

# DigitaMI: la biblioteca digitale di Milano

*Dalla collaborazione tra il Comune di Milano e la Telecom nasce un luogo di lettura virtuale*

Antonella Bassetti

Biblioteca Sormani  
antonella.bassetti@comune.milano.it

Alberto Salarelli

Dipartimento beni culturali e dello spettacolo  
Sezione beni librari, Università di Parma  
alberto.salarelli@unipr.it

Antonio Sciarappa

Istituto superiore "Mario Boella"  
Torino  
sciarappa@ismb.it

Il progetto DigitaMI nasce dalla collaborazione tra Telecom Italia (Progetto Italia) e Comune di Milano, con l'obiettivo di realizzare la prima biblioteca digitale di Milano. DigitaMI<sup>1</sup> mette a disposizione del pubblico tramite Internet documenti rappresentativi della tradizione storica e culturale della città. Il progetto, presentato al pubblico nel marzo dello scorso anno, offre ai suoi lettori una prima selezione di opere che documentano la vita del capoluogo ambrosiano di fine Ottocento, spaziando dai testi letterari di autori milanesi e lombardi a saggi sui fermenti artistici e culturali, dalle descrizioni di luoghi alle vicende storiche, senza ignorare le tradizioni popolari e la produzione dialettale.

Con DigitaMI si crea un luogo di lettura virtuale, dove la consultazione avviene a scaffale aperto anche per quei documenti rari e fragili, spesso di pregio, che a causa dello stato di conservazione o per ragioni di tutela non è possibile mettere – sempre e comunque – liberamente a disposizione del pubblico. Producendo dei surrogati digitali, la biblioteca realizza dunque il binomio conservazione/fruizione e traduce in realtà il duplice compito di diffondere e preservare il suo patrimonio.

## 1. L'infrastruttura tecnologica

La progettazione di DigitaMI si è

Fig. 1 – La homepage di DigitaMI



articolata in più ambiti di intervento e ha richiesto un lavoro di équipe per integrare le diverse competenze e professionalità richieste.<sup>2</sup> Le fasi principali hanno riguardato:

- la creazione di una infrastruttura di networking tra le diverse sedi operative;
- l'allestimento di un laboratorio di digitalizzazione presso la sede della Biblioteca Sormani e la definizione di un protocollo di acquisizione digitale;
- la definizione del set di metadati da adottare per il catalogo di DigitaMI, basato su Dublin Core, e la mappatura DC/UNIMARC ne-

cessaria per importare i metadati dall'OPAC dell'Indice SBN;

- lo studio e la realizzazione del database contenente tutte le informazioni utili all'identificazione e al recupero delle risorse tramite ricerca bibliografica (semplice e avanzata) e ricerca full-text;
- la progettazione e la realizzazione del software per generare, gestire e archiviare gli oggetti digitali e i metadati loro associati;
- l'ideazione e la realizzazione del motore di ricerca e dell'interfaccia web di accesso ai contenuti digitali e il loro inserimento all'interno del portale del Comune di Milano

nell'ambito di un piano di comunicazione coordinato.

Per quanto concerne l'infrastruttura del progetto, la Biblioteca Sormani ospita il laboratorio di digitalizzazione e l'applicativo gestionale per la lavorazione dei contenuti digitali che sono archiviati, in via provvisoria, in un server della rete interna. L'archiviazione finale e la pubblicazione su web avvengono presso i server del Data Center di Telecom Italia a Rozzano (Mi), collegato alla intranet del Comune di Milano con una connessione dedicata a 2Mbit/sec. La procedura di migrazione è programmabile utilizzando le funzioni del set di amministrazione dell'applicativo gestionale. Ad allineamento avvenuto, sul server locale resta comunque una copia del database.

La tecnologia software utilizzata è volutamente di tipo Open Source: Apache 1.3.27 (web server), Tomcat 5.0.16 (application server) e MySQL 4.0 (database).

## 2. L'acquisizione delle immagini

Il laboratorio di digitalizzazione consiste in un ambiente per la riproduzione digitale di documenti cartacei in formati e risoluzioni che consentano la loro consultazione tramite l'interfaccia web di DigitaMI e la loro archiviazione/conservazione sistematica su dvd. La soluzione di digitalizzazione adottata è basata su sistemi fotografici digitali con sensore a matrice; essi risultano più duttili, rapidi e in grado di affrontare agevolmente anche oggetti tridimensionali.<sup>3</sup> La soluzione comprende:

- uno stativo da riproduzione a colonna dotato di un sistema di illuminazione a luci fredde;
- una fotocamera digitale con obiettivi da riproduzione di 50 e 100mm, memoria di massa di 1

Giga e interfaccia fireware di accesso al computer;

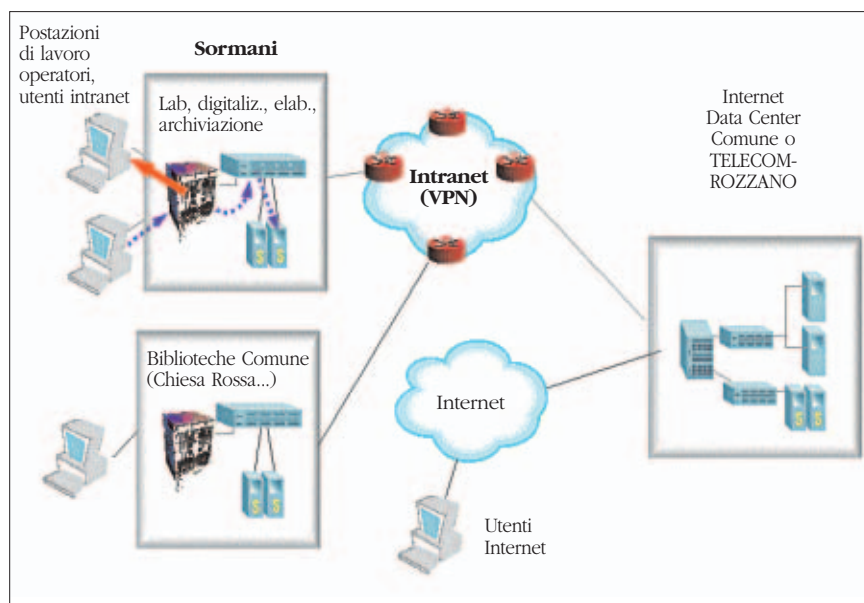
- due computer in rete, equipaggiati con 1 Giga di RAM, su cui sono installati specifici applicativi per l'acquisizione e l'elaborazione delle immagini.

Per quanto concerne la visualizzazione delle immagini delle pagine trasmesse via web, si è scelto di utilizzare il formato JPEG (Joint Photographic Experts Group).<sup>4</sup> Dette immagini sono ricavate operando sui file TIFF (Tagged Image File Format) ottenuti dalla scansione degli originali a una de-

qualitativamente migliori dei raw, anche se occupano più spazio.

Il formato JPEG utilizza una compressione di tipo "lossy", il che significa che per diminuire la dimensione dell'immagine alcuni pixel vengono eliminati, per poi essere ricostruiti (non esattamente uguali) in fase di decompressione e quindi l'immagine risultante non è più quella originale. JPEG tuttavia è stato progettato in modo tale che i pixel eliminati siano quelli meno percettibili all'occhio umano, utilizzando comunque 24 bit per pixel (cioè potendo quindi sce-

Fig. 2 – Schema dell'infrastruttura tecnologica



finizione non inferiore a 300dpi.<sup>5</sup> Si è scartata l'idea di utilizzare il formato raw, cioè del file che si ottiene direttamente dai CCD di una macchina digitale senza passare attraverso alcun tipo di processo di controllo e correzione da parte della macchina stessa.<sup>6</sup> Questi processi – successivi allo scatto – sono spesso fondamentali ai fini della qualità dell'immagine, in quanto lì avvengono, per esempio, i bilanciamenti dei bianchi, le accentuazioni dei contorni e dei contrasti eccetera. Per questo motivo le immagini post-processate (TIFF) sono

gliere tra oltre 16 milioni di colori). Il formato TIFF è un formato immagine di tipo raster sviluppato da Aldus, poi divenuto Adobe. Il TIFF è stato scelto come formato di archiviazione delle immagini sia sui server sia su dvd.

Le specifiche del formato TIFF permettono una notevole flessibilità:

il formato è largamente utilizzato per lo scambio di immagini raster fra stampanti e scanner perché permette di specificare numerose indicazioni aggiuntive come le tabelle di gamut o informazioni sulla calibratura del colore.<sup>7</sup>

### 3. Il trattamento degli oggetti digitali

DigitaMI ha definito le proprie linee operative nel rispetto degli standard biblioteconomici nazionali e internazionali, quali:

- Dublin Core per la descrizione bibliografica;<sup>8</sup>
- MAG per la descrizione dei Metadati amministrativi gestionali;<sup>9</sup>
- Z39.50/SRU e UNIMARC per l'interoperabilità con l'OPAC di riferimento.

Detti standard costituiscono il contesto di riferimento di MetaMAG, l'applicativo utilizzato per il trattamento degli oggetti digitali. Si tratta di un software specificamente progettato

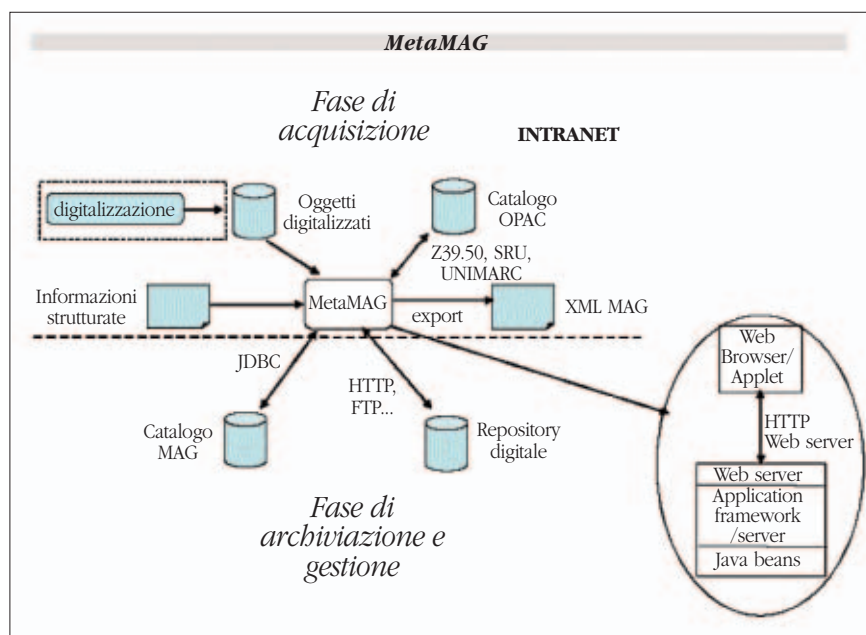
per la creazione e la gestione di una digital library, struttura "globalizzante" nella quale possono trovare collocazione, in forma digitalizzata, tutti i beni culturali (libri, carte d'archivio, musica, foto, filmati, musei virtuali, ricostruzioni tridimensionali ecc).<sup>10</sup>

MetaMAG è stato sviluppato da Inera per corrispondere alle funzionalità previste per DigitaMI; è

un prodotto web-oriented e server-oriented:<sup>11</sup> per l'utilizzo è sufficiente un browser che, collegandosi al server, accede alle funzioni disponibili, ovvero:

- comunicazione con l'OPAC dell'Indice SBN al fine di ricercare la notizia bibliografica corrispondente alla risorsa da trattare e importare i metadati utili alla sua catalogazione;<sup>12</sup>
- attribuzione dei metadati amministrativi gestionali;
- aggregazione di tutti i file relativi a un'opera per la creazione dell'oggetto digitale (Master Images; Alternative Images, OCR);
- preview degli oggetti digitali e delle relative informazioni per una verifica immediata del lavoro in corso;
- gestione dell'apparato editoriale che accompagna l'oggetto digitale;
- gestione dei diritti di accesso alle risorse digitali;<sup>13</sup>
- gestione dell'amministrazione di sistema;
- archiviazione degli oggetti digitali nel repository, aggiornamento dell'OPAC locale, memorizzazione ed esportazione dei dati in XML (Extensible Markup Language).

Fig. 3 - Configurazione del sistema MetaMAG



Quest'ultimo punto è particolarmente importante in quanto l'utilizzo di XML come linguaggio di marcatura da adottare nell'implementazione di una biblioteca digitale è fortemente auspicabile, perché tramite esso si rende possibile la creazione di una serie specifica di etichette che, per il fatto di essere conforme alle specifiche SGML, presentano comunque la caratteristica di essere comprese da tutti, infatti:

l'indipendenza dei dati [la separazione dei contenuti dalla loro rappresentazione] è la caratteristica essenziale del linguaggio XML. Poiché un documento XML descrive i dati, può essere facilmente elaborato da qualsiasi applicazione. L'assenza di istruzioni di formattazione rende un documento XML facile da analizzare dal punto di vista sintattico. Ne consegue che XML forma l'ambiente ideale per l'intercambio dei dati.<sup>14</sup>

Inoltre, a questo proposito, è importante sottolineare come MetaMAG sia pienamente compatibile con le specifiche OAI-PMH (definite dall'Open Archives Initiative).<sup>15</sup>

### 4. La consultazione

L'interfaccia di accesso alle opere digitalizzate, progettata da Argonet<sup>16</sup> unitamente al modello di fruizione dei contenuti e allo studio della comunicazione grafica, è stata pensata per risultare semplice e intuitiva; prevede l'utilizzo di un browser standard e non richiede l'installazione di plug-in aggiuntivi. La risoluzione minima supportata è di 800x600 pixel, lo standard di codifica seguito è il WAI - livello A.<sup>17</sup> L'impaginazione dei contenuti mediante i fogli di stile (CSS), evitando l'utilizzo di tabelle particolarmente ostiche per gli utenti di dispositivi screen reader, agevola la fruizione anche da parte di utenti disabili, secondo quanto pre-

Fig. 4 – La consultazione di un'opera in formato immagine



visto dalla legge Stanca sull'accessibilità.<sup>18</sup>

Nell'intento di rispondere ai diversi approcci e alle diverse esigenze dei visitatori, l'architettura informativa del sito offre varie modalità di accesso ai contenuti, dove i tradizionali strumenti di ricerca bi-

bliografica sono integrati nel catalogo con funzioni di browsing guidato e di ricerca testuale.

Il browsing consente di percorrere DigitaMI attraverso l'organizzazione logica dei documenti suddivisi secondo categorie e percorsi tematici, allo scopo di orientare nell'e-

Fig. 5 – Il risultato di una ricerca testuale



splorazione dei contenuti del sito secondo i criteri che informano i cataloghi per materie e per soggetti. Infatti, se le categorie articolano la collezione secondo grandi ambiti disciplinari (per es. letteratura, storia, arte), suddivisi fino a tre ordini di sottolivelli, i percorsi le attraversano per cogliere da esse solo i documenti significativi rispetto a un determinato soggetto. I percorsi rappresentano, per l'utente, un itinerario di scoperta all'interno del sito, con in più l'offerta di bibliografie di base su un determinato argomento; per la biblioteca, invece, si tratta dell'opportunità di valorizzare il proprio patrimonio anche in relazione a quei documenti meno noti, a volte dimenticati che, così contestualizzati e messi in relazione con altri, acquistano nuovo senso e rilevanza.

La ricerca testuale è di tipo full-text e consente di trovare parole o intere frasi all'interno di una sin-

gola opera, così come dell'intera collezione digitale. È resa possibile grazie alla presenza della versione elettronica dei testi pubblicati, ottenuta tramite OCR e offerta in formato TXT. Di conseguenza le opere presenti in DigitaMI, possono essere consultate – grazie a uno specifico set di strumenti – in versione grafica o testuale, sfogliando una pagina alla volta oppure posizionandosi in un punto preciso del volume.

La decisione di mettere a disposizione anche il testo elettronico è stata determinata dal desiderio di operare un'indicizzazione spinta dei documenti, in modo da offrire il massimo supporto possibile agli studiosi. Tutte le parole contenute nei volumi sono considerate keyword: il risultato che il motore restituisce è l'elenco delle opere e delle pagine che contengono il testo ricercato, e quest'ultimo viene evidenziato all'interno del suo

contesto, ovvero del brano ove è stato reperito, così da facilitare l'individuazione dell'occorrenza preferita (vedi figura 5).

Oltre al TXT pagina per pagina, per la maggior parte delle opere è possibile scaricare in locale il testo completo in formato RTF.<sup>19</sup>

Altre opzioni di accesso alle risorse sono costituite dalle due modalità di ricerca bibliografica, semplice e avanzata, e dalla possibilità di visualizzare l'elenco, ordinato alfabeticamente per titolo, di tutte le opere. Un sottoelenco, generato automaticamente dal sistema, consente di evidenziare gli ultimi dieci documenti pubblicati nel sito.

## **5. Lo stato dell'arte e gli sviluppi previsti**

Allo stato attuale DigitaMI propone circa settanta opere, tra monografie, spogli di periodici e brani scel-

ti da opere antologiche; gli scaffali di questa digital library si arricchiscono continuamente di nuovi documenti, secondo le indicazioni contenute nella carta delle collezioni digitali, disponibile sul sito.

Gli sviluppi futuri riguardano da una lato la manutenzione migliorativa del progetto, dall'altro la sua evoluzione in nuove direzioni.

Dal punto di vista infrastrutturale, si avverte l'esigenza di apportare alcuni interventi che prevedono, tra gli altri:

- una diversa disposizione dei link funzionali dell'interfaccia;
- l'introduzione di meccanismi di ricerca per liste;
- la pubblicazione dei file XML;
- l'attribuzione a ogni opera di un PURL per una maggiore integrazione con l'OPAC SBN;
- la possibilità di ottenere dei report statistici sugli accessi e sulle ricerche compiute;
- l'aggiornamento dei metadati amministrativi gestionali secondo l'ultima versione elaborata dall'ICCU.<sup>20</sup>

Per quanto riguarda i contenuti, la biblioteca ha due progetti in corso di studio. Il primo riguarda la digitalizzazione del fondo stendhaliano e, più in particolare, delle annotazioni autografe che lo scrittore ha lasciato sui testi della sua biblioteca, conservata presso la sezione Manoscritti della Sormani. La peculiarità del materiale richiede un'archiviazione e un ambiente di fruizione specifici, appositamente studiati per abilitare alcune funzionalità aggiuntive rispetto a quelle attualmente disponibili in DigitaMI e tuttavia con un'interfaccia coerente e integrabile. Il secondo progetto attiene alla digitalizzazione di parte dell'archivio audiovisivo della biblioteca. In collaborazione con il Dipartimento di scienze dell'informazione dell'Università degli studi di Milano è stato condotto uno studio di fattibilità per la conservazione in formato digitale di contenuti fonografici,

che si vorrebbero integrare in DigitaMI. In questo modo si punta a perseguire uno degli aspetti nodali del progetto: la possibilità, per l'utente, di recuperare contemporaneamente, da un unico punto di accesso, materiale di diversa natura (libri, riviste, filmati e contenuti musicali), cosa che non è possibile nella sede reale della biblioteca, dove il patrimonio è fisicamente distribuito in settori distinti a seconda della tipologia.

### Note

<sup>1</sup> <<http://www.digitami.it>>.

<sup>2</sup> Il gruppo di lavoro, coordinato da Antonio Sciarappa, insieme a esperti e tecnici di Telecom Italia, Argonet, Inera e dei Settori informatica e comunicazione istituzionale del Comune di Milano, comprende bibliotecari e tecnici della Biblioteca Sormani di Milano, in specifico: per il gruppo automazione Sergio Seghetti (responsabile), Antonella Bassetti, Maurizio Vinci; per la catalogazione e la metadattizzazione Sergio Graffi (responsabile), Patrizia Perotti e Filippo Salotti; per la conservazione e il laboratorio di digitalizzazione Bianca Girardi (responsabile), Mauna Gilardini, Nicola Nicodemi e Salvatore Matarrese. L'équipe si è inoltre avvalsa della collaborazione di: Anna Maria Tammaro e Alberto Salarelli dell'Università di Parma (consulenti per gli aspetti scientifico-biblioteconomici), Massimo Tosello dell'Accademia di belle arti di Urbino (consulente per l'allestimento del laboratorio di fotoproduzione e per la definizione degli standard di elaborazione digitale), Giovanna Rosa dell'Università degli studi di Milano (consulente scientifica per il percorso letterario).

<sup>3</sup> Cfr. *Manuale di buone pratiche per la digitalizzazione del patrimonio culturale*, versione 1.3 (3 marzo 2004), a cura del Gruppo di lavoro 6 del Progetto Minerva, p. 35, <[http://www.minervaeurope.org/structure/workgroup/goodpract/document/buonepratiche1\\_3.pdf](http://www.minervaeurope.org/structure/workgroup/goodpract/document/buonepratiche1_3.pdf)>.

<sup>4</sup> <<http://www.faqs.org/faqs/jpeg-faq/>>.

<sup>5</sup> Cfr. *CDL Digital object standard: meta-*

*data, content and encoding*, The University of California, 2001, p. 7, <[http://www.cdlib.org/news/pdf/CDL\\_ObjectStd-2001.pdf](http://www.cdlib.org/news/pdf/CDL_ObjectStd-2001.pdf)>; si veda anche *Manuale...*, cit., p. 38.

<sup>6</sup> Cfr. ALBERTO SALARELLI, *La fotografia digitale in biblioteca. Riflessioni su tecnica e descrizione delle immagini*, "Biblioteche oggi", 22 (2004), 8, p. 31-40: 33.

<sup>7</sup> <<http://it.wikipedia.org/wiki/TIFF>>.

<sup>8</sup> Cfr. *Manuale...*, cit., p. 46.

<sup>9</sup> Prodotto nell'ambito del Gruppo di studio sugli standard e sulle applicazioni di metadati nei beni culturali promosso dall'ICCU, <<http://www.iccu.sbn.it/strumeta/>>.

<sup>10</sup> Cfr. <<http://www.inera.it>>.

<sup>11</sup> Cfr. <<http://www.metamag.it/overview.html>>.

<sup>12</sup> DigitaMI ricerca nell'OPAC secondo i campi BID, Autore, Titolo e Soggetto. La derivazione automatica, resa possibile grazie al mapping DC/UNIMARC, compila la griglia di metadati definita dai bibliotecari, che ne completano i contenuti con le informazioni specifiche dell'esemplare digitalizzato.

<sup>13</sup> Sebbene le attività di digitalizzazione della biblioteca riguardino di preferenza documenti per i quali i diritti d'autore sono scaduti, l'infrastruttura di DigitaMI prevede un doppio livello di accesso alle risorse: pubblico (via Internet) per i documenti non coperti dal diritto d'autore; riservato (via intranet) delle biblioteche comunali, per quei documenti ancora sotto protezione del copyright.

<sup>14</sup> HARVEY M. DEITEL et al., *XML. Corso di programmazione*, Milano, Apogeo, 2002, p. 7.

<sup>15</sup> <<http://www.openarchives.org/>>.

<sup>16</sup> <<http://www.argonet.it>>.

<sup>17</sup> Cfr. *Web content accessibility guidelines 1.0*, W3C Recommendation 5 May 1999, <<http://www.w3.org/TR/WAI-WEBCONTENT/>>.

<sup>18</sup> Legge n. 4 del 9 gennaio 2004, *Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici*, <[http://www.innovazione.gov.it/ita/intervento/normativa/pubblicazioni/Acc\\_quad\\_2\\_04.pdf](http://www.innovazione.gov.it/ita/intervento/normativa/pubblicazioni/Acc_quad_2_04.pdf)>.

<sup>19</sup> Per la redazione dei testi elettronici ci si avvale della collaborazione del "Progetto Manuzio", promosso da Liber Liber, <<http://www.liberliber.it/>>.

<sup>20</sup> Versione 1.5 del 6 aprile 2004; cfr. <<http://www.iccu.sbn.it/schemag.htm>>.