culturali e Ambientali, 1997, pp. 125-154.
- LESK M., (1997) [1] *How much information is there in the world?* [<http://community.bellcore.com/lesk/ksg97/ksg.html>].
- LESK M., (1997) [2] *Practical digital libraries*, San Francisco, Morgan Kaufman, 1997.
- LISCIA R., (1998) (cur.) *Il mercato dell'editoria multimediale*, Milano, Guerini e Associati, 1998.
- MARCHIONINI G., (1998) *Research and development in digital libraries* [<http://www.glue.umd.edu/~march>].
- MUKAIYAMA HIROSHI, (1997) *Technical Aspect of Next Generation Digital Library Project* [<http://www.dl.ulis.ac.jp/ISDL97/proceedings/hiro/hiro.html>].
- NATION D.A. - PLAISANT C. - MARCHIONINI G. - KOMLODI A., (1997) *Visualizing Websites Using a Hierchical Table of Contents Browser: WEBToc* [<http://www.uswest.com/web-conference/proceedings/nation.html>].
- PEARCE C. - NICHOLAS C., (1996) «TELLTALE: Experiments in a Dynamic Hypertext Environment for Degraded and Multilingual Data», *Journal of the American Society for Information Science*, 47, 4, pp. 263-275.
- PLAISANT C. - MARCHIONINI G. - KOMLODI A., (1997) *Overviews and previews for the Library of Congress National Digital Library Program*, [TR based of LC report – August 7th, 1997 draft] [<http://www.cs.umd.edu/projects/hcil>].