

METAMAG

1 IL PRODOTTO

Metamag è un prodotto che permette l'acquisizione, l'importazione, l'analisi e la catalogazione di oggetti digitali per materiale documentale (quali immagini oppure file di testo (OCR)), e contestualmente consente la creazione ed il trattamento dei relativi metadati.

Con le risorse digitali infatti i metadati gestionali assumono una sempre più ampia rilevanza: la vita stessa della risorsa digitale dipende da questi metadati.

La prima istanza è riferita a risorse disponibili nell'ambito bibliografico, ma può estendersi al trattamento di quanto realizzato in archivi, musei, istituzioni culturali, ecc.

L'obiettivo è di costituire un valido strumento per:

- supportare le numerose campagne di digitalizzazione in atto
- garantire l'omogeneità descrittiva indispensabile all'integrazione delle informazioni sul materiale digitale nei servizi attualmente in essere
- favorire l'evoluzione dei sistemi della cultura verso la valorizzazione dei prodotti culturali.

2 IL CONTESTO

I riferimenti prioritari di MetaMAG sono gli standard descrittivi, nazionali e internazionali, di scambio dati e di servizio affermati nell'ambito dei beni culturali e di seguito elencati:

- MAG (per la descrizione standardizzata dei metadati MAG: Metadati Amministrativi Gestionali, in formato XML)¹
- Z39.50/SRU (per l'interoperabilità in fase di ricerca sul catalogo bibliografico)
- OAI: Open Archive Iniziative (per la raccolta dei metadati da parte di un livello centrale).

Altri standard che possono essere utilizzati nell'ambito di estensioni/personalizzazioni della soluzione sono:

- ODBC (per l'accesso alle Basi dati catalografiche non interoperabili)
- UNIMARC (per lo scambio di informazioni di catalogo ed eventuale alimentazione di un catalogo bibliografico locale).

La scelta degli standard consente

- una omogeneità descrittiva dei documenti digitali
- di ricostituire la visibilità dei cataloghi bibliografici di riferimento a prescindere dalle scelte architetture e di standard operate nei singoli sistemi.

Questa ultima caratteristica costituisce l'elemento caratterizzante del prodotto e ne determina la possibilità di distribuzione in vari contesti catalografici per favorire sia la derivazione automatica degli elementi descrittivi che la integrazione con i servizi di informazione bibliografica locale e centrale.

3 FUNZIONALITÀ PREVISTE

La gestione dei metadati è consentita attraverso una unica interfaccia amichevole che consente attualmente di:

¹ Prodotto nell'ambito del Gruppo di studio sugli standard e sulle applicazioni di metadati nei beni culturali promosso dall'ICCU, <http://www.bncf.firenze.sbn.it/progetti/mag/index.html>

1. Gestire ed attivare funzioni di amministrazione dell'applicazione e controllo accessi alle varie funzionalità previste
2. Configurare l'accesso al catalogo Bibliografico di riferimento e le modalità di cattura (derivazione) automatica dei dati da riportare sulla sezione BIB del DB Metadati e sviluppare funzionalità di interrogazione Client verso il DB Catalogo per l'individuazione della scheda di catalogo e la memorizzazione dei dati catturati: è possibile ricercare sul catalogo sia tramite bid che tramite altri punti di accesso come Autore, Titolo, Soggetto e selezionare la notizia da cui recuperare i parametri BIB
3. Attribuire gli elementi comuni del fondo oggetto di digitalizzazione per la gestione dell'anagrafe digitale (sezione GEN)
4. Gestire i Metadati tramite emissione di schermi per il completo trattamento dei metadati in formato MAG con pre-impostazione di dati relativi a set di oggetti digitali (ad es. immagini)
5. Attivare funzioni aggregate per la fruizione degli oggetti digitali (ad es. visualizzazione) e delle relative informazioni
6. Interrogare il Database dei Metadati con possibilità di richiamo alle funzioni di Gestione Metadati (modifica, cancellazione, inserimento)²
7. Esportare i file XML del Database Metadati in formato MAG-SCHEMA

4 CARATTERISTICHE DELLA SOLUZIONE

La soluzione è 'web-oriented' e 'server-oriented'. In questo modo l'utente non deve installare niente sul proprio PC ma utilizzare un Web Browser. Tramite un indirizzo (Intranet) accederà alle funzioni messe a disposizione del server tramite pagine HTML.

La tecnologia utilizzata è Java in termini sia di pagine JSP che di servlet Java Beans residenti sul server. La realizzazione in Java permette la sua portabilità in ambienti diversi come Windows, Macintosh, Linux, Unix.

Il software realizzato è basato su librerie disponibili con licenza Open Source.

Sul server risiede anche il database che consente la tenuta dei metadati MAG e di quanto è necessario per una corretta acquisizione e archiviazione dei dati digitali e metadati nell'ambiente custom.

La figura seguente indica l'architettura logica del prodotto

² I metadati di tipo descrittivo (cioè i Dublin Core identificati con BIB nella struttura MAG) possono non essere memorizzati nel database dei MAG perché comunque presenti nel catalogo bibliografico da cui sono derivati.

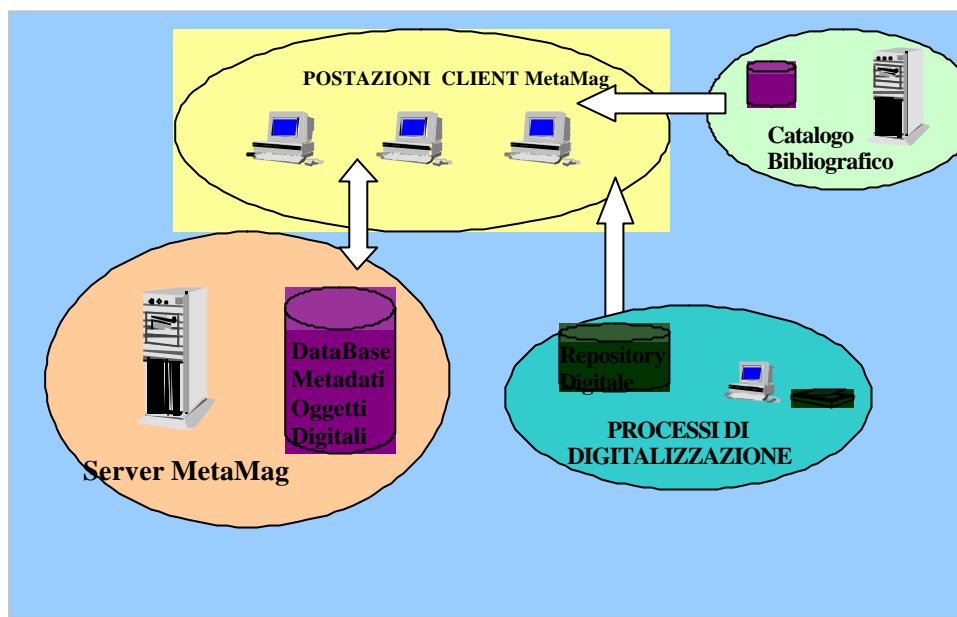


Figura A Architettura logica di riferimento del prodotto

L'architettura di riferimento è costituita da un ambiente tipicamente Intranet (ad es. una LAN) con:

- Postazioni Client su P.C. caratterizzate da:
 - Accesso agli oggetti digitali da trattare tramite Web Browser
 - Eventuale connessione tramite applet java ad un dispositivo di digitalizzazione (scanner, videocamera digitale, ecc.) con interfaccia conforme allo standard TWAIN per acquisire gli oggetti in digitale
- un Server di LAN contenente il web server (Apache Tomcat), l'application server (Tomcat), il software MetaMAG e il Database dei Metadati (RDBMS MySQL) con disponibile:
 - Connessione JDBC con il catalogo locale dei Metadati per la loro tenuta (Interrogazione, inserimento, variazione, Export XML)
 - Connessione al catalogo bibliografico tramite TCP/IP in modalità Z39.50 e/o SRU per la derivazione dei metadati BIB a fronte di ricerche attivate dal Client.

5 IL FLUSSO DI LAVORO

La configurazione prevista è indicata nello schema seguente; sono previste due fasi principali di elaborazione:

- *Acquisizione*, il cui output è la produzione di oggetti digitali e/o l'analisi tecnica nel caso di oggetti pre-esistenti, accompagnati da file XML
- *Archiviazione*, che prevede la memorizzazione dei dati XML su database, l'archiviazione degli oggetti digitali sul repository, l'aggiornamento dell'OPAC.

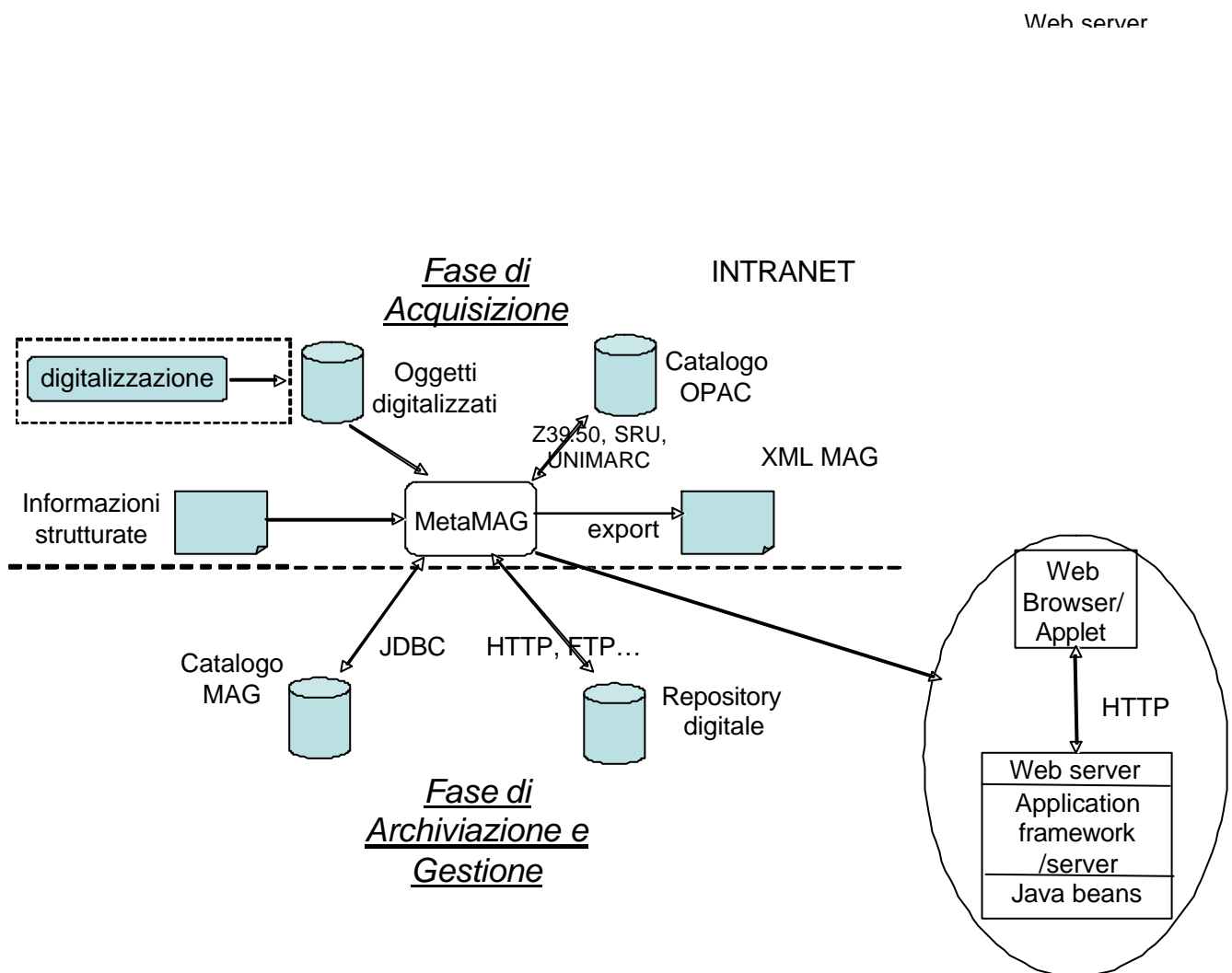


Figura B Configurazione del sistema MetaMAG

5.1 ACQUISIZIONE

La fase di acquisizione è composta da più sessioni di lavoro: nell'ambito di un singolo progetto di scansione (corrispondente ad una sezione GEN dei MAG) ogni sessione di lavoro è relativa ad una

opera diversa, cioè ad un nuovo BIB. Ogni sessione produce su una diversa cartella del server un file XML completo di tutti i componenti più le immagini associate.

Alla fine della fase di acquisizione sul server ci saranno tanti file XML quante sono state le sessioni, ognuno accompagnato da più oggetti digitali relativi.

5.1.1 Digitalizzazione

Il processo di digitalizzazione precede o può far parte della fase di acquisizione (nella figura precedente è messa in un riquadro tratteggiato).

5.1.1.1 Digitalizzazione nell'ambito dell'acquisizione

La figura seguente dettaglia il processo di digitalizzazione nel caso che esso faccia parte della fase di acquisizione, nel caso ad esempio di acquisizione di oggetto cartaceo.

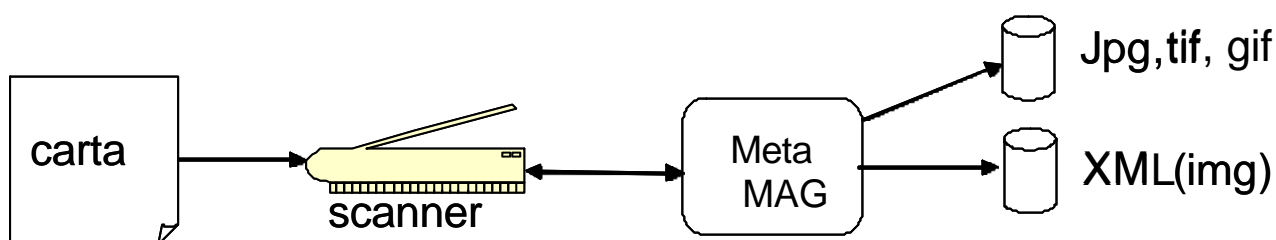


Figura C Il processo di digitalizzazione nell'ambito dell'acquisizione

5.1.1.2 Digitalizzazione precedente all'acquisizione

Se il processo di digitalizzazione invece precede (o ha preceduto in un periodo antecedente) la fase di acquisizione l'utente effettuerà un 'upload' dei file digitali verso il server che esaminerà il contenuto dei file estraendone le informazioni. Inoltre tramite l'accesso al catalogo e il completamento dei dati da parte dell'utente sarà disponibile il file XML MAG e la memorizzazione nella banca dati.

5.2 ARCHIVIAZIONE

La fase di archiviazione comprende le seguenti fasi:

1. caricamento dei file XML nel database o catalogo metadati MAG e relative funzioni di tenuta
2. upload degli oggetti digitali nel repository su LAN (in genere un sistema NAS, Network Attached Storage): il trasferimento dei file digitalizzati su NAS può essere fatto in modalità batch; altre funzioni relative sono: fruizione, spostamento, rinomino, cancellazione di oggetti digitali nel repository digitale
3. aggiornamento del catalogo (OPAC) con l'informazione che consente il recupero degli oggetti digitali associati alla notizia bibliografica (ad es. identificativo della risorsa digitale ovvero tag UNIMARC 956).

Solo la prima fase è attualmente coperta dal prodotto MetaMAG. Le altre anche per la loro estrema variabilità e configurabilità, verranno gestite come 'add-on' al prodotto valutabili singolarmente.

6 INFORMAZIONI ED INTERFACCE

Le informazioni gestite da MetaMAG sono sia relative alla struttura definita dall'XML MAG Schema che locali e di supporto al progetto, per quanto concerne l'acquisizione, l'archiviazione e la gestione, e su queste informazioni e la loro struttura è possibile effettuare delle 'customizzazioni', in relazione alle esigenze del cliente e/o del progetto. Le informazioni locali sono mantenute, insieme a quelle presenti nell'XML dello schema MAG, nella base dati di MetaMAG.

L'interfaccia Web integrata di MetaMAG consente contestualmente di effettuare:

- sessioni di lavoro per operazioni di scansione, esame dei file digitali, inserimento informazioni, creazione dei file XML e immissione nella banca dati
- configurazione ed amministrazione, da utilizzare sia per impostare i parametri che devono rimanere fissi nell'ambito delle sessioni di lavoro, che per inserire/modificare/cancellare i dati presenti sulla banca dati.

7 EVOLUZIONI

7.1 ESTENSIONI 'CUSTOM' E STANDARD AI MAG

In fase di configurazione sarà possibile aggiungere e gestire informazioni aggiuntive rispetto a quelle previste dallo standard MAG che si riveleranno via via importanti per le singole esigenze progettuali.

Inoltre è prevista la possibilità di integrare i metadati descrittivi con schemi standard come i TEI (Text Encoded Initiative) oppure di specializzare i DC con una struttura qualificata.

7.2 WORKFLOW DI CREAZIONE/GESTIONE DI BIBLIOTECHE DIGITALI

Funzione utilissima nell'ambito di progetti di una certa dimensione in cui più attori convergono in fasi diverse alla creazione e gestione della biblioteca digitale in termini di banca dati e risorse digitali.

Comprende anche la gestione dell'accesso degli utenti al sistema

7.3 FUNZIONI DI EDITING E OCR DELLE IMMAGINI ACQUISITE

Il prodotto MetaMAG permette di utilizzare al meglio le risorse che il driver TWAIN utilizzato mette a disposizione. Funzioni aggiuntive di editing e di ritocco immagine (ad es. modifiche della dimensione, dei DPI...) non sono previste nella versione base di MetaMAG. Esse possono essere fornite come 'add-on', tramite integrazione delle funzioni offerte da prodotti di foto ritocco di mercato come Adobe Photoshop o public domain come GIMP, utilizzando i loro 'Scripting languages'.

Similmente nel caso si voglia passare all'OCR di immagini acquisite ovvero di parti di esse: si provvederà ad integrare, lato client per consentire una maggiore libertà e controllo all'utente nell'impostare i parametri di riconoscimento testuale, librerie/prodotti di mercato (ad es. javaocr) che garantiscano validi risultati e completezza di funzioni.

Possibilità per gli utenti Internet di accedere al sistema MetaMAG (in ottica open source):

- sia come possibilità di ricerche di tipo full-text accoppiate a quelle catalografiche sui metadati descrittivi, sull'archivio di MetaMAG che assume quindi la funzione di un OPAC 'specializzato' nel digitale; chiaramente la ricerca full-text implica l'esistenza di documenti in formato testuale (ad es. derivati da elaborazione OCR), opportunamente indicizzati;
- che come possibilità di fruizione evoluta degli oggetti digitali (tramite l'utilizzo di applet che consentano all'utente di zoomare, stampare, salvare, ruotare, utilizzare effetto lente sulle immagini, oppure di sfogliare il libro 'virtuale' come se fosse un documento cartaceo).

7.4 RECUPERO DI PRECEDENTI DIGITALIZZAZIONI

Viene automatizzato il più possibile il processo di recupero di quanto già disponibile in formato digitale tramite processi batch asincroni, impostando lato client i parametri necessari e rendendo disponibili al server gli oggetti digitali da elaborare.

La scrittura di metadati nel db verrà comunque ove necessario posticipata alla visualizzazione e approvazione da parte dell'utente

L'immissione contestuale dei dati nel db dei metadati consente di modificare anche in corso d'opera i dati immessi durante i processi.

7.5 INTEGRAZIONE CON INDICI CENTRALI COME L'OPAC INDICE

Una naturale evoluzione di questa architettura prevede l'interoperabilità con strutture centrali che gestiscono indici (territoriali, tematici, ecc.) con finalità di ricerca e indirizzamento. L'obiettivo è:

- rendere accessibili e visibili oggetti presenti in realtà altrimenti confinate nell'ambito locale ovvero degli addetti ai lavori (ad es. quanto è fatto in biblioteche, archivi, musei),
- indirizzare i risultati delle ricerche dall'indice sugli archivi/cataloghi fonte ove oltre agli oggetti digitali si avranno informazioni aggiuntive, come ad es. i metadati MAG.

In ambito SBN si prenderà in considerazione l'Indice SBN e in particolare il catalogo centrale dei metadati MAG in sinergia con l'OPAC dell'Indice SBN.

L'interoperabilità viene garantita dall'utilizzo di alcune modalità di riversamento dei dati dalla periferia al centro.

La modalità standard è l'utilizzo del protocollo OAI PMH.

L'Indice centrale si configura come un service provider OAI: gli OAI data provider sono le singole biblioteche specificatamente nei cataloghi dei metadati MAG.

I metadati scambiati sono DC non-qualificati più eventuali aggiunte necessarie alla individuazione e fruizione degli oggetti digitali, derivati dai metadati MAG.

Altre modalità possibili di alimentazione dell'Indice sono ad es. import batch di tali metadati (via FTP).