

UNIVERSITÀ DEGLI STUDI DI MILANO
FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
CORSO DI LAUREA TRIENNALE IN
SCIENZE E TECNOLOGIE DELLA COMUNICAZIONE MUSICALE



Progettazione di Database Multimediali per Partiture Manoscritte

Relatore:
Prof. Goffredo HAUS

Correlatore:
Dott. Luca A. LUDOVICO

Correlatore:
Dott. Adriano BARATÈ

Tesi di Laurea di:
Cristina CASIRAGHI
Matr. n. 646644

ANNO ACCADEMICO 2004-2005

Indice

1	Archivi e Archivistica	3
1.1	Cos'è un archivio	5
1.2	Cos'è l'archivistica	6
1.2.1	La normativa archivistica	8
1.3	Archivi informatizzati	9
1.3.1	Vantaggi di un archivio informatizzato	11
2	Progetti	14
2.1	Progetto D.A.M.	14
2.1.1	Architettura	15
2.1.2	Recupero ed informatizzazione degli archivi del Teatro alla Scala	18
2.1.3	Interfaccia	20
2.1.4	Risultati ottenuti e conseguenze	23
2.2	Progetto ADMV	23
2.2.1	Architettura	25
2.2.2	Standard utilizzati	26
2.2.3	ADMV o ADM?	28
2.2.4	Criticità dell'ADMV	29
2.2.5	Interfaccia	30
2.2.6	Prospettive di sviluppo	31
3	Le basi di dati	33
3.1	Database multimediali	33
3.1.1	Architetture per database multimediali	34
3.1.2	Oggetti multimediali, feature e concetti	36
3.1.3	Differenze tra DBMS e MMDBMS	38
3.1.4	Applicazioni dei MMDBMS in ambito musicale	40
3.1.5	Database multimediali per teatri d'opera	41
3.1.6	Database multimediali per centri di conservazione e divulgazione musicali	46
3.1.7	Database multimediali per partiture manoscritte	48
3.1.8	Optical Music Recognition	54

3.2	Progettazione di un database in ambito musicale	58
3.2.1	Raccolta ed analisi dei requisiti	59
3.2.2	Progettazione concettuale	59
3.2.3	Progettazione logica	63
3.2.4	Progettazione fisica	68
4	Diffusione e fruizione	69
4.1	Metadati	69
4.1.1	L'utilità dei metadati	70
4.1.2	Le diverse categorie di metadati	72
4.1.3	Interoperabilità ed integrazione	74
4.2	Standard	74
4.2.1	Dublin Core	77
4.2.2	UNIMARC	80
4.2.3	MAG (Metadati Amministrativi-Gestionali)	83
4.3	La Rete della Musica Italiana (ReMI)	86
4.3.1	L'architettura di ReMI	87
4.4	Internet Culturale e Network Turistico Culturale	90
4.4.1	Lo sviluppo del progetto	92
4.4.2	Il portale Internet Culturale	93
4.4.3	Cultura, qualità e costi	96
5	Conclusioni e sviluppi futuri	100
	Ringraziamenti	104

Capitolo 1

Archivi e Archivistica

Ognuno di noi, in quanto portatore di una memoria che si vuole comunicare ad altri e trasmettere nel futuro, è potenzialmente il produttore di un archivio. Fino ad oggi, nella cultura occidentale, il supporto principale della memoria è stato il documento cartaceo, il quale ha costituito una sorta di estensione fisica delle possibilità di conservazione della memoria individuale e collettiva. La necessità di fermare sulla carta i propri ricordi, le proprie volontà, le relazioni tra gli individui o i fondamenti della vita politica e sociale, ha fatto nascere e crescere nel tempo una specie di “sedimento fisico” di memoria.

In tal modo nascono e si configurano gli archivi, depositi documentari caratterizzati dal fatto che ogni documento è strettamente legato all'altro da una rete di relazioni. Tale significativa caratteristica li rende profondamente diversi da una biblioteca, dove ogni singolo libro è invece un'entità conclusa in se stessa. Comprendere il linguaggio con cui il documento si esprime e le relazioni che lo legano agli altri è ciò che consente di comprendere l'archivio nel suo insieme e, quindi, anche di trovarvi ciò che si cerca.

Gli archivi hanno una storia: essi sono cresciuti nel tempo, sono stati conservati e trasmessi, talvolta sono stati distrutti. Essi non sono delle strane entità impersonali, amorphe e misteriose: ogni deposito documentario, ogni archivio, è infatti sempre espressione di qualcuno, che può essere un individuo o un'intera famiglia, un'istituzione, un ente, un'associazione.

Il produttore, generalmente, considera l'archivio una sua proprietà e lo organizza internamente per rispondere, in modo funzionale ed efficace, all'attività che egli svolge. Per questo motivo, l'archivio riflette non solo la storia del suo produttore, ma anche il modo in cui esso decide di organizzare e conservare il proprio patrimonio di memoria. Talvolta il produttore pretende che l'archivio rimanga segreto e che nessuno possa consultarlo, se non lui stes-

so. Non solo, esige, altresì, che non venga interamente trasmesso al futuro, preferendo scegliere attentamente ciò che va conservato e ciò che va distrutto.

Corrisponde poco alla realtà storica la tradizionale idea della lenta e progressiva sedimentazione degli archivi presso i loro produttori e del loro tranquillo e ordinato passaggio agli istituti di concentrazione che ne curano successivamente la consultazione. Nell'avvicinarsi degli enti e delle generazioni, gli archivi si sono trasmessi secondo procedure codificate (sostituzione di un ente all'altro, successione di un discendente legittimo) o seguendo percorsi non sempre prevedibili e lineari. Nella storia delle istituzioni, si riscontrano frequenti momenti di rottura o soluzioni di continuità che hanno avuto un riflesso sul destino dei rispettivi archivi.

Talvolta la volontà di alleggerire il carico di memoria ha condotto alla selezione o all'eliminazione di ciò che pare ormai inutile allo scopo per cui l'archivio è stato fatto; ma, in altre circostanze, c'è una precisa intenzione di non far sapere, di tramandare un'immagine "purgata" di sé o, comunque, gradita a chi si ritiene il proprietario della memoria. Esistono istituzioni pubbliche che nel tempo si sono arrogate il compito di conservare, selezionare e trasmettere i depositi documentari utili all'esercizio del potere. Motivo per cui, a quest'ultimo, si legano la storia della conservazione e della trasmissione documentaria, efficaci per la comprensione di ciò che viene tramandato e della ragione per cui viene archiviato.

La storia degli archivi conosce una novità fondamentale quando, soprattutto dalla fine del XVIII secolo, si afferma progressivamente la concezione che essi costituiscono un patrimonio pubblico e non una proprietà esclusiva.

Un'altra grande trasformazione interessa, contemporaneamente, anche il loro uso: quegli stessi documenti che nel passato erano serviti alla politica e all'amministrazione privata e pubblica divengono materiali indispensabili al lavoro degli storici. Gli archivi si trasformano, quindi, in "laboratori della storia". I grandi depositi archivistici, creati nei vari stati soprattutto tra la fine del Settecento e la prima metà dell'Ottocento, si qualificano sempre più come istituzioni culturali e centri di ricerca storica. Si fa strada, quindi, la consapevolezza che gli archivi sono fonti storiche che non hanno solo valore pratico, ma anche storico-culturale. Devono di conseguenza essere consultate liberamente dai cittadini e dai ricercatori per qualunque tipo di indagine non predefinibile (necessità dell'attività di ordinamento e descrizione). Con la deliberazione della Convenzione nazionale della Repubblica francese del 1794 si afferma che gli archivi sono liberamente consultabili; la loro piena accessibilità garantisce, dunque, anche la possibilità di scavare nel passato più remoto come in quello più recente. Si apre, in Francia soprattutto, una fase nuova anche per la crisi finale delle istituzioni di Ancien Regime: gli archivi

tesori che garantivano i privilegi feudali perdono il loro ruolo giuridico e la loro segretezza. Gli archivi sedimentati (delle cancellerie, delle amministrazioni che vengono abolite) si concentrano negli archivi generali e si aprono alla ricerca storica. Nascono organi statali con funzioni solo archivistiche. Si afferma la necessità di personale formato ad hoc e nascono le prime scuole d'archivio.

1.1 Cos'è un archivio

Secondo la definizione classica, data da Eugenio Casanova¹, un archivio è la raccolta ordinata degli “atti rappresentanti l'attività di un ente e di un'epoca, la storia di qualcuno dei grandi organismi costituiti dalla convivenza sociale”. Esso rivela l'impegno delle generazioni scomparse ed è mirato a garantire l'incolumità e la conservazione di quegli atti con sempre minore fatica. Funzione che lo caratterizza e che si concretizzerà soprattutto con l'introduzione di strumenti di archiviazione digitale, è il reperimento delle informazioni attraverso una modalità semplificata e più veloce.

Un archivio nasce grazie al progressivo accumulo di documenti inerenti un'attività. I suoi requisiti essenziali sono costituiti dall'organicità e dal legame che vincola tra loro i documenti conservati, attraverso criteri logici e necessari. Tale relazione permette di riconoscere, in maniera puntuale, la fonte che ha prodotto una determinata documentazione e di risalire, quindi, alle misure adottate per la crescita e la tenuta ordinata dell'archivio, secondo uno schema prestabilito che ne rispecchi e rispetti l'attività. L'archivio è anche il locale, o il deposito, nel quale i documenti sono custoditi e conservati. Del resto, lo stesso termine sembrerebbe derivare dalla parola greca *arkeion*, indicante il palazzo in cui l'arconte, il più alto magistrato della Grecia arcaica, conservava gli atti emanati. Esistono diversi metodi per procedere alla classificazione degli archivi. Considerando che, anche un archivio non più in continuazione conserva comunque una sua utilità come fonte irrinunciabile, per l'effettuazione di ricerche storico-documentarie, si parla di:

- *archivi chiusi o definiti*: per indicare quelli non più suscettibili di accrescimento, perchè appartenenti ad enti o ad amministrazioni non più esistenti;
- *archivi aperti o in formazione*, per fare, invece, riferimento a quelli suscettibili di ulteriori accrescimenti, perchè prodotti da enti o individui ancora operanti e viventi.

¹Soprintendente d'archivio, nato a Torino nel 1867, fondatore e direttore del periodico “Gli Archivi italiani”, insegnante universitario; segretario generale del Comitato Nazionale per la Storia del Risorgimento. Autore di varie pubblicazioni, nel libro dal titolo *Archivistica* (Siena, 1928)

Tali archivi, a loro volta, si distinguono in:

- *archivi correnti o registrazioni correnti*, i cui atti risultano frequentemente consultati nella misura in cui si riferiscono ad affari in corso di trattazione;
- *archivi di deposito o registrazioni di deposito*: i cui atti si riferiscono a pratiche sospese o esaurite che, pur non essendo consultate spesso, possono essere richieste per operare raffronti e ricerche;
- *archivi storici o archivi definitivi*, in cui si raccolgono tutti gli atti che, pur non avendo più alcun valore amministrativo o burocratico, conservano, però, un interesse documentario e vengono consultati soprattutto per motivi di studio.

Un altro criterio di distinzione si riferisce, invece, allo stato giuridico degli archivi e riconosce tra:

- *archivi pubblici*, i quali appartengono a o provengono da un organismo statale o un ente pubblico;
- *archivi privati*, corrispondenti a quelli prodotti da singoli individui famiglie o imprese private.

1.2 Cos'è l'archivistica

L'archivistica è la scienza che si occupa degli archivi studiandone l'origine, la formazione e la relativa regolamentazione giuridica. Essa non si occupa unicamente della loro tenuta, ma abbraccia anche l'ampio campo della costruzione e della manutenzione dei locali e delle suppellettili lì racchiuse, del loro ordinamento e della loro comunicazione nel presente e nel futuro.

In particolare, oggi, si parla di:

- *archivistica elementare*, quando oggetto di studio sono le nozioni di base di tale materia;
- *archivistica tecnica*, quando si trattano i vari metodi di classificazione, ordinamento, conservazione e riproduzione degli atti (dove per "atto" si intende il comportamento e la manifestazione di volontà che produce effetti giuridici, a differenza del documento che costituisce la rappresentazione e la testimonianza dell'atto);
- *archivistica superiore*, quando vengono studiati i metodi di ricerca storica su documenti di archivio e i nuovi ordinamenti legislativi.

La distinzione più classica, però, è quella operata da Eugenio Casanova, per il quale l'archivistica si può suddividere in tre branche, corrispondenti ad altrettanti settori di interesse:

- *archivistica pura o teorica*, che si occupa di tutto quanto concerne la formazione, l'ordinamento, l'inventariazione, la conservazione e lo studio dei documenti che costituiscono gli archivi;
- *archiveconomia o archivistica pratica*, che si interessa della organizzazione fisica degli archivi, quindi della sistemazione del materiale archivistico, della disposizione interna e dell'arredamento dei locali e di tutte le tecnologie archivistiche in genere;
- *legislazione archivistica*, che, oltre a disciplinare l'aspetto giuridico della conservazione e il controllo dei fondi archivistici, detta, in generale, tutte le norme che regolano la vita degli archivi.

Come disciplina scientifica, l'archivistica nasce ufficialmente in Europa nella seconda metà del Settecento, ma in Italia le sue radici sono molto più profonde. Già i Romani erano sensibilmente interessati alle questioni concernenti l'autenticità e la consultabilità dei documenti; in epoca medievale vennero, poi, elaborati i primi principi sull'organizzazione e l'ordinamento dei documenti di archivio; nel Seicento, infine, comparvero i primi trattati dedicati a tale argomento.

Nello specifico, parlando di documento di archivio, ci si riferisce a qualsiasi mezzo, o cosa, che consente di tramandare la memoria di un fatto, provandone l'esattezza e le modalità. Esso è testimonianza di un dato ambiente o periodo o civiltà alle quali appartiene. Può essere usato come strumento di studio, di consultazione, di indagine o come sussidio di determinate ricerche. Se un tempo poteva trattarsi di una scrittura o di un disegno, ora rientrano a pieno titolo tra i documenti di archivio anche le fotografie, le partiture, le registrazioni sonore, nonché dati elaborati e diffusi per mezzo di apparecchiature informatiche.

A seconda della forma, un documento conservato in archivio può essere in:

- *originale*, se si tratta dell'esemplare compiuto del documento, praticamente perfetto nelle sue peculiarità formali e contenutistiche;
- *minuta*, intesa come esemplare di un documento originale spedito, il quale resta nell'archivio del mittente;
- *copia*, riferendosi alla riproduzione di un documento originale o eseguita a mano, a macchina o mediante qualsiasi apparecchio abilitato alla riproduzione.

1.2.1 La normativa archivistica

Mediante il decreto legislativo 20 ottobre 1998 n. 368, è stato istituito il nuovo Ministero per i Beni e le Attività Culturali: l'amministrazione archivistica italiana rientra, allo stato attuale, nelle competenze di questo dicastero. Tale Ministero ha ereditato funzioni e competenze che, dopo aver riguardato i dicasteri della Pubblica Istruzione e dell'Interno, erano poi confluite, a partire dal 1975, nel Ministero dei Beni Culturali e Ambientali. Rispetto a quest'ultimo, esso si distingue in modo particolare per la denominazione, visto che, alla scomparsa del termine "ambientale", fa riscontro l'aggiunta delle "attività culturali". Secondo l'art. 1, del precedentemente citato decreto legislativo, tale Ministero provvede alla tutela, alla gestione e alla valorizzazione dei beni culturali e ambientali. Inoltre, si occupa di promuovere e sostenere le attività culturali; gli spettacoli (le attività teatrali, circensi, musicali, cinematografiche, di danza, dello spettacolo viaggiante); i libri e lo sviluppo dei servizi bibliografici e bibliotecari nazionali; la cultura urbanistica e architettonica, partecipando alla progettazione di opere destinate all'attività culturale.

Nell'esercizio di tali funzioni, il Ministero privilegia il metodo della programmazione; favorisce la cooperazione con le Regioni e gli enti locali, con le amministrazioni pubbliche, con i privati e con le organizzazioni di volontariato. Opera per la massima fruizione dei beni culturali e ambientali, per la più ampia promozione delle attività culturali, garantendone il pluralismo e l'equilibrato sviluppo in relazione alle diverse aree territoriali e ai diversi settori.

L'articolo 6 del Decr. Legisl. 368/1998 ha istituito presso il Ministero l'istituto centrale per gli archivi cui sono stati attribuiti, in linea generale, compiti di "definizione delle modalità tecniche per l'inventariazione e la formazione degli archivi, di ricerca, di studio, di applicazione di nuove tecnologie" in attesa di disciplinarne l'organizzazione e le funzioni con apposito regolamento. Esso si è venuto ad aggiungere ad altri quattro istituti centrali preesistenti:

- *l'istituto centrale per il catalogo e la documentazione*, con compiti in materia di catalogazione e documentazione dei beni culturali di interesse archeologico, storico-artistico e ambientale;
- *l'istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche*, con compiti in materia di catalogazione e documentazione del patrimonio librario conservato nelle biblioteche pubbliche;
- *l'istituto centrale per la patologia del libro*, con compiti in materia di restauro di materiale bibliografico;

- *l'istituto centrale per il restauro*, con compiti di ricerca scientifica finalizzata agli interventi di preservazione, tutela e restauro dei beni culturali di interesse archeologico e storico-artistico.

Oltre agli istituti centrali appena elencati esistono istituti a loro volta dotati di particolari finalità, come la Discoteca di Stato, la quale conserva la più grande collezione pubblica italiana di documentazione sonora e garantisce la conservazione e fruizione del patrimonio sonoro nazionale, sia edito che inedito.

1.3 Archivi informatizzati

Ai fini della riproduzione sostitutiva delle serie archivistiche, le quali devono essere conservate per un periodo di tempo lungo o addirittura illimitato, esiste la possibilità di ricorrere a svariati supporti tra i quali:

- memorie magnetiche (nastri e dischi, questi ultimi digitali e arrays);
- memorie ottiche (cd-rom, dischi worm, laser disc);
- dischi magneto-ottici.

Poichè le riproduzioni risultano puntualmente complicate e spesso comportano spese elevate, è auspicabile che vengano effettuate solo quando arrecano un contributo realmente positivo alla gestione dell'archivio. L'utilizzo di supporti automatizzati da' garanzie in termini di durata e vulnerabilità, del numero di informazioni acquisite e delle modalità per procedere alla loro interpretazione e dei tempi occorrenti per accedere ai testi riprodotti.

L'informatizzazione degli archivi ha avuto come conseguenza una vera e propria rivoluzione delle mansioni dell'archivista, trasformando radicalmente i sistemi di gestione degli archivi e permettendo una maggiore efficienza e rapidità nel lavoro. Lo sviluppo dell'informatica ha consentito la creazione delle cosiddette "basi informative", offrendo la possibilità di organizzare i dati mediante appositi strumenti informatici.

L'impiego di tali mezzi nelle fonti documentarie può essere applicato agli archivi esistenti oppure a quelli nati direttamente su supporto informatico.

Vi sono sostanzialmente tre attività principali che contraddistinguono la gestione di un archivio computerizzato.

1. *Immissione dei dati:*

Si propone come momento successivo alle fasi preliminari di ricevimento, registrazione e smistamento dei supporti cartacei. Quest'attività

consiste nel riportare gli estremi e gli elementi principali dei dati nella memoria del computer, ovvero su un supporto che ne permetta la registrazione permanente.

2. *Reperimento dei dati (information retrieval):*

Si tratta della fase di lavoro che consente di effettuare sia la ricerca di dati introdotti attraverso una preventiva codificazione delle informazioni, sia il collegamento tra diversi archivi.

3. *Modifica dei dati:*

Consiste nel far sì che la base informativa possa essere successivamente modificata e aggiornata con apposite transazioni da parte degli archivisti e dei ricercatori.

Tutto il lavoro di organizzazione del database è già effettuato in fase di programmazione, per cui l'attività dell'archivista è fondamentalmente quella di inserire i dati, tramite schermate guidate e un'opportuna interfaccia grafica, e curare tutti i vari aspetti riguardanti la gestione della base informativa quindi, nello specifico, l'utilizzazione del database. Con l'ausilio di appositi menù di scelta, egli potrà ricercare e rintracciare le informazioni richieste.

Le nuove tecnologie informatiche e multimediali hanno prodotto una grande rivoluzione nel campo della conservazione e della fruizione del patrimonio archivistico italiano. Occorre però precisare che l'applicazione dell'informatica alle fonti documentarie presenta non poche difficoltà, a causa delle peculiari caratteristiche delle fonti stesse.

Il problema più rilevante in tema di informatizzazione di archivi storici consiste nella difficoltà oggettiva di standardizzare la descrizione archivistica, allo scopo di favorire l'elaborazione di descrizioni coerenti e appropriate e rendere possibile l'integrazione di quelle provenienti da istituti archivistici differenti, in un sistema informativo unificato. L'uso e la gestione del database necessitano come premessa uno specifico lavoro di progettazione dell'architettura del sistema, con implicazioni sulla scelta dei componenti hardware, in relazione alle funzioni da svolgere e privilegiando, ad esempio, la gestione di grandi quantità di dati rispetto alla velocità di elaborazione degli stessi.

Contemporaneamente, però, bisogna poter disporre anche di un software adeguato per ottimizzare la gestione dei dati in rapporto alle stesse componenti hardware.

In un sistema destinato alla gestione di grandi archivi e in cui le componenti hardware e software sono appositamente progettate per svolgere queste funzioni, il programma per l'amministrazione del database è parte integrante del software di sistema, insieme al sistema operativo. Per esigenze più ridotte, è possibile utilizzare programmi applicativi per la gestione di basi informative chiamati DBMS (Database Management System).

1.3.1 Vantaggi di un archivio informatizzato

Gli archivi digitali sono ottimi strumenti per la conservazione ed il riordinamento dei dati. Uno dei motivi per cui i processi di archiviazione sembrano diventare sempre più di importanza vitale deriva dal fatto che si ha a che fare, ogni giorno, con una grande mole di informazioni. La radio, la televisione e Internet investono gli individui, di continuo, con un flusso inarrestabile di dati in cui spesso è molto difficile orientarsi.

In questa intricata foresta dell'informazione si è tentati di seguire ora una direzione ed ora un'altra, rischiando di perdere l'obiettivo della propria ricerca. In tutte le epoche, anche i filosofi hanno affrontato il problema di trovare un metodo per mettere ordine nelle proprie cognizioni.

Nel 1600, ad esempio, il filosofo francese Cartesio si chiedeva quale fosse il metodo per fare ordine nelle conoscenze e rispose a tale quesito con una metafora. Chi è smarrito in una foresta "non deve aggirarsi ora di qua ora di là, e tanto meno fermarsi, ma camminare sempre nella stessa direzione perchè, così, anche se non va proprio dove desidera - scriveva Cartesio - arriverà alla fine in qualche luogo, dove probabilmente si troverà meglio che nel fitto della boscaglia".

Il ragionamento cartesiano esemplifica il procedimento che da secoli è alla base della classificazione: per non perdersi nell'intricata foresta dell'informazione, i dati vanno ordinati secondo un unico filo conduttore. Gli archivi sono un valido aiuto nella ricerca di questo punto di forza che ne favorisce la consultazione.

Attualmente, qualsiasi prodotto mediatico, dai libri, ai filmati amatoriali, alla musica, ai cartoni animati, può essere trasformato in un documento digitale e tradotto in un flusso di bit.

Grazie all'informatica, si dispone di strumenti di ordinamento e gestione dei dati sempre più potenti. Ed è proprio attraverso questi mezzi, gli archivi digitali, che il flusso dinamico e caotico dell'informazione può essere fermato e trasformato in un insieme ordinato. Rapidità di accesso, facilità nella consultazione e vasta documentazione disponibile in uno spazio

ristretto sono quindi i vantaggi più cospicui dell'archiviazione digitale su quella tradizionale. A tutto ciò va aggiunto, naturalmente, l'uso multimediale, in senso tecnico, che questo tipo di catalogazione offre: la possibilità di rendere utilizzabile, dentro un unico supporto, l'intera attività di un mezzo di comunicazione, dalle sue origini ad oggi. Lo svantaggio consiste nel dover cercare le informazioni da soli e ciò implica dei saperi, delle competenze che devono essere acquisite a livello individuale e che prima erano unicamente affidate alla figura dell'archivista.

Un esempio di archivio digitale può essere riscontrabile nella realizzazione delle Teche Rai. Dopo quasi 50 anni di attività nell'etere, la Rai ha messo a punto un complesso sistema di archiviazione digitale del materiale radiofonico e televisivo prodotto. Le audiovideoteche della Rai, per la prima volta, affrontano la conservazione, la catalogazione e la documentazione dei programmi realizzati, sia quelli già in archivio, sia quelli che vengono archiviati giorno per giorno.

E' stato creato anche un sito internet: www.teche.rai.it su cui è possibile trovare in tempo reale documenti, foto, immagini e registrazioni radio che hanno fatto la storia televisiva italiana. Tra gli obiettivi delle teche digitali spicca quello di favorire una produzione televisiva sempre più veloce nei tempi, alimentata da prodotti di archivio trattati con precisione e tempestività. I cambiamenti diventano cruciali nella produzione dei programmi radiofonici: il computer che gestisce la regia è collegato con l'archivio centrale da cui attingerà tutte le immagini necessarie. La modifica interesserà, dunque, anche la confezione dei programmi stessi e il montaggio. L'archivio digitalizzato, che potrà essere utilizzato anche dal pubblico, sarà particolarmente utile per le ricerche.

L'archiviazione risponde al bisogno di conservare il materiale documentario in modo razionale e uniforme per renderlo recuperabile nella ricerca. Sostanzialmente, un'archiviazione dei dati e dei documenti condotta in maniera sistematica riserva notevoli vantaggi quali:

- accesso immediato ai documenti da parte del personale autorizzato o di utenti interessati e appassionati, tramite rete;
- aggiornamento costante dell'archivio digitale on-line;
- riduzione dei costi di ricerca dei documenti;
- sicurezza nella conservazione dei documenti;
- riferimento temporale specifico del documento informatico: ossia la data e l'ora associata al documento e sottoscritti elettronicamente, al

fine di garantirne l'autenticità e l'integrità. E' possibile quindi mantenere un certo grado di informazioni riferite allo stato dell'arte in un dato periodo;

- memorizzazione fatta in modo tale da garantire la leggibilità e la consultabilità nel tempo, anche grazie al tipo di supporto tramite il quale i processi di conservazione e salvataggio vengono portati a compimento;
- eliminazione di archivi cartacei e realizzazione di flussi documentali, i quali migliorano l'efficienza di amministrazioni e, comunque, la gestione di patrimoni culturali;
- risparmio di costi dovuto, non solo agli enormi spazi che si libereranno, ma anche alla possibilità di una gestione decentralizzata degli archivi;
- ricerca di un documento, in tempi brevi, tra un gran numero di esemplari, inserendo degli "indici" i quali permettono di ottenere in pochi attimi tutta la corrispondenza relativa e disponibile;
- impossibilità di accesso a dati non autorizzati. Ciascun utente ha un proprio profilo, con una serie di abilitazioni e poteri: documenti che può vedere, spedire, limiti di tempo, livello di riservatezza.

Tali potenzialità, inoltre, sono rafforzate dal fatto che la legge non impone l'utilizzo di un supporto informatico specifico, il quale potrà essere scelto liberamente, purchè permetta il rispetto dei requisiti. Per archiviare in modo semplice è fondamentale iniziare con la gestione dei documenti, la quale viene effettuata determinando:

- la tipologia (per es. articolo di giornale, regolamenti, partiture, libretti, foto...);
- la classificazione (per indici, numeri, lettere, ecc.);
- la costruzione del database per ordinare e indicizzare le informazioni.

Si può immaginare una banca dati come una grande tabella nella quale ogni riga rappresenta una serie di informazioni attinenti le une con le altre. All'interno di ogni riga sono disponibili diversi campi, ognuno destinato a un tipo di informazione differente, per esempio, nome della testata, data, fonte normativa, tipologia, argomento. L'uso delle banche dati è indicato per l'archiviazione di grosse quantità di informazioni in quanto consente l'indicizzazione dei campi e il recupero rapido delle informazioni attraverso maschere di ricerca.

Capitolo 2

Progetti

Nel capitolo precedente sono stati forniti inquadramenti storici e definizioni relativi agli archivi e alla disciplina scientifica che se ne occupa: l'archivistica.

Esempi significativi di progetti avviati e giunti ad un certo livello di completamento, molto recenti ed attuali, sono il Progetto DAM (Digital Asset Management) del Teatro alla Scala di Milano e l'Archivio Digitale della Musica Veneta (ADMV).

Essi sono tra loro molto simili nelle finalità, ma differenti come implementazione. Compito di questo capitolo sarà darne una descrizione.

2.1 Progetto D.A.M.

La Fondazione Milano per La Scala è di supporto alla grande professionalità scaligera e si occupa di sostenere progetti speciali di particolare interesse per la tutela del patrimonio artistico e culturale del Teatro come il Progetto D.A.M., Digital Asset Management, ossia gestione delle risorse digitali, che nel caso specifico del Teatro Alla Scala, assume anche il significato di Depositi Archivi Magazzini.

Esso è nato e si è sviluppato nel Giugno del 1999 per perseguire un preciso scopo: digitalizzare e rendere fruibile, mediante moderne tecnologie informatiche, il materiale presente negli archivi del Teatro, incluse fotografie, costumi, attrezzeria, bozzetti, figurini e audio, anche in un'ottica di commercio elettronico. Oggi, grazie al DAM, è possibile accedere in tempo reale al patrimonio artistico scaligero per sfruttare al meglio le conoscenze e l'esperienza del passato anche nella realizzazione di nuove produzioni.

Gli obiettivi che il Progetto DAM in principio si è posto sono:

- la creazione di un'applicazione Intranet chiamata *LaScalaDAM*, capace di rendere fruibile e facilmente accessibile ogni archivio DAM da qualsiasi postazione interna autorizzata, non solo a Responsabili degli archivi e dei magazzini, ma anche alla Sovrintendenza, alle Direzioni ed a tutti i Reparti del Teatro;
- la garanzia di poter consentire l'accesso dal Teatro e dalla tournèe a tutte le informazioni relative all'attività artistica del Teatro alla Scala dal 1950 ad oggi: dati di locandina di opera, balletti, concerti, recital e altre manifestazioni, oltre che a tutti i file multimediali relativi al materiale di scena presente negli archivi scaligeri;
- la fruizione di tali dati anche al pubblico esterno, tramite postazioni DAM all'interno del Sistema Scala e una sezione "archivio" specifica sul sito Internet del Teatro.

Dal 2000 in poi, con i primi archivi digitali consegnati, fino ai giorni nostri, il Teatro alla Scala ha così potuto attingere sempre più dai contenuti multimediali inseriti e catalogati al fine di soddisfare un vasto fronte di esigenze, fra cui, le principale sono:

- ricerca di asset archiviati/conservati per riutilizzo su nuovi allestimenti: la rapidità della ricerca e la completezza delle informazioni fornite hanno permesso un miglioramento delle percentuali di riutilizzo e snellito di molto il processo già in essere;
- selezione e spedizione di contenuti relativi agli spettacoli della stagione in corso e/o di quelle precedenti ad editori;
- alimentazione snella, supportata da un workflow di autorizzazione e di pubblicazione per alimentare intere aree del sito web istituzionale (la Stagione, l'Archivio Web).

Nel 1999, ha preso forma anche il progetto *LaScalaWeb*: il Teatro ha chiesto pieno supporto per la concezione, il design e l'implementazione del sito web il quale permette l'accesso on-line al ricco archivio multimediale della Scala.

2.1.1 Architettura

Nel DAM confluiscono i dati, provenienti da altri sistemi informatici, i quali vengono utilizzati da altri sistemi, tra cui il sito Web, le applicazioni periferiche (ossia quelle specifiche per Attrezzeria, Archivio fonico, Archivio fotografico,) e, infine, l'applicazione DAM centrale.

Quest'ultima utilizza il DAM solo in consultazione: l'output del DAM costituisce l'input per l'applicazione DAM centrale; il flusso inverso, ossia l'immissione di dati nel DAM tramite l'applicazione in oggetto, non è consentito.

L'applicazione interroga in maniera integrata ed evoluta il database del DAM, estraendo dati ed informazioni sulla base delle richieste (query) dell'utente. In particolare, l'interfaccia grafica nasconde tutti i dettagli implementativi e svincola l'utente dalla necessità di conoscere un linguaggio di interrogazione, quale l'SQL, per reperire informazioni dal database. Le possibilità di accesso ai dati sono concordate con la direzione del Teatro e determinano profili personalizzati per alcune macro-categorie di utenti. Il concetto, espresso in termini semplici, è il seguente: non tutti gli utenti hanno accesso a tutto il materiale. Quanto è pubblico e pubblicato su Internet è accessibile a tutti: ad esempio, i dati di locandina e i file PDF tratti dai programmi di sala.

I responsabili dei singoli archivi hanno la piena visibilità su quanto è di loro competenza. Ad esempio, i responsabili dell'archivio fotografico hanno accesso agli ingrandimenti delle foto e alla scheda che ne descrive l'identificazione, le caratteristiche e la collocazione fisica. Nel mezzo tra la totale visibilità e la possibilità di consultare solo quanto di pubblico dominio esistono numerose sfumature e il meccanismo può cambiare a seconda delle esigenze e delle esperienze maturate con l'uso dello strumento software.

La protezione, sia da parte del database che dell'applicazione, dovrebbe prevenire la visualizzazione non autorizzata di dati sensibili. In particolare, alcuni espedienti grafici rendono chiari all'utente i sentieri percorribili e quelli non percorribili. Il più delle volte, si ricorre alla colorazione grigia dei link disabilitati e alla disattivazione del collegamento ipertestuale.

Un'altra funzione fondamentale del software consiste nell'aggregazione di dati eterogenei. Se, ad esempio, l'utente è interessato a un titolo d'opera legato a determinati esecutori, l'applicazione estrae l'elenco degli allestimenti e delle serate ad esso riferiti, mostrando anche tutti gli oggetti multimediali relativi (tra cui fotografie di scena, bozzetti, figurini, programma di sala, eventuali registrazioni e via dicendo).

L'applicazione DAM centrale è un software concepito per l'uso interno al Teatro. Essa utilizza le infrastrutture di rete poste tra le sedi del Teatro e all'interno di ciascuna sede. L'intervallo di indirizzi IP attivati è limitato alla rete privata Scala, quindi l'applicazione non può essere utilizzata da una macchina esterna.

La struttura della rete ha incidenza sulle prestazioni globali dell'applicazio-

ne. Il server è fisicamente collocato nella webfarm di Fastweb, attualmente connessa alla sede del Piermarini da una linea a 10 Mbit/s. Le altre sedi (Arcimboldi, Ansaldo, uffici del CNR, ecc.) sono poi collegate a tale nodo centrale con linee da 2 Mbit/s, secondo una topologia a stella. Per questa ragione, da una sede periferica, la trasmissione delle pagine e dei dati e la navigazione risulteranno più lenti che da una postazione collegata al nodo centrale. Contrariamente, i tempi di risoluzione delle query non dipendono dal luogo di consultazione dato che il calcolo della risposta alle interrogazioni viene effettuato sul server nella sede Fastweb.

Nel valutare i tempi di attesa, dunque le prestazioni, si deve tener conto del carico di lavoro cui sono sottoposti il server e la rete. Tali misure non sono calcolabili a priori e dipendono dal tipo di attività in corso e dall'orario di utilizzo del servizio. Ad esempio, il fatto che un solo utente o numerosi utenti stiano effettuando interrogazioni incide anche notevolmente sui tempi di risposta del server; analogamente, la percentuale di banda passante libera tra i vari snodi della struttura di rete consente un download più o meno rapido delle pagine e dei contenuti multimediali.

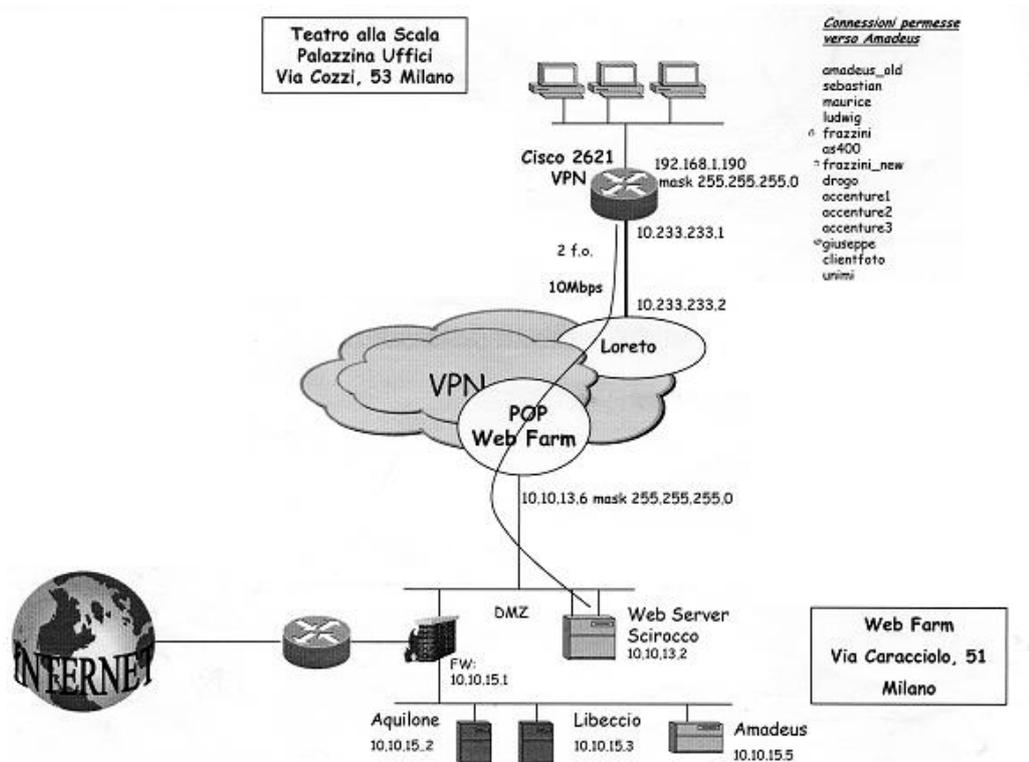


Figura 2.1: Architettura del progetto DAM

La visualizzazione e parte dei controlli di navigazione sono delegati al browser. A questo proposito, l'applicazione è stata ottimizzata per Internet Explorer 5.5 (e versioni successive) a una risoluzione di 800x600 pixel o superiore. Per apprezzare a pieno i contenuti fotografici digitalizzati, si consiglia una profondità di colore non inferiore ai 16 bit, pari a 65536 colori.

Tale dotazione minimale è presente su tutte le postazioni informatiche del Teatro e, quindi, non costituisce una limitazione all'usabilità del prodotto.

2.1.2 Recupero ed informatizzazione degli archivi del Teatro alla Scala

Il Laboratorio di Informatica Musicale (LIM) del DSI/DICo dell'Università degli Studi di Milano è impegnato dal 1997 in un progetto di recupero del prezioso archivio fonico del Teatro alla Scala, un patrimonio di grandissimo interesse musicale, musicologico e storico, oggi purtroppo a rischio di rapido deterioramento. L'archivio comprende registrazioni che vanno dal 1951 ad oggi, con "copertura" quasi completa del cartellone dal 1974 ai giorni nostri, per un totale di circa 5.000 nastri (open-reel analogici dal 1951 al 1990 e DAT dal 1991) ognuno dei quali contenente registrazioni dei più famosi musicisti di questo secolo (cantanti, direttori, esecutori strumentali, orchestre, cori). Lo stato non ottimale di conservazione dei nastri e la preoccupazione di garantirne la salvaguardia hanno dato il via all'operazione di salvataggio di questo prezioso patrimonio culturale.

L'operazione di recupero è svolta mediante la realizzazione coordinata di innumerevoli attività, sinteticamente enumerabili come segue: pulizia dei nastri analogici originali; trattamento termico per i nastri che si disgregano per l'età, la cattiva conservazione o la scarsa qualità dei materiali originali; digitalizzazione dell'informazione fonica analogica per la conservazione e il trattamento numerico; masterizzazione di due copie di CD-R (Compact Disc Recordable) per ogni materiale fonico digitalizzato, una copia destinata all'archivio e un'altra, di backup, da conservare in caveau alla COMIT (banca COMmerciale ITaliana); classificazione e descrizione delle registrazioni originali e dei nuovi supporti mediante circa 40 attributi organizzati in un apposito database musicale in ambiente distribuito e multipiattaforma Oracle8 a oggetti, comprendente materiali fonici e partiture musicali integrati; progettazione e realizzazione di un ambiente di database dotato di strumenti per l'interrogazione dell'archivio musicale del teatro basati su frammenti audio e di partiture, oltre che sui tradizionali frammenti alfanumerici.

Nel 1999 ha inizio, ufficialmente, il progetto DAM il cui coordinamento scientifico è affidato al Dott. Goffredo Haus e la sua esecuzione ad Accen-

ture per gli archivi multimediali e al LIM per l'archivio musicale (fonico e partiture). Negli anni successivi ha preso forma la fase di integrazione degli archivi del Teatro in un unico ambiente informatico (affidata al LIM). Nell'autunno 2004, si è avuta infine la completa integrazione delle teche multimediali periferiche dei singoli archivi in un unico database centralizzato che servirà l'Intranet del teatro.

L'idea di fondo che ha guidato le procedure di digitalizzazione dei supporti audio è che i contenuti musicali dovessero essere acquisiti senza interventi di restauro dell'informazione fonica, da rimandare eventualmente ad un secondo momento, in presenza di una politica editoriale definita o comunque rivolta alla fruizione, esterna o interna al Teatro. Tuttavia, per i casi di supporti che oltre a non presentare uno stato di conservazione soddisfacente si fossero presentati in condizioni critiche, correndo il rischio di deteriorarsi completamente in breve tempo, le procedure standard per il riversamento conservativo dei contenuti fonici contemplavano alcune operazioni concernenti il restauro dei supporti deteriorati.

Accanto al trattamento dei materiali per la conservazioni dei contenuti fonici, è stata progettata e sviluppata l'applicazione M.A.I.S. (Musical Archive Information System): si tratta di un software fortemente innovativo, che oltre alle normali funzionalità di gestione e interrogazione di database, permette di accedere tanto alle registrazioni audio musicali quanto alle corrispondenti partiture in un ambiente integrato tra audio e partitura, consentendo, per fare un esempio, di trovare una partitura canticchiando un motivo o di reperire una registrazione scrivendo qualche nota sullo schermo.

I nuovi obiettivi sono perciò divenuti i seguenti:

- digitalizzazione degli archivi fonico, partiture e fotografico;
- schedatura fotografica di costumi, attrezzeria, bozzetti e figurini, e digitalizzazione di tali fotografie;
- informatizzazione dei cataloghi di tutti gli archivi arricchiti di nuove informazioni testuali e di tutti i materiali multimediali sviluppati nell'ambito del progetto D.A.M;
- progettazione, sviluppo e avviamento a regime di applicazioni software per i singoli archivi;
- progettazione, sviluppo e avviamento a regime di un'applicazione software di integrazione tra gli archivi digitali del progetto, che consente l'accesso a tutti gli archivi da ogni postazione informatica del Teatro.

2.1.3 Interfaccia

L'attuale versione del software opera sul server Amadeus. Ciò consente di realizzare, rispetto alla vecchia versione, interrogazioni inimmaginabili a causa della complessità computazionale e dei lunghi tempi di risposta, nonché di ottimizzare il codice preesistente e di effettuare *stress test*.

L'applicazione DAM centrale offre grande flessibilità nella navigazione, permettendo l'utilizzo di percorsi diversi per una stessa ricerca. Di pagina in pagina, è consentito un flusso informativo efficace ed efficiente, il quale connette tra loro contenuti eterogenei, ma correlati. Tutti i media relativi ad un allestimento o ad una serata sono accessibili da tale pagina; è, inoltre, possibile effettuare il percorso inverso, ossia risalire da un oggetto multimediale all'allestimento o alla serata cui questo fa riferimento.

La navigazione si può basare sugli strumenti standard messi a disposizione dal browser (pulsanti [BACK] e [FORWARD]) oppure sui controlli appositamente implementati, per facilitare la visita delle pagine.

Il software tiene costantemente traccia delle pagine visitate e, durante la navigazione, va a comporre un percorso "a tappe" che permette di tornare ai risultati della ricerca in modo veloce. Tale immediatezza si realizza evitando all'utente la pressione ripetuta del pulsante [BACK] per tornare alla pagina dei risultati, memorizzando quest'ultimi in una cache, in modo da renderli immediatamente disponibili all'utente dall'interrogazione precedente. In tal modo, si risparmiano i fisiologici tempi di risposta del sistema, dovuti all'esecuzione della query e all'invio dei risultati.

La pagina di autenticazione effettua il controllo di accesso da parte degli utenti, il quale viene eseguito prima di caricare ciascuna pagina. Una volta superato lo step di autenticazione, è possibile muoversi liberamente all'interno del sito. Non tutti gli utenti del DAM saranno autorizzati a vedere ogni singola pagina. L'applicazione, in ogni caso, nasconde dinamicamente i collegamenti "proibiti", sulla base della coppia nome utente-password immessa. I responsabili dei singoli archivi, ad esempio, hanno la piena visibilità su quanto è di loro competenza. Una volta effettuata l'autenticazione, l'utente si ritrova nella pagina di ricerca per brano e allestimento, la quale rappresenta anche la pagina di ricerca di default, vista la significatività dei risultati che permette di ottenere. I campi di ricerca previsti contemplano: tipo spettacolo, titolo, stagione, autore, artista, personaggio, luogo e oggetto. In un modulo di ricerca, essi possono essere compilati in qualsiasi numero e in qualunque combinazione. Teoricamente, è possibile lanciare una ricerca anche senza specificare alcun criterio restrittivo, il che porta alla visualizzazione dell'intero database. Tale possibilità è stata inibita solo nella

ricerca per evento.

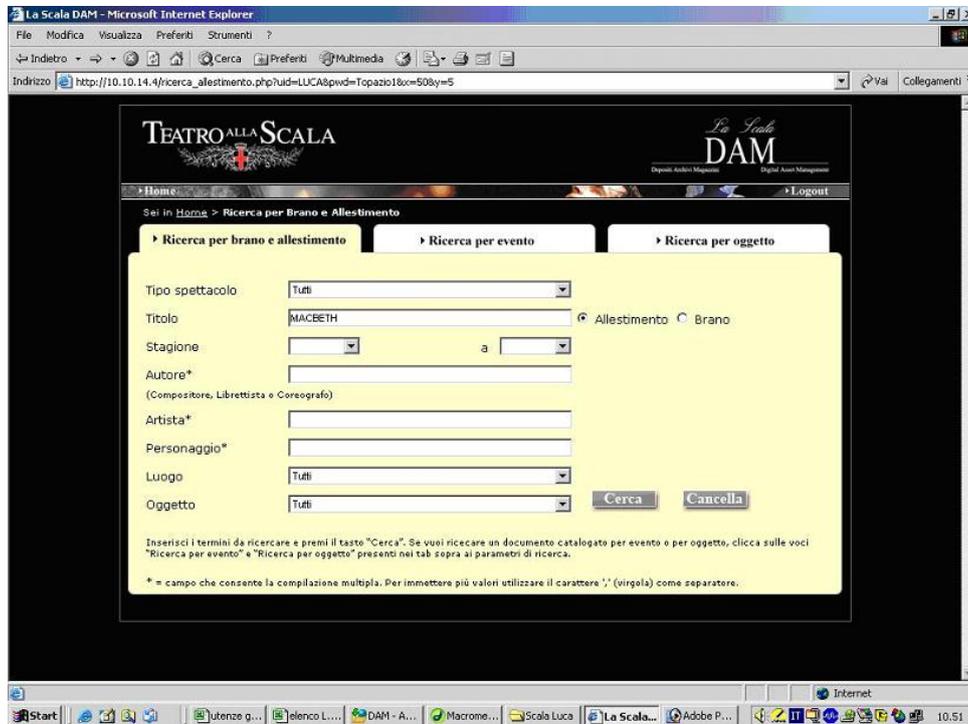


Figura 2.2: Pagina di ricerca per brano e allestimento

L'applicazione gestisce la ricerca per sottostringhe; è possibile, ad esempio, cercare tutti gli allestimenti in cui un personaggio contiene le lettere *ca*. Questa possibilità risulta essere utile nei casi in cui sia dubbia la corretta grafia. Ad esempio, il *Tannhauser* potrebbe trovarsi nel database con i titoli alternativi di *Tannhuser* o *Tannhaeuser*. Nell'incertezza, è possibile eseguire la ricerca immettendo come autore *Richard Wagner* e come titolo la sottostringa *tannh*. Questo problema in generale non si verifica per il titolo della composizione, dato che nel database vengono memorizzate più versioni per ogni allestimento (ad esempio, *Zauberflöte* ma anche *Flauto magico*); piuttosto, spesso è il nome del compositore a creare problemi, soprattutto nei casi di traslitterazione dal cirillico all'italiano (si pensi a *Tchaikowskij* nelle sue numerosissime varianti). In tali situazioni, l'estrazione di una sottostringa significativa e comune alle varie grafie dovrebbe portare al risultato desiderato.

Un numero via via crescente di campi compilati per la ricerca fornisce risultati sempre più precisi e puntuali, ma a scapito dei tempi di risposta del sistema.

I risultati della ricerca vengono riportati, di decina in decina, così come

mostra la 2.3, sotto rappresentata.

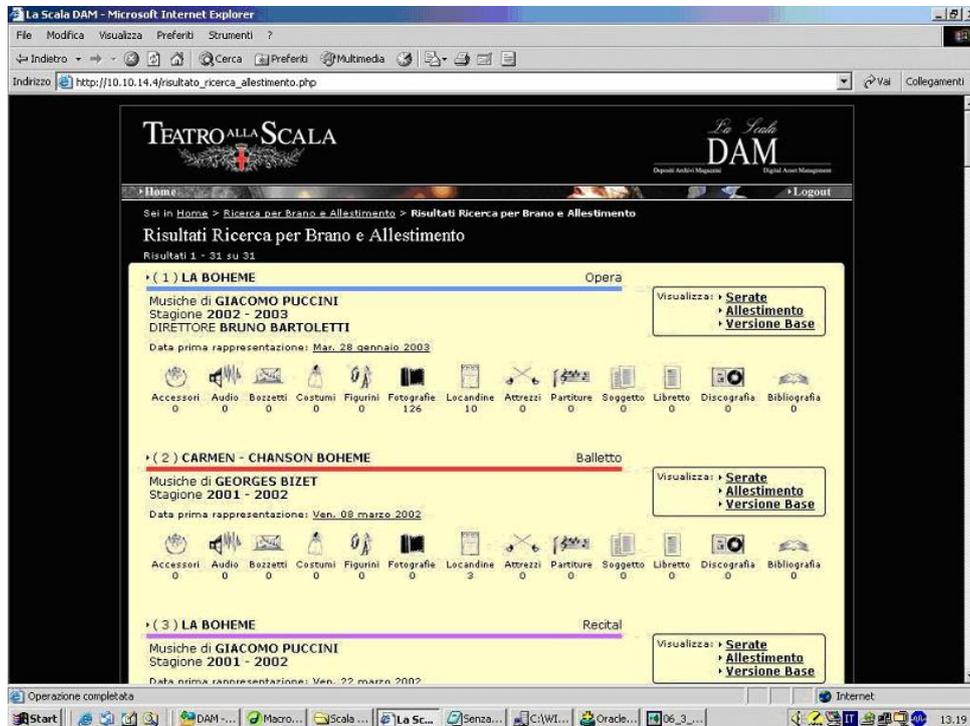


Figura 2.3: Risultati della ricerca per brano e allestimento.

A seconda del tipo di interrogazione, una stessa pagina di risultati può contenere tipologie diverse di allestimento: ad esempio opere, concerti e balletti. E' il caso della ricerca con il titolo *Boheme*, che estrae l'opera omonima, i recital e i concerti in cui ne sono stati cantati degli estratti, il balletto *Carmen Chanson Boheme* e le manifestazioni relative all'evento. Per permettere all'utente di distinguere a colpo d'occhio la tipologia di spettacolo, si è scelta una convenzione sui colori delle fascette che sottolineano i titoli (ad esempio, azzurro per le opere e giallo per i concerti).

Una parte fondamentale della visualizzazione è nella barra delle icone che accompagna ogni allestimento e ne enumera i contenuti multimediali associati, secondo quanto precedentemente visto. Nel caso in cui il materiale multimediale faccia riferimento alle serate e non all'allestimento (è il caso delle locandine, ad esempio), il numero si riferisce alla somma di tutti i materiali delle serate che compongono l'allestimento.

2.1.4 Risultati ottenuti e conseguenze

L'intervento conservativo per l'archivio fonico del Teatro alla Scala ha permesso la messa a punto di metodologie per la catalogazione e il trattamento dei materiali originali, consentendo inoltre la definizione di tecniche e la sperimentazione di sistemi che nell'insieme costituiscono una vera e propria tecnologia per la conservazione degli archivi fonici.

Per quanto la problematica della conservazione degli archivi multimediali sia sempre più attuale, i centri specializzati in questo campo sono ancora pochi. Gli esiti dei progetti informatici per il Teatro alla Scala hanno quindi richiamato una notevole attenzione, anche a livello internazionale, testimoniata anche dai numerosi articoli apparsi su riviste divulgative e scientifiche.

L'insieme delle attività progettuali svolte nell'area della conservazione dei beni informativi, artistici e non, ha avuto ricadute su almeno tre fronti: la formazione di giovani specialisti per nuove professionalità, il trasferimento di know-how verso le aziende pubbliche e private che affrontano l'esecuzione di progetti di conservazione e informatizzazione, la standardizzazione di forme di codifica dell'informazione multimediale, in particolare musicale.

2.2 Progetto ADMV

Il progetto dell'Archivio Digitale della Musica Veneta (ADMV) si propone di sperimentare e di mettere a regime un modello di servizio integrato per la ricerca e la consultazione di documenti che contengono musica notata. Esso si è prefisso di istituire un servizio di accesso, disponibile in rete, a tali documenti, offrendo la possibilità, attraverso una stretta integrazione funzionale con la registrazione catalografica della partitura musicale pubblicata su un OPAC¹, di navigare dal record bibliografico, relativo ad una partitura, alla sua immagine digitalizzata e all'eventuale documento sonoro digitalizzato corrispondente, attraverso tecnologie di distribuzione in rete di immagini e suoni.

Volgendo lo sguardo al ruolo dei produttori di contenuti che le tecnologie digitali e le reti consentivano alle istituzioni della memoria, i responsabili della Discoteca di Stato, della Biblioteca Nazionale di Torino e della Biblioteca Marciana di Venezia hanno cercato di individuare un terreno di cooperazione fra i tre istituti. Qualcosa che permettesse loro di valorizzare le loro collezioni ed al contempo di mettere in piedi un servizio di biblioteca digitale innovativo. Il contributo della Discoteca di Stato consiste principalmente nella messa a disposizione delle registrazioni sonore, mentre le due

¹La sigla OPAC, acronimo di On-line Public Access Catalog, identifica in modo generico i sistemi in rete che consentono la libera consultazione di cataloghi bibliografici in linea.

Biblioteche contribuiscono con le descrizioni bibliografiche e le immagini digitalizzate di manoscritti e musica a stampa.

Il primo problema affrontato, nell'intraprendere un progetto di tale portata, è stata la selezione dei materiali, ovvero stabilire "cosa" digitalizzare e "perchè". I criteri seguiti sono stati conseguenza immediata degli obiettivi definiti dal progetto, che potrebbero essere così sintetizzati:

- un obiettivo di natura culturale: ricondurre virtualmente ad unità collezioni omogenee altrimenti disperse, nel tentativo di aprire, nella migliore delle ipotesi, nuove prospettive di ricerca;
- un obiettivo di servizio: creare un servizio innovativo di biblioteca digitale con un forte accento sulla multimedialità, attraverso tecnologie di distribuzione in rete di immagini e suoni.

La "sostenibilità" del progetto era requisito fondamentale: solo servizi rivolti ad un bacino di utenza ampio e non specialistico potevano assicurarli la capacità di attrarre finanziamenti e nuovi partner per mantenersi nel tempo, al di là delle prime realizzazioni.

Sulla base di quest'impostazione iniziale è possibile motivare l'origine di alcune scelte tecnologiche e funzionali distanti dalla sperimentazione pura e più orientate a modelli tecnologici già disponibili e consolidati.

Nella fase iniziale, tale Archivio, ha visto l'integrazione delle basi di dati delle tre istituzioni coinvolte:

- quelle della *Biblioteca Nazionale Universitaria di Torino*, occupatasi della catalogazione e della scansione digitale del proprio fondo di partiture, autografe e non, di Antonio Vivaldi, la più ricca ed organica esistente (27 volumi per un totale di 7786 carte);
- della *Biblioteca Nazionale Marciana*, che, conservando un fondo manoscritto musicale di eccezionale importanza, soprattutto per il Sei e Settecento veneto, ha trattato in maniera analoga le partiture manoscritte di Alessandro e Benedetto Marcello (65 volumi), di Baldassarre Galuppi e Alessandro Stradella e immagini come quelle relative ai "Balletti sabaudi";
- della *Discoteca di Stato*, che ha digitalizzato e reso disponibili in rete un certo numero di esecuzioni sonore degli autori citati.

L'utente viene messo nella condizione di poter entrare in contatto diretto con le esecuzioni sonore riferite alla partitura tramandata. La relazione tra esecuzioni sonore e partiture musicali non è di natura filologica; questo perchè le esecuzioni avvengono, generalmente, su trascrizioni moderne.

Il contesto in cui ci si trova è quello di *biblioteca digitale* nell'accezione che, a questo termine, danno le istituzioni della memoria. Ci si riferisce, quindi, a tutte le attività necessarie a realizzare la scansione digitale di preesistenti oggetti analogici e a mettere in servizio gli oggetti digitali così realizzati a vantaggio degli utenti, con ricadute positive anche sulle attività di conservazione. Precisazione importante, in quanto non va dimenticato che l'accezione di biblioteca digitale propria, ad esempio, delle istituzioni della ricerca e delle università, è diversa ed è molto più attenta alle modalità di produzione, validazione, disseminazione ed uso dei documenti che nascono digitali.

2.2.1 Architettura

L'utente, locale o remoto, accederà, tramite un normale browser di rete con protocollo HTTP al web server ADMV, il quale si occuperà di implementare il sito corrispondente con tutte le informazioni sui servizi previsti. Egli potrà quindi navigare verso gli OPAC locali (Biblioteca Nazionale Marciana, Biblioteca Nazionale Universitaria di Torino, Discoteca di Stato) contenenti i dati di bibliografia musicale o condurre la ricerca simultanea su tali OPAC tramite un client z39.50. L'utente avrà le medesime possibilità anche accedendo, in base ad accordi da formalizzare con l'Istituto Centrale per il Catalogo Unico (ICCU), all'OPAC indice SBN, senza passare per il web server ADMV.

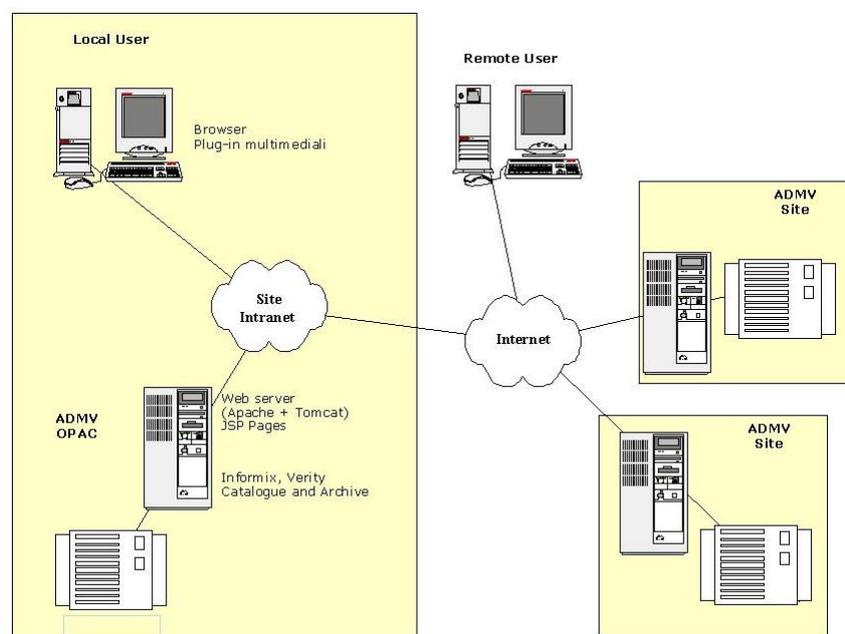


Figura 2.4: Architettura ADMV

Perciò, l'archivio non prevede la realizzazione di una base dati catalografica e di immagini centralizzata, ma di tante basi dati locali per il recupero dei record catalografici e degli oggetti digitali che siano consultabili unitariamente dall'utente finale. Il sistema si caratterizza come un insieme virtuale di archivi locali, secondo un modello distribuito. Ciascun polo ha piena autonomia organizzativa nella gestione delle collezioni, ha il suo catalogo, le sue attrezzature tecnologiche, i suoi applicativi. L'esito della ricerca sarà un record bibliografico, ad esempio la descrizione di una partitura musicale, a partire dal quale sarà possibile visualizzare l'immagine della partitura e delle sue parti componenti. Qualora si voglia accedere al documento sonoro, la navigazione porterà, comunque, alla visualizzazione dei record bibliografici relativi ai dischi o nastri che contengono le esecuzioni musicali e di qui, individuata l'esecuzione di interesse, all'ascolto del documento sonoro digitalizzato. E' prevista anche una ricerca per incipit musicale, suonato su una tastiera MIDI, che individuerà il record bibliografico del documento sonoro.

2.2.2 Standard utilizzati

Argomento centrale per ADMV è l'uso corretto e coerente degli standard, sia di carattere tecnologico, dettati sostanzialmente dal mercato, che funzionale.

Fra questi ultimi è possibile distinguere fra:

- Standard relativi ai record bibliografici;
- Standard relativi agli oggetti digitali.

Per quanto riguarda i primi, il progetto prevede l'adozione di UNIMARC, nel formato di alimentazione dell'OPAC di indice SBN, per le varie tipologie di materiali. A questo proposito, sono state concordate con ICCU alcune integrazioni alla griglia, al fine di gestire i codici di genere musicale, il codice di riproduzione su microforma, i collegamenti alla descrizione bibliografica del libretto e/o del documento sonoro corrispondente ad una partitura e i collegamenti con gli oggetti digitali.

La corretta formalizzazione del collegamento del record bibliografico con l'oggetto digitale consentirà di gestire in modo efficace le fasi di esportazione del record dall'applicativo gestionale e di importazione nell'applicativo OPAC, garantendo l'operatività delle funzioni di navigazione dall'OPAC verso la base dati degli oggetti digitali.

Per quanto riguarda questi ultimi, il processo di definizione di standard di natura funzionale non è ancora giunto ad esiti ben definiti. Il gruppo ADMV sta attualmente lavorando alla definizione di un certo numero di metadati, intesi come insiemi di informazioni relative agli oggetti digitali, funzionali

soprattutto all'uso che di tali oggetti verrà fatto. In questo senso, si ritiene che l'individuazione dei metadati necessari vada fatta in base a principi di economia, non di ridondanza, per facilitarne la successiva gestione e che vadano ridotti al minimo i dati di natura descrittiva. Esigenze di questo tipo devono venire soddisfatte al livello dei record bibliografici.

I metadati devono essere acquisiti e mantenuti al fine del loro utilizzo nel processo di digitalizzazione il quale può essere suddiviso in almeno tre fasi:

- acquisizione degli oggetti digitali;
- archiviazione permanente degli oggetti digitali;
- gestione degli oggetti digitali (per l'accesso, la conservazione etc.).

Dunque ADMV non è un software, ma obbliga l'adozione dei medesimi standard sia per i record bibliografici che per gli oggetti digitali. Per quanto riguarda i primi, gli applicativi di catalogazione utilizzati dai partner, pur differenti, devono disporre di funzionalità di export in formato UNIMARC standard; per i secondi, è necessario che gli oggetti digitali siano corredati di metadati amministrativi e gestionali in formato MAG, a prescindere dall'applicativo utilizzato per produrli. Non ci sono invece vincoli, relativamente agli standard di risoluzione o compressione delle immagini o del suono, i quali seguono l'evoluzione tecnologica.

I Metadati Amministrativi e Gestionali sopracitati hanno segnato il discrimine fra due diverse generazioni di progetti di digitalizzazione. La prima concentrava l'attenzione sulla riproduzione digitale. La seconda, invece, ha posto l'accento sui valori d'uso delle basi dati di oggetti che venivano creati a seguito della scansione digitale.

Il rapporto fra catalogazione e digitalizzazione è un problema che si ripresenta spesso nei progetti di scansione digitale, in quanto, quest'ultimo è un processo di natura seriale, mentre la catalogazione è un'attività parcellizzata, atomica. I tempi dell'una e dell'altra sono drammaticamente diversi. Vale sempre il principio generale secondo il quale non si deve procedere alla scansione digitale di materiale che non sia già stato catalogato e per il quale non si disponga di registrazioni standardizzate e disponibili al pubblico su sistemi di recupero delle informazioni.

I MAG, questo insieme di dati da associare a ciascun oggetto digitale, forniscono tra l'altro le informazioni necessarie per l'attivazione delle procedure di conservazione digitale (Digital Preservation).

I MAG, e gli schemi di metadati in genere, sono componenti necessarie

di un archivio digitale che rispetti lo standard OAIS (Open Archival Information System), il quale si autodefinisce come “un archivio consistente in un’organizzazione di persone e sistemi, che ha accettato la responsabilità della conservazione dell’informazione e del renderla disponibile per una determinata comunità”. A tale scopo “esso individua termini e concetti rilevanti per l’archiviazione di documenti digitali, identifica le componenti e i processi chiave comuni alla maggior parte delle attività di conservazione digitale e propone un modello logico di riferimento per gli oggetti digitali e i metadati loro associati, che comprende la creazione e l’uso dei metadati utili a gestire il materiale elettronico, dalla fase di acquisizione a quella della conservazione, fino alla fase di accesso” (Cirocchi, 2001). Di fatto è lo standard emergente per la conservazione delle risorse digitali. OAIS si pone però come modello logico di riferimento di valore generale ed astratto e prescinde da una specifica implementazione.

ADMV ha goduto della fiducia e dell’appoggio concreto da parte della Direzione Generale per i beni librari ed ha consentito di avviare, su vari aspetti, una collaborazione più che proficua con l’ICCU; da ultimo, è stato apprezzato e, in qualche modo, adottato dal Comitato Guida della Biblioteca Digitale Italiana, che ha disposto quanto necessario per coinvolgere un gruppo di nuovi partner, creando, tra l’altro, le condizioni per la caduta di quella V finale, a segnare l’estensione a documenti musicali diversi da quelli riconducibili all’area veneta.

2.2.3 ADMV o ADM?

Il contesto tecnico e funzionale di un polo ADM scaturito dal modello di architettura sopra delineato prevede tre componenti logiche e, generalmente, anche fisiche:

- un applicativo gestionale, con relativo database, per la catalogazione dei documenti musicali, in grado di esportare i record nel formato UNIMARC definito per il progetto. Gli applicativi attualmente utilizzati sono Archimusic/Archimedia in Marciana, nella versione completa delle funzioni di supporto alla creazione dei MAG, Cadmus, un applicativo sviluppato dalla Biblioteca Nazionale Universitaria di Torino su base SBN-musica, e WinDJ, l’applicativo di catalogazione della Discoteca di Stato;
- un applicativo di Information Retrieval OPAC, con relativo database, in grado di importare quei record, gestire servizi differenziati per tipologie di utenti e rendere disponibile, tramite navigazione, l’oggetto digitale (immagine o suono);
- un applicativo di gestione della teca digitale, che contiene gli oggetti

digitali e i relativi MAG. Il modello logico-funzionale della teca dovrebbe essere coerente con gli standard più avanzati per la conservazione permanente di tali oggetti (in particolare: OAIS, Open Archival Information System standard ISO 14721:2003), per non vanificare nel tempo gli investimenti per la digitalizzazione.

Tra le componenti tipiche è solito un servizio di scansione digitale dei documenti-fonte, in grado di produrre oggetti digitali completi dei relativi MAG.

Per quanto riguarda l'OPAC sono attualmente disponibili:

- una modalità di ricerca a testo libero su di un unico campo, che lancia una ricerca sull'intero record catalogafico, del tipo di quelle offerte dai motori di ricerca *full text*, con possibilità di interrogazione in linguaggio naturale (con la sintassi propria dello specifico motore utilizzato dall'applicazione) o booleano;
- una modalità di ricerca per campi, corrispondenti ai campi predefiniti sulla base dati per ogni record catalogafico. All'inserimento di valori in più di un campo corrisponde una ricerca che fornirà all'utente risultati soddisfacenti, in base a tutte le condizioni da lui richieste (AND logico).

2.2.4 Criticità dell'ADMV

In corso d'opera sono emersi alcuni problemi derivati da aspetti sottovalutati all'inizio e che hanno comportato un allungamento dei tempi di realizzazione del progetto. Ci sono tre aspetti da considerare:

- *UNIMARC*: uno dei capisaldi dell'ADMV è l'indipendenza dagli applicativi di catalogazione, la quale obbliga gli stessi ad una gestione uniforme degli export UNIMARC. All'inizio, solo l'applicativo utilizzato dalla Biblioteca Marciana disponeva di funzioni native di export UNIMARC. Si sono resi quindi necessari un lavoro accurato di mappatura di ciascun applicativo sugli altri e di tutti questi sullo standard UNIMARC e una serie di interventi sui medesimi applicativi per renderli conformi allo standard e realmente interoperabili. Tutto questo in un contesto da un lato di carenza e conseguente instabilità dello standard internazionale, specie relativamente ai manoscritti musicali, e dall'altro di ritardo nello sviluppo teorico e applicativo di quello stesso standard a livello nazionale. Attualmente, il formato UNIMARC, gestito dai partner ADM (UNIMARC-ADM), recepisce molte delle nuove proposte UNIMARC di imminente pubblicazione, ma non coincide con il formato UNIMARC gestito dall'attuale Indice 2 di SBN (UNIMARC Italia). Allineare i due standard significherebbe, per chi

si occupa di tale progetto, fare un passo indietro con costi insostenibili. L'augurio è che si reperiscano le risorse necessarie per aggiornare il formato UNIMARC SBN di Indice 2 per la musica alla più recente versione dello standard internazionale.

- *Bonifiche catalografiche*: è, tradizionalmente, un aspetto critico, di cui di solito non si ha piena consapevolezza in fase di progettazione. Si tratta di un'attività tanto più importante quanto più resta centrale il ruolo dell'OPAC e, quindi, delle rappresentazioni formalizzate degli oggetti-fonte, come strumento di reperimento degli oggetti digitali associati alle rappresentazioni stesse. In ADMV, questa attività ha spaziato dalla correzione alla creazione di record di monografie o di spogli. Anche qui si è impiegato del tempo, ma la qualità dell'informazione ne ha tratto beneficio. Sarebbe sempre preferibile avviare la digitalizzazione dopo la conclusione delle bonifiche catalografiche e la pubblicazione dei record su OPAC; in subordine, si può tenere presente che per la digitalizzazione, che oggi non può prescindere dalla creazione contestuale di una serie di metadati, è, generalmente, sufficiente disporre di un buon record in formato Dublin Core.
- *Metadati amministrativi e gestionali (MAG)*: si tratta di una criticità positiva di natura non catalografica, come si conviene ad un progetto di scansione digitale. Lo standard MAG, adottato con convinzione dal gruppo di lavoro, è stato definito quando ADMV era già un progetto maturo e la sua implementazione ha comportato una serie di interventi non previsti inizialmente sia sugli applicativi di catalogazione che sul software di supporto alle operazioni di scansione digitale, in modo da suddividere fra queste due componenti l'onere della generazione dei MAG, ottimizzandone il flusso produttivo.

2.2.5 Interfaccia

Le interfacce sono di tipo tradizionale e rispondono ad abitudini di ricerca consolidate. Si è ritenuto opportuno consentire immediatamente in ambedue le maschere di ricerca (semplice e avanzata) l'uso di filtri per selezionare a priori solo i record con immagini collegate o solo quelli con audio collegato; in entrambe è disponibile un filtro di ricerca per organico che consente, tramite l'apertura di una specifica finestra, di selezionare o escludere gli strumenti desiderati. Sono poi ben evidenti nelle maschere di prospettazione analitica dei record le icone di collegamento con gli oggetti digitali, per facilitare chi fosse meno interessato agli apparati descrittivi e volesse puntare direttamente alle immagini o al suono.

Dovendo comunque fare i conti con registrazioni catalografiche spesso ricche e dettagliate, data anche la natura e l'eccezionalità dei materiali, si è

pensato di articolare la prospettazione analitica dei record in tre macroaree informative, secondo uno schema empirico: “cosa dove altre informazioni”:

- *cosa* (in interfaccia: dati identificativi) ha la funzione di *individuare e definire* con certezza (e con il minimo di dati necessari) il documento reperito;
- *dove* (in interfaccia: localizzazione) permette di localizzare con esattezza il documento;
- *altre informazioni* permette di ottenere altre informazioni descrittive di contestualizzazione, approfondimento e dettaglio.

Già la prima macroarea da sola dovrebbe soddisfare le esigenze della maggior parte degli utenti interessati alla fruizione diretta degli oggetti digitali. L’utente può salvare i record reperiti in un’area personale e riservata (carrello), al fine di recuperarli successivamente, stamparli, trasmettere l’esito della ricerca tramite la posta elettronica o rieseguire la strategia di ricerca a fronte di modifiche avvenute nella base dati.

Non sono ancora disponibili, ma sono in via di realizzazione, le modalità di ricerca per incipit musicale, con il meccanismo di trascinamento delle note su un pentagramma a video.

2.2.6 Prospettive di sviluppo

E’ prevista l’adesione alla rete della Biblioteca Nazionale Centrale di Roma e della Biblioteca Estense Universitaria di Modena. Queste, in attesa di dotarsi delle piattaforme tecnologiche necessarie, trasferiranno oggetti digitali e MAG, prodotti con recenti campagne di digitalizzazione, nel sistema sperimentale attivo presso l’ICCU, dotato di OPAC e teca digitale ADM.

Il contesto di un possibile sviluppo di ADM è dato dal progetto Biblioteca Digitale Italiana/Network Turistico Culturale (BDI/NTC), che propone la qualificazione culturale dell’offerta turistica del nostro paese, e in particolare una linea tematica dedicata ai contenuti musicali, denominata Rete della Musica Italiana (ReMI). Si tratta anche, naturalmente, di inserire le istituzioni della memoria, in quanto detentrici di contenuti culturali, valorizzabili anche dal punto di vista economico, in un ambiente in cui sia più facile reperire finanziamenti o produrre ritorni economici. Il primo prodotto tangibile di BDI/NTC è un portale per l’accesso pubblico ai contenuti digitali nazionali, in grado di veicolare anche i contenuti ADM e che comprende funzionalità di e-commerce per l’acquisto di copie dei documenti.

Da un punto di vista architettuale, il sistema prevede la centralizzazione fisica di alcuni metadati standard MAG, come strumenti di reperimento degli

oggetti digitali. Per trasportare i MAG dalla periferia al sistema centrale, si utilizza il metodo dell'harvesting, implementando il protocollo OAI-PMH, che si sta rivelando uno standard efficace anche al di fuori dell'ambiente dell'Open Access in cui è stato originariamente sviluppato. A questo fine, le teche digitali ADM, che in termini OAI-PMH si comportano come data provider, si sono attrezzate per "esporre" i metadati secondo tale protocollo e divenire passibili di harvesting da parte di NTC, che funge da service provider. In questo modo NTC è in grado di reperire e consegnare, a domanda di un utente, gli oggetti digitali presenti nelle teche locali, gestendo in aggiunta la transazione commerciale. L'utilizzo di OAI-PMH consente fra l'altro alle teche di essere indagabili e fruibili anche da parte di altri service provider che abbiano implementato il medesimo standard. Ovviamente le transazioni commerciali riguardano copie digitali di alta qualità, mentre non sono previsti costi per la fruizione degli oggetti liberamente disponibili in internet.

Capitolo 3

Le basi di dati

3.1 Database multimediali

L'evoluzione tecnologica degli ultimi anni ha permesso l'acquisizione, la memorizzazione e la distribuzione a costi ridotti di sorgenti di informazione innovative. Immagini, filmati, file audio e frammenti di testo scritto a mano rappresentano gli esempi più noti di queste nuove tipologie di dati. Le nuove tecnologie hanno portato alla riduzione dei costi necessari alla gestione ed all'acquisizione di queste nuove sorgenti.

Alla sempre maggiore disponibilità di sorgenti multimediali ed all'evoluzione dell'hardware necessario alla loro gestione non è seguito un uguale sviluppo degli strumenti software forniti agli utenti per recuperare, nel minor tempo possibile, le informazioni conformi alle loro richieste. I database che memorizzano dati classici, come numeri, stringhe di testo e campi dati, si avvalgono, per il recupero di tali informazioni, di potenti linguaggi di interrogazione come SQL, e di strumenti per la gestione di Basi di Dati Relazionali (RDBMS) per ottimizzare i tempi di ricerca. Questi tipi di dati sono noti dall'introduzione dei database ed è ovvio che le tecniche di recupero siano molto più evolute di quelle sui nuovi dati multimediali. Negli ultimi anni, si è verificato il crescente bisogno di realizzare sistemi in grado di interrogare ed elaborare vaste quantità di questi dati.

Un sistema per la gestione di database multimediali MMDBMS (Multimedia Database Management System) altro non è che un ambiente che organizza differenti tipi di dati, i quali possono essere rappresentati in una moltitudine di formati e risiedere fisicamente su media diversi.

Per poter funzionare in maniera efficiente, un MMDBMS deve soddisfare le seguenti caratteristiche:

- deve avere la possibilità di interrogare in maniera uniforme i dati

memorizzati in diversi formati. Ad esempio, un MMDBMS deve fornire la possibilità di interrogare ed integrare, in maniera costante, i dati provenienti da database relazionali, file di dati non strutturati oppure dati memorizzati in DBMS di tipo Object Oriented o spaziali. In particolare, interrogazioni che spaziano su sorgenti memorizzate in DBMS diversi devono essere in grado di fondere i risultati e presentarli in maniera uniforme;

- deve essere possibile interrogare dati appartenenti a media diversi. Ad esempio, un MMDBMS deve essere in grado di interrogare contemporaneamente sia un database di immagini che uno audio e/o uno relazionale ed integrare i risultati in maniera adeguata;
- deve rendere possibile il recupero di oggetti appartenenti a media diversi, memorizzati localmente, con continuità e senza evidenti ritardi. Siccome oggetti multimediali, come i video, spesso occupano grosse quantità di memoria, anche in formato altamente compresso, bisogna tenere in considerazione il fatto che questi dati possono essere registrati su sistemi di memorizzazione che garantiscono prestazioni diverse;
- deve, partendo dalla risposta di un'interrogazione, sviluppare una presentazione audiovisiva di questi risultati. Indipendentemente dal fatto che l'interfaccia ed il contenuto di ciascuna presentazione possa variare da un'applicazione all'altra, l'utente deve avere la possibilità di specificare la struttura ed il contenuto della risposta che vuole ottenere dal sistema;
- deve, quando la presentazione è stata delineata, secondo le specifiche del punto precedente, essere possibile distribuirla in maniera tale che venga rispettata la qualità di trasmissione del mezzo impiegato. Ad esempio, se al punto precedente, il sistema che sovrintende alla presentazione ha deciso che un filmato ed un file sonoro devono essere presentati in parallelo, il MMDBMS deve essere in grado di garantire che questa non subisca rallentamenti o sfasamenti dovuti ai requisiti diversi che i due flussi di dati devono soddisfare.

3.1.1 Architetture per database multimediali

Poichè i risultati delle interrogazioni vengono trasmessi attraverso i nodi della rete, per raggiungere i vari utilizzatori, il sistema di gestione della base di dati multimediale deve garantire sempre una banda di rete adeguata e la disponibilità dei buffer necessaria per mantenere una qualità di servizio tale da trasmettere senza interruzioni la presentazione dei risultati prodotti dal sistema. Negli ultimi decenni, sono stati sviluppati i linguaggi di interrogazione, le tecniche di indicizzazione, gli algoritmi di recupero e i metodi di

aggiornamento impiegati nelle basi di dati relazionali, orientate agli oggetti, spaziali e temporali ed altri tipi di database. Ognuno di questi sistemi ha esteso i linguaggi, gli algoritmi e le conoscenze precedenti allo scopo di incorporare nuovi ed importanti tipi di dati. In quest'ottica, i dati multimediali non sono differenti dai precedenti e le novità devono integrarsi nel panorama esistente senza richiedere di ricreare tutto di nuovo.

Quando si realizza un database multimediale che rappresenta un'ampia varietà di oggetti appartenenti a media diversi, occorre effettuare un confronto con diversi aspetti riguardanti l'organizzazione del loro contenuto informativo.

Le possibili architetture, per l'organizzazione del contenuto di un database multimediale, adottano i seguenti principi:

- *Principio di autonomia*: raggruppamento di tutte le immagini, tutti i video e tutti i documenti e creazione di un indice per ciascun raggruppamento, in modo da ottenere la massima efficienza possibile per la rispettiva sorgente. Si parla di autonomia poichè ciascuna sorgente mantiene un'organizzazione totalmente separata dalle altre.
- *Principio di uniformità*: utilizzo di una singola struttura astratta impiegata per tutti i tipi di media supportati dal sistema; creazione di un indice unico che indirizza tutte le entità del sistema indipendentemente dalla loro natura.
- *Principio di organizzazione ibrida*: adozione di una soluzione ibrida. Alcuni tipi hanno la loro struttura specifica, secondo l'approccio autonomo, mentre altri invece, in accordo con l'approccio omogeneo, adottano una struttura unificata. La scelta di quei dati che devono essere trattati in maniera uniforme o, viceversa, in maniera autonoma, dipende dalle scelte progettuali, le quali analizzano le caratteristiche dell'organizzazione dei dati che si vogliono privilegiare.

Le architetture basate sul principio dell'autonomia richiedono la creazione di algoritmi e strutture dati per ogni tipo di media supportato. Inoltre, sono necessarie tecniche per riunificare i risultati parziali ottenuti, interrogando le singole strutture dati. D'altro canto, creare strutture specializzate che possano accedere efficientemente a ciascun tipo di dato permette di ridurre drasticamente i tempi di accesso e di risposta alle interrogazioni.

Quando occorre integrare basi di dati esistenti con strutture dati ad algoritmi ereditati, il principio di autonomia ne permette l'utilizzo immediato, senza sviluppare ulteriore codice. I sistemi orientati agli oggetti implementano il principio di autonomia trattando ciascun tipo di dato come un oggetto i cui metodi sono accessibili dal MMDBMS globale. In contrasto con il principio di autonomia, quello di uniformità richiede la determinazione di una

struttura comune che possa contenere le informazioni relative a immagini, filmati, suoni e così via. Quest'architettura richiede di rappresentare quello che tutti i media hanno in comune con gli altri e di creare un unico indice su questi dati. In pratica, il principio di uniformità viene realizzato attraverso i concetti di annotazioni o metadati, dove le informazioni sul contenuto di ciascun elemento multimediale vengono espresse attraverso un metalinguaggio comune. Il vantaggio principale dell'adozione del principio di uniformità sta nella facilità di implementazione dei relativi algoritmi e nella assenza di agglomerazione dei risultati parziali. Lo svantaggio consiste nella modalità con cui vengono create le annotazioni, che può essere manuale o automatica. L'operazione di annotazione manuale può risultare molto costosa ed in questo caso le annotazioni risultano dipendenti da chi le ha inserite. L'annotazione automatica, invece, rischia di essere troppo superficiale trascurando informazioni importanti per la caratterizzazione della sorgente. Il principio dell'organizzazione ibrida si avvantaggia dei punti positivi di entrambe le architetture precedenti, cercando contemporaneamente di attenuare molti aspetti negativi.

3.1.2 Oggetti multimediali, feature e concetti

Benchè, a tutt'oggi, non esista un modello standard per la descrizione dei dati multimediali, è opinione comune che tutte le applicazioni non banali debbano considerare una rappresentazione su (almeno) quattro livelli:

- *raw data*: descrive i dati multimediali veri e propri, indipendentemente dal loro contenuto. Questo livello è rilevante per gli aspetti di memorizzazione, ma non di ricerca;
- *MM objects description*: a questo livello sono definiti gli "oggetti" multimediali di interesse (ad esempio parti di immagini o elementi di testo);
- *feature (caratteristiche)*: le feature di un oggetto multimediale ne descrivono il contenuto in termini di grandezze misurabili (ad esempio colore di un'immagine, spettro di un suono, ecc.). La definizione delle feature di interesse è dipendente dal dominio applicativo e la sua estrazione è tipicamente svolta in maniera automatica. Il confronto tra i valori delle feature è normalmente di tipo non esatto e richiede la definizione di criteri di similarità;
- *concepts*: il livello dei concetti usa una rappresentazione semantica del dominio di interesse (background/domain knowledge) per interpretare il contenuto degli oggetti multimediali ("cosa sono" = "quali entità del mondo reale rappresentano"). Il processo di classificazione è spesso manuale o semiautomatico.

Oltre a memorizzare i raw data e le informazioni che li accompagnano, è necessario provvedere ad identificare gli oggetti di interesse, ad estrarre le feature ed a riconoscere i concetti.

I dati sono tipicamente non strutturati. Se si vuole analizzare il contenuto di un documento multimediale e rappresentarlo, è necessario scegliere le caratteristiche importanti (metadati) da associare ai raw data nella base di dati. La rappresentazione può avvenire secondo due modalità per riferimento esterno o interno. Il grosso vantaggio della prima rispetto alla seconda è l'assenza di limiti nella grandezza del contenuto, dettati soltanto dalla capacità dei supporti fisici sottiacenti. Al contrario, la modalità interna permette una gestione consapevole (dal punto di vista della base di dati) dei contenuti, la quale porta a livelli elevati di protezione, di salvataggio e di categorizzazione.

- *Rappresentazione tramite riferimento esterno:*
il database memorizza un riferimento ai file che contengono i dati multimediali, i quali non vengono gestiti dal DBMS. Tale metodologia prevede l'utilizzo di campi di tipo BFILE in cui sono mantenuti soltanto dei locator ovvero delle referenze alla locazione fisica, dove i dati multimediali sono effettivamente memorizzati. In tal modo, è possibile utilizzare qualsiasi tipo di supporto (hard disk, media ottici...) indipendentemente dalle capacità del database di organizzare, ad esempio su più supporti, i contenuti delle proprie tabelle.
- *Rappresentazione tramite riferimento interno:*
i dati multimediali vengono memorizzati nel DBMS in campi di tipo LOB (Large Object); il DBMS gestisce questi dati dal punto di vista dell'accesso, ripristino e autorizzazione.

I LOB facilitano la memorizzazione di dati multimediali (documenti, immagini, audio, ecc.). Il DBMS non associa alcuna interpretazione a questi dati.

Le tipologie di dati LOB consistono di un LOB locator (un puntatore ai dati) e di un valore LOB (il dato attuale) e possono essere così classificate:

- *Binary LOB (BLOB):* usato per immagazzinare raw data (binari) come testo formattato, immagini o contenuti audio e video;
- *Character LOB (CLOB):* caratteri non formattati, i quali utilizzano il set di caratteri predefiniti del database;
- *National Character LOB (NCLOB):* caratteri non formattati, i quali utilizzano il set di caratteri predefiniti del database nazionale.

I LOB sono supportati da SQL-99 e non sono fisicamente memorizzati esternamente alle tabelle, ma internamente al database. Un BLOB è un qualsiasi ampio singolo blocco di dati memorizzati in un database. Essi possono contenere tutti i dati in forma binaria, per esempio, grandi file di testo, documenti per elaborazione dei dati, grafici, immagini, video e file di musica. I BLOB sono la scelta migliore quando l'informazione memorizzata non contiene alcun dato che possa essere utilizzato nelle operazioni di ricerca: il contenuto di questi campi non ha alcun significato per il DBMS e viene gestito senza poterne ispezionare porzioni del contenuto.

Un CLOB, invece, è una stringa di caratteri di dimensione variabile che ha il vantaggio di permettere al database di effettuare operazioni di ricerca su parti del contenuto del campo. Così facendo, l'utente può specificare nella stringa di ricerca condizioni riguardanti il contenuto del CLOB, utilizzando direttamente il linguaggio naturale di descrizione dell'informazione.

Per ovviare alle problematiche di internazionalizzazione dei caratteri e dei dati, il NCLOB permette la specifica del set di caratteri con cui interpretare il contenuto binario dei dati.

3.1.3 Differenze tra DBMS e MMDBMS

DataBase Management Systems e MultiMedia DataBase Management Systems hanno alcuni obiettivi comuni quali:

- persistenza, privacy, integrità, non ridondanza, ripristino;
- indipendenza dei dati: separati dall'applicazione e dalla presentazione;
- controllo di concorrenza: transazioni e prestazioni;
- supporto per le interrogazioni;
- controllo delle versioni.

I MMDBMS esprimono alcuni requisiti supplementari a causa della tipologia dei dati memorizzati:

- gestione di grandi quantità di memoria;
- integrazione, composizione e presentazione di media diversi;
- interfaccia multimediale / multimodale.

Alcuni requisiti sono molto diversi (e più complessi):

- ricerca approssimata in opposizione alle interrogazioni in stile SQL (ricerca esatta) possibili nei DBMS;

- funzionalità di interrogazione (e metodi di accesso) multimediale;
- domini applicativi ampi e differenziati;
- livello di interazione e tempo di presentazione dei risultati.

Rispetto ai DBMS tradizionali, un MMDBMS affronta problematiche differenti:

- dimensioni maggiori del database;
- gestione di media continui (audio, video);
- complessità di modellazione degli oggetti multimediali;
- ricerche non necessariamente “esatte”;
- supporto di fast-forward e di rewind;
- richiesta di estrazioni di sequenze di frame e/o frame singoli.

Argomenti importanti nella progettazione di MMDBMS sono:

- modellazione di dati multimediali;
- memorizzazione di oggetti multimediali;
- integrazione, diffusione, presentazione multimediali;
- indicizzazione multimediale, retrieval, browsing;
- supporto per interrogazioni multimediali;
- database distribuiti.

L'informazione multimediale è strutturalmente semplice, ma il significato è complesso ed a volte sfumato:

- la struttura dipende dai media (immagini, video, audio);
- la dimensione dei dati è importante;
- il significato è esterno ai dati.

Le interrogazioni invece dipendono da livelli di linguaggi “intermedi” che possono essere complessi:

- bassa (o nessuna) corrispondenza tra forma e significato di un oggetto;
- query by example;
- similarity search vs. corrispondenza esatta (booleana).

La classificazione e la ricerca dei dati codificati si basa sull'interpretazione convenzionale del loro significato, in funzione del contenuto oggettivo:

- modelli di definizione basati sulla struttura e sulle relazioni reciproche;
- linguaggi di interrogazione basati sulla composizione di condizioni esatte.

La classificazione e la ricerca di informazioni generiche si appoggiano a strutture informative supplementari che consentono di ricondursi all'obiettività dell'interpretazione. Tutto questo implica l'utilizzo di tecniche non necessariamente deterministiche e certe per fornire a sistemi informatici la capacità di maneggiare tali ricerche. Questo perchè il concetto di informazione connessa ai dati deve essere indotto artificialmente sintetizzando le caratteristiche attraverso cui categorizzarli.

L'interpretazione di dati non codificati si appoggia a convenzioni e conoscenze da loro non indotte, ma che seguono:

- presupposti culturali e ambientali;
- ambiguità del linguaggio naturale;
- contesto specifico;
- consuetudini e regolamenti.

Le strade percorribili sono due: la prima, più semplice, consiste nell'estrapolare metadati da associare al file multimediale ed effettuare la ricerca limitatamente ai campi contenenti i metadati; la seconda, più complessa ma altrettanto potente, prevede la strutturazione dell'informazione multimediale in oggetti che la rappresentano formalmente e la ricerca basata su questi ultimi oggetti. In tal modo, ci si avvicina sempre maggiormente alla ricerca per contenuti, la quale potrebbe permettere, in futuro, di interrogare il sistema canticchiando una melodia o fornendo un semplice audio pre-registrato.

3.1.4 Applicazioni dei MMDBMS in ambito musicale

La musica, nelle sue varie forme e rappresentazioni, costituisce un patrimonio dell'umanità piuttosto eterogeneo: non si compone unicamente di accordi e partiture, ma anche di suoni, di voci, di immagini (fotografie e registrazioni video), di oggetti artigianali (costumi, accessori, strumenti per il palcoscenico) o, comunque, dovuti alla attività manuale dell'uomo (acconciature e make-up).

A seconda della tipologia di istituzioni coinvolte, ad esempio un teatro d'opera piuttosto che un centro di editoria musicale o, ancora, centri di produzione e divulgazione musicali, la struttura del database multimediale sarà intrinsecamente legata ai loro specifici contenuti. Un database ben strutturato per un teatro d'opera è intrinsecamente diverso da uno fatto per altri scopi. I materiali di interesse, in questo contesto, includono sia metadati che dati e quest'ultimi possono essere eterogenei. Tali database devono immagazzinare un ampio numero di documenti di diverso tipo (per esempio: foto, testi, immagini e partiture), ma anche contenuti multimediali i quali, naturalmente, richiedono un MMDBMS.

Qui di seguito, saranno presentate tre diverse tipologie di database multimediali, i quali interessano gli ambiti e materiali di cui si è accennato, il teatro d'opera e i centri di conservazione e divulgazione musicali e le partiture manoscritte.

3.1.5 Database multimediali per teatri d'opera

La fruizione di un lavoro musicale dal vivo, all'interno di un ambiente teatrale, è un'esperienza molto difficile da ricreare in un contesto diverso o da rimaneggiare a posteriori. Tuttavia i teatri operistici rappresentano importanti centri per la produzione e la diffusione della cultura. Essi pongono garanzie affinché il lavoro non vada immediatamente perso a seguito della performance.

I processi, che in un ambiente così caratterizzato sono molto eterogenei e complessi, includono, oltre all'amministrazione economico-gestionale e le attività logistiche che hanno luogo negli uffici, anche le produzioni artistiche sul palcoscenico e i risultati di attività manuali svolte in laboratori. Probabilmente, queste ultime sono le più interessanti dal punto di vista di un discorso di eredità culturale. Affinchè essa possa essere ottenuta, attraverso la trasformazione di performances musicali, è necessario mettere in luce quali informazioni dovrebbero essere prese in consegna, strutturate e ricercate. Ragion per cui, le possibilità tecnologiche sempre in crescita e la necessità di organizzare l'ampio materiale a disposizione in database hanno recentemente coinvolto anche gli ambienti teatrali.

Nel caso di teatri d'opera, in particolare, i database possono essere impiegati non solo per immagazzinare dati amministrativi e gestionali, ma anche per preservare e trasmettere l'attività artistica stessa. La tecnologia, in continua evoluzione, si muove nei più disparati campi.

A partire da ciò, un crescente numero di istituzioni culturali ha preso in considerazione il bisogno di controllare e maneggiare questa vasta quantità

di dati e di digital media assets ¹, le quali rimandano concettualmente alla locuzione “Digital Asset Management” (o semplicemente DAM), con riferimento al progetto del Teatro alla Scala di Milano.

La progettazione e l’implementazione di un efficace ed efficiente Digital Asset Management per ambienti teatrali, in grado di conservare e organizzare le informazioni, permette agli utenti di salvare tempo e denaro. In tal contesto, sono inoltre cancellati i lunghi tempi di ricerca effettuati in archivi tradizionali.

Organizzazione del database

Per quanto concerne la tipologia di informazioni che dovrebbero trovare posto all’interno di un database progettato per un teatro d’opera, ne considerano due differenti categorie:

- i dati relativi all’amministrazione e alla gestione del teatro (informazioni sul personale, sugli stipendi o gli indirizzi);
- i dati e i metadati sulla loro produzione artistica (registrazioni, foto o il cartellone).

Spesso, i due differenti ambiti sono strettamente correlati. Ad esempio, informazioni precise circa la presenza di un artista in un certo numero di performances potrebbero essere usate non solo per ricostruire il cast della rappresentazione, ma anche per calcolare il suo salario. Molti altri esempi potrebbero essere citati: le relazioni tra i costi di un particolare allestimento e l’andamento di una rappresentazione, la disponibilità delle registrazioni ad essere messe in vendita e così via. Per quanto riguarda il secondo punto, esso è peculiare in un ambiente teatrale.

Un comune teatro d’opera, di solito, ospita differenti tipologie di spettacoli: opere, balletti, concerti sinfonici, concerti da camera e recitals, o anche altre iniziative quali lezioni, conferenze e presentazioni. All’interno di un siffatto ambiente vi sono tre concetti chiave da tenere presente nello sviluppo del database di riferimento: *la versione base, l’allestimento e la serata*.

La *versione base* rappresenta l’opera così com’è stata concepita dai suoi autori in termini di strutturazione (suddivisione in atti, scene) e composizione dell’organico. A questi dati si aggiungono quelli sugli stessi autori, edizione, prima rappresentazione assoluta, eventuali note e via dicendo. Dunque, le informazioni presenti nella versione base sono comuni anche a tutte le sue riprese, ossia agli allestimenti e alle serate. Questa definizione è molto

¹Forme di media convertite in fonti binarie. Il termine “asset” è usato per indicare il fatto che file così fatti hanno una sorta di valore intrinseco che li rende utili da maneggiare.

significativa per le opere, per i balletti e i concerti e non per altri tipi di manifestazioni quali le conferenze.

L'*allestimento* è una particolare messa in scena di una composizione. Tale concetto aggrega i dati che caratterizzano ogni specifica riproposizione della versione base. Si tratta di un set di performances caratterizzate dallo stesso programma musicale, appartenenti alla stessa stagione e aventi in comune caratteri inerenti la produzione, l'allestimento e il cast. Ad esempio, il regista, lo scenografo e il costumista di un'opera rimangono solitamente invariati per tutte le rappresentazioni che compongono un allestimento in una determinata stagione.

Esiste anche la possibilità che un allestimento sia la ripresa di uno precedente, ma per motivi cronologici questo legame stretto si perde e i due allestimenti vengono ricondotti a due diverse serate. Nel corso di una sola stagione, l'allestimento è generalmente unico, anche quando ha luogo in posti differenti, ma questo non si verifica nella totalità dei casi. Per quanto riguarda i concerti, per esempio, il concetto di allestimento sottintende ogni singola composizione (solitamente compiuta) che fa parte del concerto stesso. Ad esempio, il concerto di Natale del 23 dicembre 2003 si compone di due allestimenti: la *Missa Solemnis* di F.X. Süssmayr e i *Vesperae Solennes de Confessore* di W.A. Mozart. Pertanto, cercare un allestimento dal titolo *Concerto di Natale* non porta ad alcun risultato. Per superare questo problema si è introdotto il concetto di serata. L'identificativo dell'allestimento viene generalmente attribuito all'intera composizione compiuta e non alle sue sottoparti. Bisogna però prestare attenzione ai casi in cui vengono eseguite solo alcune parti di una composizione, caso tipico dei recital (arie d'opera, estratti, ecc.).

La *serata* (o *evento*) è legata ad una particolare esecuzione della composizione, in una certa data e in un determinato luogo. Essa presenta un legame univoco con la locandina. Una delle informazioni più interessanti della serata è quella relativa al cast, dato necessariamente associato all'evento: nel caso di opere e balletti che prevedono più rappresentazioni, il cast non è unico per l'allestimento, ma lo è sicuramente per la serata. Quest'ultima rappresenta l'implementazione di una versione base (o di più versioni base o di una loro selezione), in accordo con un luogo preciso, in una particolare data e con un cast ben definito di interpreti.

Il concetto di serata non compare nella stessa struttura gerarchica di cui fanno parte la versione base e l'allestimento, ma darà il nome ad un tipo di ricerca che si pone in alternativa a quella effettuata in base alla tipologia di messa in scena. Si prefigura, fin dal principio, una differenza tra il concetto di allestimento e quello di serata.

Esistono dei casi in cui i titoli di allestimento e serata coincidono. Ciò accade quando il titolo della composizione dà anche il nome alla serata; si tratta, comunque, di circostanze che si verificano tipicamente in occasione delle opere e dei balletti.

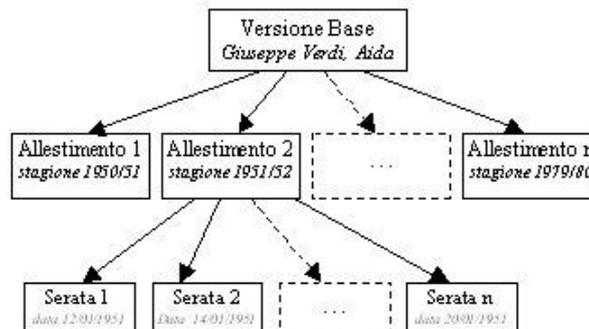


Figura 3.1: Struttura gerarchica tra la Versione Base, l'Allestimento e la Serata.

Una situazione particolare ha luogo quando una performance può essere suddivisa in un determinato numero di blocchi, ognuno completo e indipendente dagli altri; in questo caso, si assume che ognuno di questi appartenga ad una differente versione base. Tutta l'informazione contenuta in quest'ultima è concettualmente comune anche all'allestimento e alla serata: il titolo della composizione, i suoi autori, l'edizione, la data della prima rappresentazione, l'elenco dei personaggi, l'organico base e via dicendo. Alcune di tali informazioni vengono omesse ai livelli inferiori per evitare ridondanza. È comunque sempre presente il collegamento alla cosiddetta versione base. Altre informazioni sono invece maggiormente dettagliate nell'allestimento e nella serata: è il caso dei personaggi, i quali vengono affiancati dai nomi degli interpreti.

Ovviamente, molte informazioni testuali specifiche sono proprie dell'allestimento e non della versione base: ad esempio la stagione, il luogo e il numero di rappresentazioni.

Infine, la serata contiene qualcosa di ancor più dettagliato: la data dell'evento, il numero progressivo della rappresentazione e, come già detto, l'elenco dei partecipanti.

Per quanto riguarda i media presenti all'interno del database, invece, questi necessitano un discorso a parte. Se, da un lato, bozzetti, figurini, costumi, accessori, attrezzi, soggetto, libretto, partiture, discografia e bibliografia so-

no riconducibili al singolo allestimento, se non, addirittura, ad una molteplicità di questi (si pensi al libretto, che, salvo casi particolari, accomuna tutti gli allestimenti di una determinata composizione); dall'altro, audio, fotografie e locandine sono legate ad una singola e specifica serata.

Anche in quest'ultimo caso esistono delle eccezioni di cui tenere conto: ad esempio, il fatto che, talvolta, un'unica locandina descriva più serate ed eventi o che determinate fotografie in realtà non siano legate ad alcuna serata e rientrino, di conseguenza, in una stagione generica.

Contenuti del database

Il primo passo per effettuare la costruzione della base di dati consiste nel catturare tutte le informazioni di un qualche interesse per il proprio obiettivo. Un teatro d'opera è un ambiente informativo molto ricco e complesso che include materiali e documenti come:

- spartiti e rappresentazioni simboliche della musica;
- registrazioni audio;
- registrazioni video;
- volantini, locandine e posters;
- foto;
- bozzetti;
- costumi ed accessori ad essi correlati;
- equipaggiamenti, strumenti e mappe per il palcoscenico;
- altro materiale testuale, come la bibliografia, la discografia, il libretto, le descrizioni brevi e le recensioni di lavori musicali.

Questa lista illustra l'eterogeneità dei dati e dei metadati che il database dovrebbe immagazzinare e gestire.

Digitalizzazione, conversione e limiti

Spesso, la forma originale del materiale preso in esame deve essere convertita. Ciò si verifica per una ragione che di primo acchito potrebbe apparire scontata, ma che in realtà si pone come premessa e base della progettazione del database: solo gli oggetti digitali possono essere immagazzinati all'interno di una base di dati. Alcuni contenuti della lista sopracitata, come per esempio le foto o le registrazioni audio e video, possono essere, fin dal principio, disponibili in formato digitale. Il resto deve essere sottoposto ad

un processo specifico di conversione.

La digitalizzazione è il processo che determina il passaggio dal campo dei valori continui a quello dei valori discreti e viene oggi comunemente intesa nei termini di un passaggio dall'analogico al digitale. Tale conversione, prevista per alcuni dei contenuti a disposizione, prevede, sotto particolari condizioni, una perdita di informazione; è il caso di un'immagine scannerizzata, di audio analogico convertito, di risultati della digitalizzazione video o della scrittura di documenti digitali su un disco fisso (testi e partiture). In alcuni casi, tale perdita è accettabile, in quanto si guadagna in semplicità di rappresentazione o, in altri, non è comunque percepita. In tale contesto, la digitalizzazione consiste praticamente in un processo di trasformazione di un'immagine, di un suono o di un documento in un formato digitale, interpretabile da un computer.

Lo strumento usato in fase di digitalizzazione di materiale cartaceo è generalmente lo scanner, il quale funziona sostanzialmente come una fotocopiatrice, riproducendo sullo schermo dell'elaboratore un'immagine digitale assolutamente analoga a quella dell'esemplare d'origine. Un'immagine digitale altro non è, quindi, che la rappresentazione al computer di un oggetto, la quale dopo essere stata acquisita nella memoria dell'elaboratore, tramite lo scanner (o un altro dispositivo di cattura), può essere in vari modi manipolata dall'operatore.

Infine, ci sono oggetti fisici che possono essere catturati solo da certi punti di vista e un numero limitato di volte. In questo caso, la copia digitale non permette di rendere giustizia e di apprezzare tutte le sfaccettature degli oggetti digitali. Per esempio, se si prendono in esame gli attrezzi da palcoscenico o i costumi, possono essere scattate e immagazzinate notevoli quantità di foto digitali, affinché sia possibile mostrarli, ma l'intera sequenza fotografica non è sufficiente per provvedere ad una visione tridimensionale dell'oggetto originale.

3.1.6 Database multimediali per centri di conservazione e divulgazione musicali

Un ulteriore esempio, relativo all'utilizzo di database multimediali, ma in un ambito differente rispetto al precedente caso, lo si può ricercare all'interno del progetto "La Musica Italiana in Rete" volto a catalogare, digitalizzare e mettere on-line gran parte dei documenti più significativi dell'archivio Storico di Casa Ricordi, riguardanti Giuseppe Verdi e Giacomo Puccini. Il progetto è promosso dal Dipartimento per i Beni archivistici e librari e si pone come obiettivo, attraverso la preziosa opera di conservazione digitale, la salvaguardia di beni quali partiture, libretti, manifesti, foto d'epoca

ed epistolari custoditi nella Biblioteca Nazionale Braidense, rendendo disponibile su Internet tutto il materiale all'intera comunità scientifica, agli appassionati di musica e ai semplici cittadini. Nello specifico, grazie al web, il progetto intende valorizzare una parte importante della cultura nazionale: l'opera e il melodramma. Oltre alle opere più importanti di Verdi e Puccini, a partire dal "Falstaff", andato in scena alla Scala nel 1893, e dalla "Bohème", presentata al Regio di Torino nel 1896, "La musica italiana in Rete" si occuperà infatti di raccogliere anche i risultati delle altre campagne di digitalizzazione sviluppate a Torino per Vivaldi, a Venezia per i fratelli Marcello e a Roma per la musica sacra e profana. Collegato al Servizio Bibliotecario Nazionale, lo speciale archivio musicale entrerà così a far parte del nascente e ambizioso portale Internet Culturale della "Biblioteca digitale italiana" e del Network Turistico Culturale (NTC). Il futuro della cultura musicale italiana è su Internet. Solo così nulla delle più importanti fonti musicali italiane dell'800 e del '900 andrà perduto.

Contenuti del database

L'archivio Storico Ricordi è una delle più importanti raccolte musicali private. Si è formato parallelamente al nascere e crescere di Casa Ricordi, di cui conserva ancora oggi i documenti della sua fondazione, avvenuta nel 1808, e raccoglie, nel corso di quasi due secoli, un patrimonio che ha fatto la storia musicale e teatrale italiana conosciuta a livello mondiale.

La storia dell'archivio ha inizio quando Giovanni Ricordi fonda l'omonima società e comincia a raccogliere e catalogare documenti funzionali al proprio lavoro: manoscritti autografi, libretti, bozzetti, lettere, fotografie, manifesti. Nel corso di quasi due secoli, la raccolta continua ad arricchirsi dando vita ad un inestimabile patrimonio che rappresenta la storia della musica e del teatro italiano. Lo straordinario valore dell'archivio, seppur legato alla musica e all'arte, deriva proprio dall'ampiezza dello scenario dei documenti conservati, i quali offrono uno sguardo d'insieme sulle diverse sfaccettature della cultura, dell'industria ed una straordinaria testimonianza dei momenti più importanti e cruciali della storia italiana degli ultimi secoli.

Attualmente l'archivio comprende 3.593 partiture dal '700 al primo '900 di cui 2.246 autografe, circa 15.000 lettere di musicisti e librettisti, oltre 10.000 bozzetti e figurini, più di 9.000 libretti a stampa e manoscritti, 4.000 foto d'epoca, manifesti liberty firmati dai grandi della grafica. È nell'archivio Storico Ricordi che sono conservati gli originali di 23 delle 28 opere scritte da Verdi, tutte le opere di Puccini (con la sola eccezione di *Rondine*), che si affiancano a moltissime opere di grandi autori di musica classica: Rossini, Bellini, Donizetti, Paganini, Liszt, fino ai più moderni come Respighi, Malipiero, Casella, Pizzetti, Nono, Maderna, Donatoni.

All'intero dell'archivio Storico della Ricordi sono presenti, oltre a quanto già elencato, anche:

- manoscritti musicali;
- frontespizi a stampa;
- copialettere;
- disposizioni sceniche;
- piante del palcoscenico;
- caricature;
- tavole e attrezzi;
- note di sartoria;
- giornali.

A digitalizzazione ultimata, il materiale sarà reso disponibile grazie ad una base di dati multimediale studiata ad hoc. Questa utilizza i BFILE per l'archiviazione dei contenuti in modo tale da delegare la loro gestione al sistema operativo, in quanto ogni contenuto multimediale viene memorizzato nel filesystem come un singolo file e la base di dati non deve far altro che memorizzare il loro indirizzo fisico. In questo caso specifico, è preferibile la scelta dei BFILE rispetto ai BLOB, in quanto i contenuti digitali del database dell'archivio Storico della Ricordi sono di grosse dimensioni e l'utilizzo dei BLOB appesantirebbe in maniera eccessiva la base di dati e, di conseguenza, la sua gestione.

3.1.7 Database multimediali per partiture manoscritte

Dopo aver presentato i database multimediali e visto il loro utilizzo all'interno dei teatri d'opera e dei centri di conservazione e divulgazione musicali, in questo paragrafo, si andrà ad analizzare il processo di digitalizzazione delle partiture manoscritte, le quali rappresentano uno degli elementi fondamentali di tali istituzioni. Una partitura è l'insieme di molti righi musicali condensati in una sola parte complessiva, ad uso del compositore o del direttore d'orchestra.

Nel linguaggio comune, si utilizzano spesso i termini partitura o spartito per indicare il mezzo di rappresentazione scritta della musica. In realtà i termini parte, spartito e partitura hanno significati ben precisi nel linguaggio musicale.

Per “parte”, generalmente, si intende la notazione musicale riservata ad uno strumento solista, come un pianoforte solo, oppure uno strumento in formazione cameristica, come un violino in duo con un pianoforte.

Mozart
Eine Kleine Nachtmusik, K. 525
Violin I

Allegro

Figura 3.2: Esempio di parte per violino.

Per “partitura” si intende la sovrapposizione di molti rigli musicali, eseguiti da diversi esecutori, in una sola parte complessiva, ad uso del compositore o del direttore d’orchestra.

Lento sostenuto (♩. = 50)

dolce

sempre legato

Figura 3.3: Esempio di partitura.

Per “spartito” si intende, invece, la riduzione per voci e pianoforte di una composizione originariamente concepita per voci e orchestra.

Vois sur ton chemin
(Christophe Barratier / Bruno Coulais)

♩ = 96

Soprano

Alto

Piano

mp

Vois sur ton che-min Ga-mine ou - si - es - e - ga - nie

Figura 3.4: Esempio di spartito.

Il sistema di notazione musicale normalmente in uso nella musica occidentale per strumenti musicali si serve di un pentagramma su cui vengono segnate le note musicali ed altri simboli come le chiavi; è un sistema di notazione comunemente stampato su carta. Questo non è del tutto corretto per fondi antichi presenti in archivi musicali, i quali sono scritti a mano. Affinchè materiali così datati e, per questa ragione, delicati possano essere conservati e tramandati alle generazioni future, è necessario convertire tali opere in formato digitale.

È possibile ritenere archiviata l'epoca in cui digitalizzare significava produrre una certa quantità di scansioni da collocare su una serie di CD-ROM, difficili da maneggiare e destinati a permanere a lungo sugli scaffali. Gli obiettivi di queste operazioni variavano dalla necessità di ridurre l'uso degli originali analogici a quella di disporre di copie da cui trarre riproduzioni o pubblicazioni.

Prima di intraprendere tali progetti di digitalizzazione, occorre chiarire chi sono i destinatari e analizzarne la tipologia. Digitalizzare una fonte può avere un duplice significato: la digitalizzazione come riproduzione di una copia esatta dell'originale (facsimile) e quindi la creazione di un file di immagine; la digitalizzazione come operazione di riconoscimento ottico della fonte di partenza e quindi conversione della fonte cartacea in set di caratteri ASCII, cioè file di testo.

La digitalizzazione si configura, quindi, come un processo di memorizzazione di un documento attraverso la sua conversione in una sequenza di caratteri numerici binari, ciascuno dei quali corrisponde ad una porzione del documento originale.

Lo strumento usato nella fase di digitalizzazione è, come già detto, lo scanner, il quale funziona come una fotocopiatrice, riproducendo sullo schermo dell'elaboratore un'immagine digitale assolutamente analoga a quella dell'esemplare d'origine. L'immagine digitale altro non è, quindi, che la rappresentazione al computer della partitura, la quale, dopo essere stata acquisita nella memoria dell'elaboratore tramite lo scanner (o tramite un altro dispositivo di cattura), può essere in vari modi manipolata dall'operatore.

Un'immagine digitale ha le seguenti caratteristiche:

- *dimensione*: l'unità base di misura della grandezza di un'immagine digitale è il pixel, il quale rappresenta il singolo punto con associate tutte le informazioni, quali colore, luminosità, ecc.. Molteplici pixel vengono giustapposti l'un l'altro per generare quella che i nostri occhi vedono come un'immagine, ma che, in realtà, è soltanto un insieme di

quadratini unitari affiancati gli uni agli altri; la dimensione di un'immagine è solitamente espressa con la coppia formata da larghezza e altezza le quali descrivono il minimo rettangolo inscrivente la figura;

- *risoluzione (dpi)*: qualunque sia il supporto di visualizzazione, le immagini vengono rappresentate da pixel, più o meno densi a seconda del supporto. Il numero di pixel per unità di spazio è detta “risoluzione” ed è una misura di densità in una dimensione. La dimensione in pixel dell'immagine è legata alla dimensione dell'immagine e alla sua risoluzione dalla semplice relazione: dimensione in pixel = grandezza fisica x risoluzione;
- *profondità di colore (bit)*: ad ogni pixel è sicuramente collegato il suo valore in colore che, solitamente, è descritto attraverso i sistemi RGB o CMYK, i quali utilizzano dei colori base che vengono poi miscelati proporzionalmente all'informazione associata al singolo pixel. Tale dato è descritto con una stringa di bit che è più o meno lunga a seconda del modello utilizzato e/o del numero di colori presenti nell'immagine. Ovviamente, maggiore è la profondità di colore associata ad ogni pixel, maggiore sarà la definizione dell'immagine e, conseguentemente, anche la dimensione che essa occupa;
- *compressione*: visti i problemi sopra citati, riguardanti la dimensione, sono state create differenti codifiche in grado di eliminare o ridurre la ridondanza dell'informazione che descrive l'immagine, pur mantenendo una sufficiente aderenza visiva all'originale. Le codifiche di compressione si dividono in: codifiche con perdita (lossy) o codifiche senza perdita (lossless). Le prime, quali l'universale JPEG, eliminano le irrilevanze percettive (noi vediamo solo alcune caratteristiche delle immagini), ossia inserisce delle approssimazioni laddove esse risultano meno percepibili. In particolare l'occhio umano è meno sensibile alle variazioni di colore che alle variazioni di luminosità. Dall'immagine compressa non è più possibile ricostruire l'immagine originale in tutto e per tutto. Esse implicano quindi una perdita anche in materia di qualità dell'immagine. Le seconde, quali per esempio il TIFF, sono ottenute mediante l'eliminazione delle ridondanze intrinseche del segnale (alcune immagini hanno caratteristiche ripetitive) e non implicano perdite definitive. La qualità viene salvaguardata ed è possibile riottenere l'immagine di partenza. La scelta di una tecnica rispetto ad un'altra viene fatta in base alle necessità.

JPEG è un formato standard di compressione di tipo lossy dei file grafici in formato bitmap (tale grafica richiede grandi quantità di memoria e le immagini che lo utilizzano risultano quindi molto ingombranti). JPEG è la sigla di Joint Photographic Experts Group, il nome del comitato che ha

scritto le specifiche. Attualmente JPEG è lo standard internazionale di compressione delle immagini fotografiche più utilizzato, in quanto comprime in maniera ottimale immagini che possiedono variazioni graduali (sfumature) di colore, ma non quelle con aree di colore omogeneo, per le quali è preferibile utilizzare altri formati, come il GIF o PNG. Parte dell'informazione contenuta nei pixel viene ignorata in base a opportuni criteri percettivi e, in fase di decompressione, le informazioni mancanti vengono ricalcolate in modo approssimato. Le informazioni eliminate sono quelle meno percettibili all'occhio umano, che risulta essere meno sensibile alle piccole variazioni di colore che alle piccole variazioni di luminosità. Le estensioni più comuni utilizzate per questo formato sono .jpg o .jpeg.. Esso tratta solo immagini statiche, ma esiste un altro standard correlato, MPEG, per i filmati.

Il formato di compressione (lossless) TIFF è stato sviluppato nel 1986. Tag Image File Format è il più usato, più flessibile ed affidabile metodo per memorizzare immagini bitmap in bianco e nero, a scala di grigio, a scala di colore, a colori RGB, CMYK, YCbCr, Lab. Un file TIFF può essere di ogni dimensione (in pixel) e di ogni profondità di bit. Oltre ai dati dei pixel TIFF può contenere qualunque metainformazione in locazioni di memoria chiamate tag. Le più comuni sono la risoluzione, la compressione, il tracciato di scontorno, il modello di colore. Sebbene sia considerato un formato standard, alcune applicazioni inseriscono dei tag proprietari che talvolta impediscono ai file di essere aperti da altre applicazioni. Le immagini possono essere memorizzate, oltre come linee di scansione, anche in riquadri: questo permette di avere un rapido accesso ad immagini di grosse dimensioni.

Non esistendo uno standard nel campo della compressione delle immagini digitali si necessita di continui aggiornamenti, per adeguarsi alle progressive evoluzioni tecnologiche nel settore.

Lo svantaggio più pericoloso nel campo della compressione del materiale digitale è l'inevitabile perdita di informazione che tale operazione comporta: una volta compressi i dati, all'atto della decompressione non tutti i file fondamentali al recupero dell'immagine d'origine vengono ripristinati; la perdita di dati informativi è uno degli inconvenienti che più di frequente limita il lavoro dello studioso e dovrebbe essere dunque sempre ridotta al minimo. Di fronte a questi svantaggi esistono, però, innegabili vantaggi: le immagini digitali non decadono nel tempo, sono infatti destinate a durare a lungo, nella copia, che non comporta perdita di qualità, e nell'uso, possono essere cioè consultate molteplici volte senza subire alcun danneggiamento. Nonostante la partitura digitalizzata, in quanto immagine, abbia un'interpretazione nella realtà, il file corrispondente rappresenta una serie di dati e non un'informazione. Affinchè l'immagine porti con se un carico informativo, si rende necessario l'utilizzo di metadati che la descrivano.

I diversi oggetti digitalizzati sono organizzati in una struttura gerarchica a livelli, così formata:

- *collezione digitale*: un insieme di oggetti digitali omogenei per contenuto, per titolarità o altre caratteristiche salienti (ad esempio: il fondo musicale di Antonio Vivaldi della Biblioteca Nazionale Universitaria di Torino, ma anche le collezioni botaniche seicentesche delle biblioteche di Roma). Una collezione digitale può identificarsi con un aggregato o può essere composta di più aggregati;
- *aggregato*: un insieme di oggetti digitali omogenei per tipologia del contenuto e per titolarità (ad esempio le partiture autografe di Vivaldi della Biblioteca Nazionale Universitaria di Torino);
- *oggetto primario*: un oggetto digitale definibile come un intero coerente, in genere corrispondente ad un'unità fisica (ad esempio: un libro, un disco, o una singola partitura autografa di Vivaldi);
- *oggetto intermedio*: una particolare vista o formato dell'oggetto primario (ad esempio: un manoscritto in formato immagine TIFF o in formato testo codificato in XML), oppure una parte componente dell'oggetto primario (ad esempio: la digitalizzazione in real audio di due facciate di un disco a 78 giri ciascuna con vari brani, la digitalizzazione delle etichette);
- *entità digitale (terminal object)*: il singolo file recante un'unità elementare di contenuto digitale (ad esempio: una singola pagina di una partitura autografa di Vivaldi, il frontespizio di un libro).

Si assume che ciascun oggetto digitale di un determinato livello abbia come attributi:

- i metadati specifici per gli oggetti digitali ascrivibili a quel livello;
- i metadati specifici per gli oggetti digitali ascrivibili al livello immediatamente superiore.

Le partiture manoscritte sono caratterizzate dai seguenti metadati:

- tipo di partitura (partitura per voce, per accompagnamento...);
- tipo di testo presente (poesie, drammi, racconti...);
- lingua del testo;
- lingua originale del testo;

- durata;
- forma della composizione;
- organico;
- data creazione;
- data e luogo di pubblicazione;
- tecnica di stampa;
- misure dimensionali;
- segnatrice;
- note autore e loro lingua.

La scansione della partitura è caratterizzata da metadati che descrivono le specifiche tecniche dell'immagine digitalizzata per esempio l'unità di misura del campionamento, il numero di bit per campione, algoritmo di compressione e formato dell'immagine.

3.1.8 Optical Music Recognition

Il riconoscimento automatico di partiture (OMR) è un processo di acquisizione dell'informazione musicale stampata o manoscritta. Tramite la corretta interpretazione dell'immagine binaria di una partitura, esso codifica tale informazione in un linguaggio di rappresentazione musicale. Il brano, nella forma digitale così ottenuta, può essere utilizzato come input per numerose applicazioni on-line e off-line. In un sistema on-line, la macchina analizza lo spartito e produce un risultato in modo pressochè istantaneo. Data la natura del sistema, l'analisi potrebbe non essere portata a termine in breve tempo e, quindi, la generazione dell'output verrebbe anticipata. In questo caso, la macchina dovrà processare piccoli brani di partitura per generarne l'output immediatamente. Con questo approccio, errori commessi in precedenza non possono essere corretti successivamente. Nei sistemi off-line, viene prima acquisita un'immagine della partitura utilizzando uno scanner o una macchina fotografica digitale; essa viene poi analizzata dal computer e codificata in un formato binario utilizzabile come input per pacchetti software di manipolazione, stampa e riproduzione. Poichè un sistema off-line può analizzare l'intero spartito prima di generarne l'output, le informazioni globali sulla partitura possono essere utilizzate per migliorare l'accuratezza del riconoscimento. Ad esempio, può essere sviluppato un controllo sofisticato della semantica per correggere gli errori che si sospetta il precedente passo di riconoscimento abbia commesso.

Optical Musical Recognition e Optical Character Recognition

Un sistema OMR riconosce i simboli su uno spartito e produce in output il risultato in un formato binario. Questa operazione è analoga a quanto compiuto dai sistemi OCR i quali si occupano del riconoscimento del testo. Nonostante questa apparente somiglianza, sono molti gli elementi che non permettono di adottare le tecniche sviluppate per i sistemi OCR nel campo del riconoscimento ottico di partiture:

- il testo è monodimensionale, mentre un brano musicale è bidimensionale: nel primo caso ciò che ha rilevanza è la successione delle lettere in senso orizzontale, mentre sul pentagramma anche la posizione verticale fornisce un'informazione fondamentale come l'altezza della nota;
- i caratteri che compongono una parola e quindi, più in generale, una frase o un brano, sono disgiunti, mentre i simboli musicali sono sovrapposti alle linee del pentagramma e, a volte, ad altri simboli (ad esempio: nella rappresentazione di accordi);
- i simboli musicali possono essere di dimensioni molto differenti tra loro e alcuni legano il loro significato alla grandezza (ad esempio: l'appoggiatura);
- contrariamente a quanto accade nei testi, i simboli musicali sono formati da componenti che possono essere combinate in modi diversi (ad esempio: una gamba può essere unita a più teste e può essere unita, con una trave, ad un'altra gamba);
- uno stesso simbolo musicale può apparire in forme differenti (ad esempio: la lunghezza delle legature, delle travi e delle gambe dipende dal contesto);
- non c'è in ambito musicale un alfabeto standard: il numero di simboli è molto ampio e in continua evoluzione, infatti i compositori inventano nuove notazioni quando è necessario e per particolari strumenti si utilizzano notazioni specializzate.

Tipologie di problemi

I fattori che influenzano la realizzazione di un sistema OMR sono di varia natura e complessità; la loro incidenza è andata mutando con l'evoluzione tecnologica del settore. Ad esempio, rispetto ai tempi in cui la ricerca sul riconoscimento automatico di partiture ha preso il via, fine anni '60 e inizio '70, si possono considerare superate le limitazioni hardware costituite dalla lentezza dei processori, dalla poca quantità di RAM disponibile sui sistemi di elaborazione e dalle capacità inadeguate delle memorie di massa. Invece, fattori tuttora da considerare sono:

1. la qualità grafica del materiale digitalizzato;
2. la complessità del brano musicale;
3. il formato in cui si vuole convertire la partitura;
4. l'efficienza dei programmi applicativi usati per completare la riproduzione musicale.

1. Qualità grafica del materiale digitalizzato

I problemi legati alla qualità grafica della partitura digitalizzata riguardano la superficie visiva, il riconoscimento dell'oggetto e la rappresentazione musicale.

Problemi nella superficie visiva

I problemi nella superficie visiva sono per lo più imperfezioni nella stampa come, ad esempio:

- la rotazione e la curvatura dei pentagrammi, per cui le linee non sono perfettamente parallele al bordo della pagina;
- le variazioni di spessore delle linee del pentagramma;
- la posizione scorretta dei simboli (ad esempio: una nota che copre sia uno spazio che una riga).

Per quanto possano apparire questioni trascurabili per l'occhio umano, queste irregolarità possono essere destabilizzanti per un software di riconoscimento che accetta piccole variazioni, ma solo all'interno di un insieme ben limitato. Inoltre queste imperfezioni sono maggiormente presenti quando si opera con partiture non recenti, come accade frequentemente nel campo della ricerca per motivi di copyright.

Problemi nel riconoscimento dell'oggetto

Essi variano in base all'approccio adottato. I principali, comuni a molti approcci, sono:

- *le dimensioni*: a causa delle variazioni di dimensione tra oggetti uguali (ad esempio: La lunghezza delle gambe e la larghezza delle travi nelle vecchie partiture o la diversità fra un'appoggiatura e una nota comune);
- *le forme e la rappresentazione*: costituiscono un esempio tipico di inconsistenza le chiavi e le pause da un quarto che hanno ampie variazioni nella loro rappresentazione grafica;

- *la contiguità e la sovrapposizione*: ad esempio le legature che possono creare problemi tollerabili se si limitano ad attraversare una gamba, ma che se toccano una nota o attraversano un segno di dinamica creano oggetti apparenti che non vengono ritrovati nell'insieme dei simboli grafici;
- *la disgiunzione*: i simboli formati da componenti disgiunte, come la chiave di basso, richiedono l'utilizzo di un meccanismo particolare per affrontare il loro riconoscimento.

Problemi di rappresentazione musicale

Lo studio riguardante la rappresentazione della musica occidentale in notazione musicale tradizionale non ha portato all'esaurimento o alla descrizione sistematica di tutte le anomalie come, ad esempio, quando più oggetti (note di diversa durata) condividono parti in comune (la gamba) in modo improbabile. Tali esempi interferiscono con un approccio grammaticale al riconoscimento degli oggetti, poichè, nella sua rappresentazione grafica, la logica sottostante la notazione è circondata da ogni parte da eccezioni di natura imprevedibile. In più, programmi di riconoscimento possono essere confusi da oggetti che non si intende acquisire.

2. Complessità del brano musicale

I primi prototipi di programmi erano progettati per lavorare solo con parti singole su un pentagramma. I programmi più recenti si concentrano su musica per insiemi di strumenti non troppo ampi. Non si conoscono programmi in grado di acquisire il testo presente nella musica vocale e sono altrettanto rari quelli che si occupano di acquisire altri elementi testuali come ad esempio le indicazioni sul tempo. Gli utenti di software per il riconoscimento devono aspettarsi di dover supportare il materiale acquisito automaticamente con informazione addizionale, se vogliono creare materiale completo.

3. Formato di conversione dell'immagine digitalizzata

I programmi di scansione intelligente odierni di solito riproducono completamente l'immagine originale. Essi acquisiscono un numero selezionato di attributi dell'informazione musicale. Ad esempio, i programmi per la riproduzione sonora cercheranno di acquisire l'altezza e la durata, mentre, in base al repertorio scelto, i programmi per la stampa potranno aver bisogno di acquisire un maggior numero di elementi, come le legature, le gambe, le travi, le dinamiche e gli abbellimenti.

4. Efficienza dei programmi applicativi usati per completare la riproduzione musicale

Il tempo richiesto per correggere gli errori di scansione dipende non solo dal livello di accuratezza del materiale acquisito automaticamente, ma anche dalla natura ed efficienza dei programmi applicativi in cui il materiale viene letto, tralasciando la facilità di utilizzo da parte dell'utente. Mentre il supporto per la notazione in output è intrinsecamente più esigente che per il suono, molti programmi per la produzione di una stampa del materiale digitalizzato supportano solo un numero finito di elementi addizionali. Il numero di attributi supportati aumenta proporzionalmente alla crescita di competenza generale del programma.

3.2 Progettazione di un database in ambito musicale

La funzionalità di un database dipende, in modo essenziale, dalla sua progettazione: la corretta individuazione dei suoi scopi e quindi delle tabelle, da definire attraverso i loro campi e le relazioni che le legano; permette poi un'estrazione dei dati più veloce e, in generale, una gestione più efficiente di questi.

La progettazione di una base di dati costituisce una delle componenti del processo di sviluppo di un sistema informativo complesso e necessita un inquadramento in un contesto più ampio: quello del ciclo di vita dei sistemi informativi musicali.

L'obiettivo della progettazione è quello di produrre i seguenti risultati:

- uno schema di dati;
- un insieme di sottoschemi di applicazione;
- un insieme di (specifiche di) programmi applicativi.

È possibile schematizzare le varie fasi per la progettazione di un database come segue:

- raccolta ed analisi dei requisiti;
- progettazione concettuale;
- progettazione logica;
- progettazione fisica.

Ogni fase si basa su un modello che permette di generare una rappresentazione formale della base di dati ad un dato livello di astrazione. L'utilizzo del modello appropriato permette di verificare le caratteristiche del database.

3.2.1 Raccolta ed analisi dei requisiti

Per raccolta dei requisiti, si intende la completa individuazione dei problemi che l'applicazione da realizzare deve risolvere e l'insieme delle caratteristiche statiche o dinamiche che tale applicazione dovrà avere. I requisiti vengono, inizialmente, raccolti in specifiche espresse, generalmente, in linguaggio naturale e, per questo motivo, spesso ambigue e disorganizzate.

L'analisi dei requisiti consiste, invece, nel chiarimento e nell'organizzazione di tali specifiche. Si tratta, ovviamente, di attività fortemente interconnesse: l'analisi inizia con i primi requisiti ottenuti per poi procedere di pari passo con attività di raccolta, le quali, in molti casi, sono suggerite dalla stessa analisi.

Negli ambiti musicali precedentemente trattati, queste fasi sono chiaramente specifiche per le diverse istituzioni. Ad esempio, nel caso del database progettato per il Teatro alla Scala erano già disponibili una serie di archivi in parte digitalizzati; inoltre alcuni dettagli implementativi sono stati decisi dall'istituzione. Decisioni prese a seguito di una serie di colloqui con i responsabili dei vari archivi del Teatro. Un procedimento simile a questo è stato adottato per l'analisi dei requisiti per la progettazione del database dell'Archivio Storico Ricordi. Il materiale analogico di tale archivio, quantitativamente superiore e più eterogeneo del precedente, è attualmente depositato presso l'Accademia di Brera. A seguito dei colloqui con la responsabile dell'archivio, sono state definite linee guida per il processo di digitalizzazione e di sviluppo del database, tenendo conto delle caratteristiche del materiale trattato. In particolare, sono state fornite indicazioni derivate dal confronto tra il sistema d'archiviazione utilizzato fino a quel momento e le specifiche di funzionamento del database desiderato.

3.2.2 Progettazione concettuale

Lo scopo di questa fase è quello di rappresentare le specifiche informali della realtà di interesse, nei termini di una descrizione formale e completa, la quale sia indipendente dai criteri di rappresentazione utilizzati nei sistemi di gestione di basi di dati.

Il prodotto di questa fase viene chiamato schema concettuale e fa riferimento a un modello concettuale dei dati. Il progettista ha il compito di rappresentare il contenuto informativo del database senza preoccuparsi ne delle modalità con le quali queste informazioni verranno codificate in un sistema reale ne dell'efficienza dei programmi che faranno uso di queste informazioni.

La progettazione concettuale di una base di dati consiste nella costruzio-

ne di uno schema Entità Relazione (Modello E-R), in grado di descrivere al meglio le specifiche sui dati di un'applicazione. Anche nel caso questa non sia particolarmente complessa, lo schema che si ottiene può contenere molti concetti correlati in maniera piuttosto complicata.

Concetti fondamentali di tale schema sono:

- *entità*: insieme di oggetti della realtà che possiedono caratteristiche comuni;
- *associazioni*: legami logici tra entità;
- *attributi*: proprietà elementari possedute da un'entità o da un'associazione.

Relativamente alle entità e alle associazioni esistono anche:

- *istanze di entità*: oggetti appartenenti ad una certa entità;
- *istanze di associazione*: combinazioni delle istanze delle entità che prendono parte ad una relazione.

Nel modello E-R esistono due vincoli di integrità:

- *Implicito*: automaticamente verificato dal sistema e dipendente dalla semantica stessa dei costrutti del modello:
 - ogni elemento di associazione deve riferirsi ad elementi di entità;
 - elementi diversi della stessa associazione devono riferirsi a differenti combinazioni di elementi delle entità partecipanti all'associazione.
- *Esplicito*: definito chiaramente da chi progetta lo schema ER sulla base della conoscenza della realtà che si sta modellando:
 - vincoli di cardinalità (per associazioni e attributi): definiscono il massimo (o minimo) numero di istanze della relazione cui partecipa un'istanza dell'entità;
 - vincoli di identificazione (o chiave): specifica un attributo o insieme di attributi che identificano univocamente le istanze dell'entità.

Un caso esemplare di diagramma Entità - Relazione è quello disegnato su misura per il database del Teatro alla Scala di Milano. È possibile osservare, all'interno del modello, le relazioni che si vanno ad instaurare tra i contenuti digitali appartenenti alla lista di materiali presenti nell'archivio e i concetti precedentemente trattati (3.1.5) di versione base, allestimento e serata.

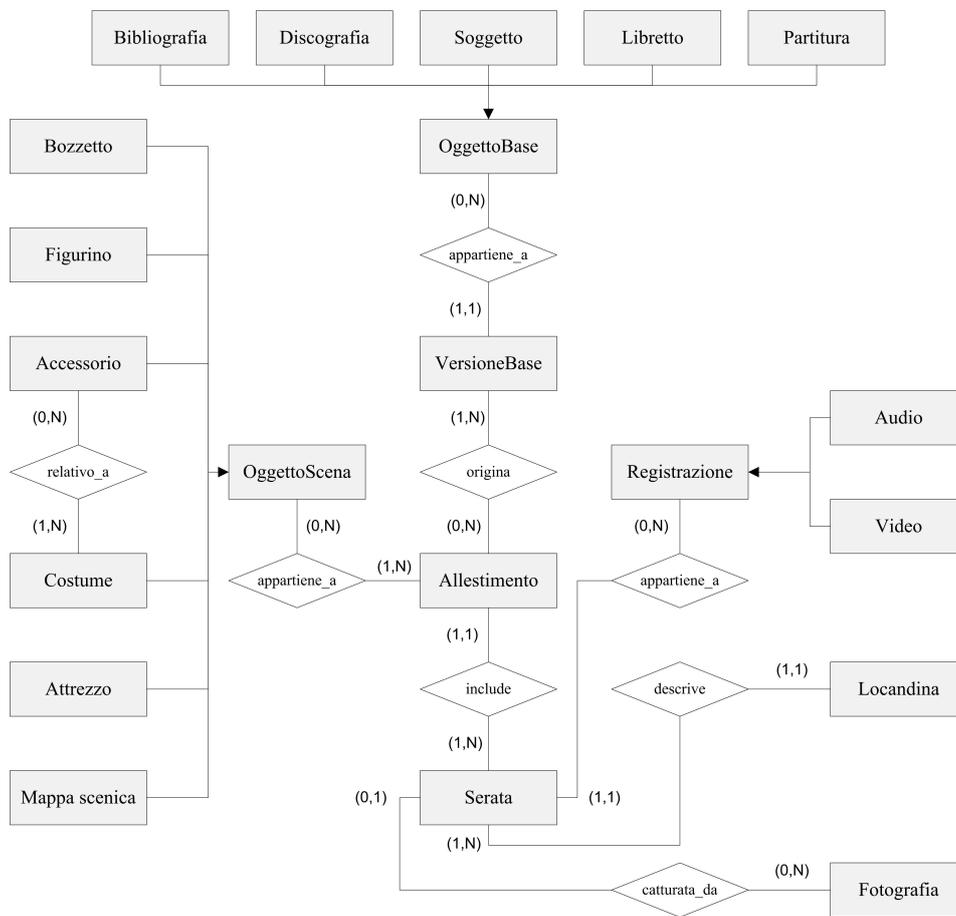


Figura 3.5: Il Diagramma Entità - Relazione per il Database Multimediale.

La disposizione degli elementi nello schema è evidente: la struttura vede la presenza dei tre concetti chiave di versione base, allestimento e serata, considerati tutti con le loro relazioni. Lo spartito musicale, così come un documento di testo (un libretto, per esempio) e il dato ad esso relativo (autore, titolo, data di pubblicazione...), può essere associato alla versione base. Ogni sotto entità presenta i propri e specifici metadati. Queste specifiche verranno considerate nell'implementazione della base di dati, ma, da un punto di vista teorico, esse possono essere tutte raggruppate sotto il concetto di Oggetti della Versione Base.

Possono verificarsi, anche, alcuni problemi di minore entità, quali, per esempio, la presenza di diverse versioni autografe di uno stesso spartito o differenti titoli di uno stesso lavoro. Fortunatamente, questi problemi possono essere facilmente risolti grazie all'instaurazione di relazioni uno a molti anziché uno a uno.

Bozzetti, costumi, mappe e equipaggiamenti per il palcoscenico sono, naturalmente, in relazione non alla versione base, bensì all'allestimento. La scelta dei materiali per il palcoscenico, infatti, dipende dalle direttive del Direttore, del costumista, dello scenografo e così via. Anche se tutti gli oggetti presenti nella lista, precedentemente menzionata, sono, generalmente, in relazione con un particolare palcoscenico, non si possono ignorare gli eventi comuni come le contaminazioni tra differenti lavori, le cui produzioni impiegano gli stessi equipaggiamenti. Questa è la ragione per cui nella 3.5, la teorica cardinalità uno a molti tra le entità Allestimento e Oggetto dell'Allestimento diventa una relazione molti a molti. In altre parole, il database, così strutturato, supporta il fatto che per una particolare messa in scena possono essere utilizzati diversi oggetti, i quali, a loro volta, possono essere presenti in differenti produzioni.

Considerando le sotto entità, raggruppate sotto il concetto di Oggetti dell'Allestimento, è possibile notare che due di loro presentano una particolare relazione: i costumi e gli accessori. Questi ultimi, includono, per esempio, anche la gioielleria e le scarpe. La cardinalità della relazione è giustificata dal fatto che alcuni costumi non hanno nessun accessorio o, in alcuni casi, questi sono immagazzinati e classificati insieme con i loro costumi; al contrario, sarebbe auspicabile che ciascun accessorio venga relazionato con il suo costume.

Esaminando la 3.5, un'asimmetria diviene da subito evidente: la Versione Base è messa in relazione con gli Oggetti della Versione Base, così come l'Allestimento lo è con i suoi oggetti, ma la Serata non è messa in relazione con un'unica entità Oggetti della Serata. La ragione risiede nelle differenti cardinalità con le quali si ha che fare nella relazione che coinvolge quest'ultima. Infatti, le registrazioni audio e video sono necessariamente messe legate ad una ed una sola serata; al contrario, un volantino è, generalmente, messo in relazione con un singolo spettacolo, ma può anche descrivere molte serate che condividono lo stesso programma musicale e gli stessi interpreti. Anche le foto possono anche essere non in relazione con una serata: questo è il caso di scatti effettuati durante eventi non musicali o ritraenti momenti privati della vita dell'artista.

A database ultimato, è possibile osservare i numerosi vantaggi di cui gode un teatro d'opera avvalendosi di un efficace, quanto efficiente, base di dati. Grazie a tecnologie informatiche, i contenuti di quest'ultimo possono essere facilmente trasmessi, sia nella dimensione spaziale che in quella temporale. La versione digitale del materiale originale può essere maneggiata, servendosi di eventuali accorgimenti, di generazione in generazione. In un certo senso, non solo i dati e i metadati, ma anche gli oggetti fisici, possono essere

trasmessi dopo una campagna di digitalizzazione ad hoc. Le tecnologie di rete, inoltre, permettono un accesso globale, ai contenuti presenti all'interno della base di dati, ad ogni utente autorizzato. Le pubblicazioni on-line di contenuti multimediali richiedono un'interfaccia ben strutturata contenente efficaci strumenti per la navigazione e appropriate forme di aggregazione dei dati. Come risultato di questo sforzo, i ricercatori, gli interpreti e gli appassionati potrebbero avere, quindi, accesso ad un importante patrimonio culturale.

3.2.3 Progettazione logica

Questa fase consiste nella traduzione dello schema concettuale, definito nella fase precedente, nel modello di rappresentazione dei dati, adottato dal sistema di gestione della base di dati a disposizione. Il prodotto di questa fase viene denominato schema logico e fa riferimento ad un modello logico dei dati. Le scelte progettuali si basano su criteri di ottimizzazione di operazioni da effettuare su di essi. Si fa comunemente uso anche di tecniche formali di verifica della qualità dello schema logico ottenuto. Nel caso del modello relazionale dei dati, la tecnica della normalizzazione aiuta tale verifica permettendo di descrivere la struttura logica dei dati assicurandone quindi l'efficienza, garantendo che si possa recuperare l'originale collezione di relazioni senza perdita di informazioni e permettendo di riconoscere le proprietà non desiderabili al fine di trasformarle come si desidera.

Sottofasi della progettazione sono: la *ristrutturazione* dello schema concettuale, la *traduzione* canonica e la valutazione ed eventuale modifica dello schema. Gli aspetti da considerare sono invece: integrità e consistenza, sicurezza ed efficienza.

La ristrutturazione consiste, come prima cosa, nell'eliminazione degli identificatori esterni o misti di un'entità dati da un'ulteriore entità collegata ad essa attraverso un'associazione. Il secondo passo è l'eliminazione di attributi composti e multivalore. L'ultimo passo è l'eliminazione della gerarchia di generalizzazione che può avvenire in tre diverse modalità:

- *eliminazione delle entità figlie*, le quali vengono inserite nella entità padre. Viene aggiunto anche un ulteriore attributo che permette di identificare quale entità figlia rappresenta una determinata istanza dell'entità padre;
- *eliminazione dell'entità padre*: la quale elimina il padre, inserendo in tutte le entità figlie gli attributi dell'entità padre;
- *sostituzione della generalizzazione con associazioni*: le entità coinvolte non vengono modificate, mentre la gerarchia di generalizzazione viene

modellata con n associazioni uno a uno, ognuna delle quali lega l'entità padre con una diversa entità figlia.

La scelta della soluzione da adottare dipende dal tipo di interrogazione che sarà necessario fare all'interno dell'entità.

Il risultato è uno schema E-R ristrutturato.

Dato lo schema E-R ristrutturato è possibile tradurlo in un equivalente schema relazionale. La traduzione implica:

- *traduzione dell'entità*: si genera una relazione per ogni entità presente nello schema E-R. La relazione avrà un attributo per ogni attributo dell'entità. La chiave della relazione sarà costituita dagli attributi corrispondenti all'identificatore dell'entità;
- *traduzione delle associazioni*: effettuata mediante un insieme di regole secondo due diverse modalità: o modellandola con una nuova relazione o aggiungendo nuovi attributi alle relazioni corrispondenti alle entità partecipanti all'associazione.

Ci sono due alternative: l'associazione può essere rappresentata inserendo opportuni attributi (chiavi esterne) in una delle relazioni rappresentanti le entità partecipanti; altrimenti, l'associazione stessa viene modellata con una relazione.

I vincoli di cardinalità influiscono sulla traduzione delle relazioni nei seguenti modi:

- *associazione binaria uno a uno*: si aggiunge la chiave di una delle due entità e un attributo per ogni attributo dell'associazione, alla relazione che rappresenta l'altra entità coinvolta. Se esiste una entità la cui partecipazione è obbligatoria, allora gli attributi vengono aggiunti nella relazione che la rappresenta. Altrimenti (tutte due obbligatorie od opzionali) la scelta ricade su una qualsiasi;
- *associazione binaria uno a molti*: si aggiunge alla relazione che rappresenta l'entità dal lato Uno, sia la chiave della relazione rappresentante il lato molti, sia un attributo per ogni attributo della relazione;
- *associazione binaria molti a molti*: per ogni relazione n -aria si genera una relazione corrispondente che contiene tutti gli attributi della associazione, più le chiavi che modellano le entità partecipanti all'associazione. La chiave della relazione è costituita dalle chiavi delle relazioni corrispondenti all'entità che partecipano all'associazione.

Un caso esemplare di schema logico è quello mostrato qui sopra. Il modello fornisce, per definizione, una serie di strutture per descrivere la realtà presa in esame. In questo caso specifico, quella dell'Archivio Storico Ricordi custodito in un centro di conservazione e divulgazione musicale, quale la Biblioteca Braidense di Milano. Le strutture, o costrutti, sono utilizzati per definire gli schemi che descrivono l'organizzazione e la struttura delle occorrenze dei dati.

I costrutti principali, come già visto in precedenza, sono:

- le entità, in questo schema esplicitate dai nomi delle tabelle;
- le relazioni, le quali rappresentano legami logici, significativi per l'applicazione, tra due o più entità e sono qui esplicitate attraverso tabelle di join;
- gli attributi, i quali descrivono proprietà elementari di entità o relazioni di interesse ai fini dell'applicazione;
- le generalizzazioni, le quali rappresentano legami logici tra entità padre e una o più entità figlie.

Prendendo in esame il materiale a disposizione e raggruppandolo in un unico concetto generale, laddove possibile, vengono identificate le entità necessarie allo sviluppo della base di dati. Attraverso l'impiego dell'operatore join, è possibile correlare i dati contenuti in tabelle diverse, sulla base di valori uguali in attributi con lo stesso dominio. Questo permette di effettuare interrogazioni complesse e incrociate all'interno del database.

Definite le entità da prendere in esame, si è partiti nella progettazione dello schema logico, considerandone due in particolare: la composizione e l'allestimento. Per composizione, si intende la versione base dell'opera così come concepita dai suoi autori, in termini di strutturazione (suddivisione in atti, scene) e composizione dell'organico. Questa entità racchiude in sé anche i concetti di edizione, prima rappresentazione assoluta, eventuali note e via dicendo. La composizione è direttamente collegata all'allestimento, in quanto le informazioni riferite alla prima, sono comuni anche a tutte le riprese, quindi ai secondi. La composizione, come entità, ha attributi specifici che la caratterizzano, i quali sono: data e luogo di pubblicazione e titolo. L'allestimento consiste in una particolare messa in scena di una composizione. Tale concetto aggrega i dati che caratterizzano ogni specifica riproposizione della versione base. Si tratta di un set di performances caratterizzate dallo stesso programma musicale, appartenenti alla stessa stagione e aventi in comune caratteri a proposito della produzione, allestimento e del cast. Un allestimento può essere anche la ripresa di uno precedente, ma per motivi cronologici, questo legame stretto si perde e due diverse messe in scena

vengono ricondotte a serate differenti. Esso ha, in quanto entità, attributi, come la composizione, relativi a data e luogo di pubblicazione.

La relazione che unisce le due entità, per prime prese in esame, è di tipo uno a molti, il che significa che una composizione può riferirsi a diverse messe in scena, ma un allestimento è associato ad una e una sola composizione.

Quest'ultima, inoltre, può avere, come è possibile notare, attraverso una tabella di join, uno o più titoli (per esempio, quello originale o quello tradotto) e può dare vita, quindi, ad un'uguale relazione: un titolo può essere associato, infatti, a più composizioni.

Esse sono formate da più movimenti, quindi, la relazione che li lega è di tipo uno a molti. Questo si verifica in virtù del fatto che alla composizione appartengono più parti (o movimenti), ma queste si legano ad una sola opera.

I movimenti, come le composizioni, possono avere uno o più titoli ad essi associati. Attributi di questa specifica entità sono, per esempio, la loro durata, l'organico che si occupa della loro interpretazione e la forma stessa della composizione che, in qualche modo, è determinante.

Entità legate alla composizione in maniera diretta, tramite una relazione uno a molti, sono le partiture. Esse possono essere di diverse tipologie: grafiche, complete o incomplete, a stampa o manoscritte, in diversi formati, per uno o una molteplicità di strumenti. Ad esse sono legati attributi come la data e il luogo di pubblicazione; l'organico, la durata e la forma della composizione, come nel caso dei movimenti; le note e la lingua in cui queste sono scritte; il tipo di segnatura utilizzata e la tecnica. Alla partitura si lega la sua scansione, attività necessaria per renderla fruibile attraverso un elaboratore. Lo stesso legame, per lo stesso motivo, si evidenzia anche per il libretto e il frontespizio. Infatti, sono necessariamente associati ad una specifica composizione, ma quest'ultima, in quanto rappresentabile in molteplici messe in scena, può figurare in diversi libretti o frontespizi.

La composizione si lega, inoltre, all'articolo, inteso come testo apparso su una rivista in cui il titolo di tale opera è citato. La relazione si esplicita tramite una tabella di join che lega le due entità attraverso un'associazione molti a molti. Un articolo, infatti, può parlare di più composizioni, le quali a loro volta possono essere contenute in numero più o meno variabile di articoli. Non solo, l'entità articolo è in relazione anche con le entità personaggio e persona.

Quest'ultima si lega, attraverso un'operazione di join, con la composizio-

ne. Con tale concetto, ci si riferisce ad una persona fisica, caratterizzata da dati anagrafici definiti, che ha un ruolo specifico all'interno dell'opera, come, ad esempio, la sua stessa composizione o la sua interpretazione. Tale concetto è in relazione anche con la lettera, nella quale può esistere un riferimento o una citazione relativa ad una persona. L'entità personaggio è strettamente correlata all'entità composizione attraverso una relazione uno a n, la quale specifica il fatto che un personaggio fa parte di una sola composizione. Utili alla definizione del personaggio sono i costumi, i quali sono caratterizzati da attributi quali i materiali, i bozzetti e figurini e un'immagine scansionata degli stessi.

Relativi all'allestimento, ma correlati ad alcuni degli elementi di cui si è parlato finora come riferiti alla composizione, ci sono entità quali gli attrezzi e le loro tavole, le fotografie e le disposizioni sceniche. Gli attrezzi, infatti, sono collegati all'allestimento, alla loro scansione e alla tavola degli attrezzi nello stesso modo in cui i costumi sono messi in relazione con i figurini, la loro scansione e le note di sartoria.

3.2.4 Progettazione fisica

In questa fase, lo schema logico viene completato con la specifica dei parametri fisici di memorizzazione dei dati. Il prodotto viene denominato schema fisico e fa riferimento ad un modello fisico dei dati. Tale modello dipende dallo specifico sistema di gestione del database scelto e si basa sui criteri di organizzazione fisica dei dati in quel sistema. Aspetti da considerare in questa fase sono: le caratteristiche fisiche di realizzazione della base di dati e la sua dimensione, il tipo di richieste e le risorse già disponibili.

Il risultato è uno schema fisico che descrive le strutture di memorizzazione e accesso ai dati (es. clustering, indici, algoritmi di ricerca,). A seguire, si ha una prima rappresentazione formale e del tutto indipendente dall'implementazione della base di dati (indipendente anche dal tipo di DBMS che sarà utilizzato).

Precedentemente (2.1) si è già visto il risultato di una progettazione fisica per database di teatri d'opera: il progetto DAM per il Teatro alla Scala. Al contrario, la progettazione del database per l'Archivio Storico Ricordi è attualmente ferma alla progettazione logica.

Capitolo 4

Diffusione e fruizione

4.1 Metadati

I metadati sono attualmente un tema cruciale in molti contesti, non solo bibliotecari. Anne Gilliland-Swetland, direttore e professore associato del Department of Information Studies, presso la UCLA Graduate School of Education and Information Studies, per illustrare il significato del termine “metadato” utilizza la seguente metafora: “the sum total of what one can say about any information object at any level of aggregation”. In pratica, si tratta di “dati usati per descrivere altri dati”, ossia informazioni strutturate, scandite in campi e relative ad un documento del quale permettono una maggiore organizzazione e un più efficace recupero.

La funzione principale di un sistema di metadati è quella di consentire il raggiungimento dei seguenti obiettivi:

- *ricerca*: individuazione dell’esistenza di un documento;
- *localizzazione*: recupero di una particolare manifestazione del documento;
- *selezione*: analisi, valutazione e scelta di una serie di documenti;
- *interoperabilità semantica*: possibilità di effettuare la ricerca in ambiti disciplinari diversi, grazie ad una serie di equivalenze fra descrittori;
- *gestione risorse*: coordinamento di raccolte di documenti, grazie all’intermediazione di banche dati e cataloghi;
- *disponibilità*: reperimento di informazioni sull’effettiva disponibilità del documento.

I metadati non devono essere unicamente percepiti come informazioni digitali: sono sempre stati creati sia dal retaggio culturale che dai professionisti dell’informazione e, fino ad oggi, sono stati gestiti all’interno delle

collezioni tradizionali. Ora, con i sistemi di informazione digitale, i metadati si esprimono in differenti modi e forme: incorporati all'interno di biblioteche digitali, incarnati nei documenti contenuti entro periodici elettronici, collegandosi sugli archivi aperti (open archives) e fornendo accesso ai contenuti scientifici. I metadati configurati in formati differenti, interni o esterni ai documenti che essi rappresentano, aprono le vie ad una comunicazione più estesa, mettendo in connessione differenti mondi con regole diverse. Grande attenzione deve essere posta al loro trasporto tra server comunicanti.

Molteplici sono le fonti da cui i metadati possono essere originati. Per molto tempo, queste "informazioni sulle informazioni" sono state create, come precedentemente accennato, da professionisti attraverso attività manuale. Ora, anche utenti meno esperti sono potenzialmente in grado di creare metadati che descrivono i documenti da loro stessi prodotti, attraverso interfacce web che consentono inserimenti manuali in stringhe preconfezionate. Attualmente, vi sono metadati che possono essere creati automaticamente dalla macchina e possono essere interconnessi attraverso complesse relazioni le quali creano legami tra risorse differenti. Da sempre, gli autori di contenuti intellettuali accademico-scientifici hanno utilizzato metadati per descrivere e indicizzare le loro creazioni nei circuiti di comunicazione, all'interno del proprio ambito disciplinare.

È importante comprendere che un metadato di un oggetto informativo può essere contemporaneamente un dato di un altro di uguale tipologia. I metadati assumono importanza strategica nello sviluppo dei sistemi di informazione digitale in rete e in tale contesto ogni previsione conduce ad un concetto di metadato ampio allargato.

4.1.1 L'utilità dei metadati

L'accesso all'universo delle risorse disponibili on-line è divenuto recentemente uno degli obiettivi primari di molte istituzioni scientifiche, le quali gestiscono risorse informative e producono informazioni di contenuto. La descrizione degli oggetti che compongono le collezioni di molte istituzioni, scientifiche e non (biblioteche, musei, archivi...), è stata gestita negli anni più recenti attraverso sistemi di automazione che hanno fornito strumenti per un controllo più efficiente delle collezioni. Tali descrizioni prima erano effettuate con sistemi non automatizzati; nel processo di transizione dalla fase tradizionale a quella automatizzata, hanno subito un processo di ricognizione strutturale e concettuale. Molti dei sistemi gestionali ad oggi esistenti sono stati costruiti senza riferimento alcuno a standard di comunità e, quindi, non sono prodotti condivisibili all'interno della rete. Molto spesso, soprattutto in ambito scientifico-accademico, l'informazione prodotta in settori disciplinari specifici è descritta con linguaggi compresi solo da

quelle specifiche comunità e, quindi, tali insiemi informativi restano isolati. Esistono perciò differenti sistemi che mantengono un proprio disegno strutturale sia per la composizione dei campi (descrittivi o di accesso), sia per la presentazione complessiva dei dati descrittivi all'interno del sistema. Tali differenze si fanno ancora più sensibili quando ci si riferisce a strumenti di supporto per il recupero dell'informazione, quali schemi di classificazione e vocabolari controllati.

L'avvento del web e la crescita esponenziale delle risorse elettroniche ha incrementato anche la domanda da parte di utenti interessati, relativamente alle effettive capacità di avere strumenti che consentano di ricercare, attraverso differenti strutture di metadati, in modo simultaneo. La necessità sempre crescente, delle fasce di utenza specializzata, di poter recuperare informazioni da contenitori integrati ricreando ambienti interdisciplinari, ha motivato molte istituzioni scientifiche a convertire i loro metadati "non standardizzati" in formati più facilmente accessibili. In molti casi, le istituzioni scientifiche hanno messo a disposizione i contenuti intellettuali a testo pieno, prodotti all'interno dei propri circuiti disciplinari, esponendone le descrizioni (metadati) di modo che queste potessero essere "raccolte" da sistemi adeguatamente messi a punto. Tali collezioni di metadati generano indici cumulativi centralizzati, i quali sono il prodotto delle raccolte dai vari server distribuiti lungo la rete e che espongono i metadati descrittivi. In tale modello distribuito, noto come Harvest, vi sono due livelli: il livello dei singoli server (repositories), dove risiede l'informazione a testo pieno prodotta dalle istituzioni scientifiche, e il livello del server che contiene i metadati.

I campi di una collezione di metadati sono costituiti da informazioni che descrivono le risorse informative a cui si applicano, con gli scopi di migliorarne la visibilità e facilitarne l'accesso. I metadati permettono il recupero di documenti primari, indicizzati attraverso sintetiche stringhe descrittive contenute nei record. In esse vengono rappresentate le caratteristiche più significative o le proprietà peculiari dei dati in questione, affinché la loro essenza possa essere catturata da un'unica concisa descrizione, che in modo sintetico e standardizzato fornisce, a sua volta, una strada di ritorno ai dati stessi. Nei progetti di digitalizzazione e nelle attività di gestione degli archivi di oggetti digitali, i metadati rivestono un'importanza crescente, tanto da venire considerati parte costituente della definizione di "oggetto digitale".

In questo specifico contesto, un oggetto digitale è qualcosa che può essere indirizzato e manipolato sia attraverso l'attività umana sia attraverso un sistema automatico. Tale oggetto può essere costituito da un singolo termine o può essere un aggregato di più argomenti. In generale, tutti gli oggetti digitali, indipendentemente dalla forma fisica o intellettuale che assumono, hanno tre caratteristiche: contenuto, contesto e struttura, le quali possono

essere ricomposte attraverso il metadato.

Il *contenuto* stabilisce un nesso tra ciò che l'oggetto contiene in sè, o nei suoi dintorni, ed è intrinseco all'oggetto informativo stesso.

Il *contesto* indica il chi, cosa, perchè, dove, come; tutti aspetti associati con la creazione dell'oggetto e, comunque, estrinseci all'oggetto informativo stesso.

La *struttura* si riferisce all'insieme formale di associazioni situate all'interno o tra gli oggetti informativi.

Tali insiemi possono essere intrinseci o estrinseci agli oggetti e indicare il contesto, la gestione, i processi, la conservazione e l'uso della risorsa che è stata descritta.

4.1.2 Le diverse categorie di metadati

I metadati, oltre ad essere riferibili alla descrizione di un oggetto, possono assumere nel circuito della comunicazione scientifica differenti connotazioni a seconda del ruolo al quale essi sono preposti o in relazione alle funzioni che essi svolgono. Essi sono mobili nel tempo e nello spazio; continuano ad accrescersi entro il sistema in cui dimorano, durante il ciclo di vita dell'oggetto informativo con il quale si relazionano. I metadati, intesi in senso moderno, non possono essere considerati "definitivi", in quanto, una volta creati, non rimangono statici, ma vengono modificati nel corso del tempo e qualche volta possono anche essere disposti in punti differenti durante l'arco della vita di una risorsa, soprattutto se digitale.

Fino ad oggi i professionisti dell'informazione e, in particolare, gli archivisti e i bibliotecari, hanno focalizzato la loro attenzione su metadati in associazione ad attività di descrizione o catalogazione. Con la parcellizzazione delle conoscenze, anche la specificità degli stessi metadati è andata via via differenziandosi, a seconda delle condizioni in cui questi si muovono insieme agli oggetti a cui si riferiscono. I metadati possono non solo descrivere o catalogare un oggetto, ma anche indicare il contesto, la gestione, i processi, la conservazione e l'uso della risorsa che è stata descritta.

Per capire meglio il concetto di metadato, risulta utile frazionarlo in categorie distinte in grado di riflettere gli aspetti chiave delle sue funzionalità le quali conducono a molteplici tipologie:

- metadati *amministrativi-gestionali* MAG (usati per la gestione ed amministrazione delle risorse informative);

- metadati *descrittivi* (dal MARC al Dublin Core);
- metadati *di conservazione* (compresa migrazione);
- metadati *tecnici* (comportamento dei metadati e funzionamento dei sistemi);
- metadati *di utilizzo* (relativi al livello e al tipo di utilizzo dell'utente).

E' possibile effettuare un'ulteriore classificazione in tre macroaree:

- *Metadati descrittivi*: utili all'identificazione ed al recupero della risorsa o degli oggetti digitali; sono costituiti da descrizioni dei documenti fonte o dei documenti nati in formato digitale; risiedono, generalmente, nelle basi dati dei sistemi di Information Retrieval all'esterno dell'archivio digitale e sono collegati a quest'ultimo tramite appositi link.

Ne sono esempio: PURL (Persistent Uniform Resource Locator), Dublin Core, MARC e Meta Tag HTML.

- *Metadati amministrativi e gestionali*: evidenziano le modalità di archiviazione e manutenzione degli oggetti digitali nel sistema di gestione dell'archivio digitale e sono necessari per una corretta esecuzione delle relative attività. Nel mondo digitale, data la labilità dell'informazione elettronica, questi tipi di metadati assumono un'importanza preponderante ai fini della conservazione permanente degli oggetti digitali: essi possono documentare i processi tecnici associati alla conservazione permanente, fornire informazioni sulle condizioni e i diritti di accesso agli oggetti digitali, certificare l'autenticità e l'integrità del contenuto, documentare la catena di custodia degli oggetti, identificarli in maniera univoca.

Ne sono esempio: MAG, Preservation Metadati for digital Collections della National Library of Australia e Metadati for digital preservation CEDARS.

- *Metadati strutturali*: collegano le varie componenti delle risorse per un'adeguata e completa fruizione, la quale avviene spesso attraverso la mappatura di schemi di metadati diversi. Questi inoltre forniscono dati di identificazione e localizzazione del documento, come il codice identificativo, l'indirizzo del file sul server, l'archivio digitale di appartenenza e il suo indirizzo Internet.

Ne sono esempio: SGML, XML, Encoded Archival Description (EAD).

Uno stesso metadato può rientrare in categorie diverse. A ben vedere, i metadati sono strumenti che assolvono determinate funzioni. Potranno,

quindi, esserci tante categorie di metadati quante sono le funzionalità che una determinata tipologia di metadati è in grado di garantire. In altre parole, l'elenco delle tipologie è aperto (può sempre nascere l'esigenza di nuove e specifiche funzionalità).

4.1.3 Interoperabilità ed integrazione

I metadati possono essere utilizzati per consentire un impiego funzionale dei documenti nell'ambito di un determinato sistema informativo (ad esempio: certificare l'autore del documento, stabilirne un periodo di validità, consentirne la lettura a determinate categorie di utenti, individuare la base di dati dalla quale è estratto o il software con cui è stato generato, identificare lo specifico formato del documento e l'ambito della sua applicabilità, stabilire legami operativi con altri documenti). Nel vastissimo panorama del World Wide Web, sotto un unico tetto confluiscono risorse eterogenee, in parte destrutturate e prive di metadati, in parte descritte da schemi di metadati differenti, tanto negli attributi assegnati alle risorse quanto nel livello di analisi. Per questo motivo, i metadati vengono dotati, oltre che di una struttura generale, anche di campi particolari pensati per assolvere ai bisogni descrittivi che un particolare tipo di risorsa necessita. A seconda della tipologia di quest'ultima, c'è la possibilità di specificare campi di metadati con caratteristiche complementari che svolgono funzioni diverse a seconda degli oggetti ai quali devono essere applicati. In un contesto di rete, la discussione sui metadati è diventata una discussione su schemi standard condivisi, per far sì che un'indicizzazione con uno schema di metadati omogeneo consenta l'interoperabilità anche tra tipi di risorse diverse (testi, audio, video, ecc.) e l'integrazione di vari sistemi informativi, sia all'interno che all'esterno dei diversi sistemi locali. Due obiettivi, questi ultimi, che devono basarsi soprattutto sull'accordo e sull'uso di un sistema di indicizzazione standard e quindi su uno stesso schema di metadati.

Importante è il fattore dipendenza tra il dato (la risorsa) e i relativi metadati: nel mondo digitale infatti l'esistenza stessa di una risorsa dipende dai metadati. Essa altro non è che un insieme di bit, ma questi possono veicolare contenuti solo se sono sempre accompagnati da una chiave che li rende comprensibili da un elaboratore oltre che da un essere umano.

4.2 Standard

Gli standard fanno parte della molteplicità di norme, che, quotidianamente, in maniera implicita o esplicita, più o meno consapevolmente, vengono condivise e rispettate affinché sia consentito agli esseri umani l'efficace ed efficiente svolgimento delle attività sociali.

Essi possiedono caratteristiche molto varie sia per quanto concerne il loro ambito di applicazione, sia per quanto riguarda la loro diffusione e accettazione. Quando si parla di standard non ci si riferisce ad un fenomeno moderno. La loro storia evolutiva permette di delineare una successione cronologica che giunge in epoca moderna all'elaborazione delle odierne norme di unificazione. La successione storica segue, a grandi linee, la serie delle tappe del progresso tecnologico. Si viene così a delineare una sorta di parallelismo, secondo il quale ad una particolare epoca tecnologica trova riscontro una relativa tipologia di standard e ad una maggiore complessità del sistema tecnologico corrisponde un'altrettanto articolata struttura di norme unificate. Tuttavia, per le epoche passate non si dovrebbe parlare propriamente di standard nel senso moderno del termine, quanto piuttosto di elementi e norme unificate e unificanti, poichè gli standard moderni contengono un'elevata quantità di caratteristiche vincolanti che ne limitano l'elaborazione, l'uso e la validità.

Dare una definizione tecnica di standard è un compito assai arduo. È possibile comunque specificare il fatto che esso si esprime attraverso delle regole, più o meno complesse, o tramite delle direttive coerentemente espresse all'interno di un documento. La standardizzazione, invece, include l'analisi, la formulazione e l'esplicitazione stessa di tale disciplina. Si può dire che uno standard consiste, sostanzialmente, in una regola, ma non è vero il contrario: non tutte le regole che governano l'operare dell'uomo sono da considerarsi degli standard, o meglio non lo sono propriamente, poichè non tutte rispondono ad altri elementi che invece caratterizzano delle norme tecniche uniformi.

Caratteristiche importanti che devono essere garantite da uno standard sono:

- la stabilità nel tempo;
- l'uso ripetuto;
- la larga diffusione.

La durata temporale di ogni standard può essere interpretata come una sorta di ciclo di vita: quando uno standard non sembra più adeguato a risolvere i problemi per cui era stato formulato, lo si sostituisce con uno nuovo in grado di soddisfare le esigenze più attuali.

Comunemente gli standard vengono emanati per risolvere dei problemi di armonizzazione. La standardizzazione si occupa di individuare delle regole che, rispondendo alle caratteristiche sopra esposte, facilitino le relazioni tra gli individui e/o tra le organizzazioni. Tali entità sono interessate al rispetto degli standard poichè si trovano a dover operare insieme o a fornire strumenti o servizi per il bene comune: devono perciò condividere oggetti, principi

e procedure o offrire prodotti fruibili generalmente.

Nell'ambito catalogafico si rende necessaria la definizione di modelli di documentazione standard per preservare a lungo termine il valore dei dati raccolti in un catalogo e per permettere il loro scambio fra le istituzioni.

Nello specifico, la rappresentazione catalogafica di un documento si compone di due fasi:

- *descrizione*: volta ad esporre, in ordine fisso e in modo omogeneo, una serie di dati relativi agli elementi di natura fisica, testuale e storica ritenuti essenziali e capaci di permetterne l'esatta individuazione patrimoniale e scientifica;
- *determinare dei percorsi* attraverso i quali un utente potrà recuperare la notizia catalogafica.

La scelta dei dati descrittivi è spesso condizionata da implicazioni di ordine amministrativo, poichè risponde alla necessità di inventariare e tutelare il patrimonio informativo e culturale di una comunità; tende, inoltre, al raggiungimento della massima condivisione possibile delle informazioni e dei documenti stessi che le tramandano. Quest'obbiettivo viene perseguito attraverso la formulazione e l'adozione di schemi specifici astratti, i quali contemplanò le informazioni necessarie ad una corretta ed efficace descrizione e li esprimono sotto forma di schede catalogafiche a stampa o di record bibliografici in forma elettronica. Proprio a questo scopo sono nati i cosiddetti standard internazionali per la descrizione bibliografica che a dispetto del loro nome non si interessano soltanto di libri, ma di tutti i documenti che posseggono al pari del libro la qualità di supporti e media informativi.

Il bisogno di un modello standard per la catalogazione dei documenti bibliografici è emerso a partire dalla metà del XIX secolo, momento in cui si è avuto l'inizio di un movimento di elaborazione e di adozione degli standard catalogafici che progressivamente ha coinvolto tutto il settore della biblioteconomia e della documentazione e che, durante gli ultimi trent'anni, ha ricevuto un notevole impulso. Da qui a rendersi conto della necessità di norme specifiche per la catalogazione dei vari materiali (ad esempio: fotografie, materiale musicale, materiale cartografico, oggetti,) il passo è stato breve. Questo perchè ogni materiale ha delle caratteristiche bibliografiche particolari. Un libro a stampa possiede un titolo proprio con il quale può essere facilmente identificato. Al contrario, la maggior parte delle opere musicali non ha un titolo proprio, ma ne possiede uno che consiste nel nome di una forma o di un genere musicale unito alle informazioni riguardanti l'organico vocale-strumentale, il numero d'opera, la tonalità e quant'altro. Solo l'insieme di queste informazioni rende identificabile una composizione.

Tre dei principali standard di catalogazione sono:

- Dublin Core: schema di metadati descrittivi;
- Unimarc: formato per la rappresentazione e lo scambio di dati catalografici;
- MAG: schema di metadati amministrativi e gestionali.

4.2.1 Dublin Core

L'evoluzione del formato

Il progetto del Dublin Core si è sviluppato in ambito OCLC (On-line Computer Library Center), la grande rete di servizi americana per le biblioteche. Nel marzo 1995, si è tenuta una conferenza a Dublin (Ohio - USA), alla quale i partecipanti, bibliotecari, archivisti, editori, ricercatori e sviluppatori di software, oltre ad alcuni membri dai gruppi di lavoro dell'IETF (Internet Engineering Task Force), hanno collaborato al fine di fornire strumenti per l'accesso alle risorse digitali. L'intenzione era quella di suggerire un insieme base di elementi descrittivi che potessero essere forniti dall'autore o dall'editore dell'oggetto digitale ed inclusi in esso o da esso referenziati. Il consorzio di utilizzatori che si è costituito ha incominciato così a sviluppare un'architettura per i metadati che in grado di andare incontro alle necessità dei venditori e dei produttori di informazioni.

Elementi costitutivi

Lo scopo primario dello sviluppo della norma Dublin Core interessava la descrizione delle risorse informative generiche che potevano essere recuperate in una sola sessione di ricerca, trattando i documenti come oggetti forniti di etichette (tag) che ne evidenziano determinate caratteristiche.

I motivi che fanno preferire Dublin Core ad altre simili proposte normative sono:

- *semplicità*: lo standard si rivolge sia a catalogatori esperti che (e soprattutto) a non esperti di catalogazione;
- *comprensibilità*: i descrittori utilizzati sono molto vicini al linguaggio naturale;
- *interoperabilità semantica*: può descrivere documenti appartenenti a qualsiasi disciplina e dovrebbe consentire al Dublin Core di diventare una lingua franca per la ricerca di risorse in rete;

- *flessibilità*: dovrebbe consentire di codificare descrizioni di risorse anche in modo più formale: non ci sono limiti formali al suo sviluppo per ulteriori e diverse applicazioni;
- *consenso internazionale*: il Dublin Core è stato preparato da esperti di varie nazioni e sono in corso numerosi progetti di utilizzo dello standard: è norma in molti siti nel mondo.

Il Dublin Core si propone come uno standard di descrizione delle risorse in formato elettronico. Esso è stato concepito allo scopo di consentire agli autori di effettuare direttamente in modo standardizzato la descrizione di risorse rese disponibili sulla rete, senza la mediazione di un'agenzia catalografica.

Esso è costituito da 15 elementi descrittivi di base. Nel tempo, si è esteso anche a sottoelementi o qualificatori: lo standard è in via di sviluppo, ma il cosiddetto "core dei 15 elementi della descrizione" è rimasto stabile. Ciascun elemento è definito usando un set di 10 attributi ricavati dalla norma ISO 11179. Essi sono, inoltre, ripartibili in tre gruppi, in base a determinate caratteristiche del documento:

- *contenuto*: titolo, soggetti o parole-chiave, descrizione, fonte da cui deriva, lingua, relazione con altri documenti, copertura spaziale o temporale;
- *proprietà intellettuale*: nome del creatore, dell'editore, del contributore e fonte giuridica;
- *esemplarità, o caratteri intrinseci*: data, tipo, formato, identificatore.

I 15 elementi descrittivi sono i seguenti:

1. titolo (title): nome con il quale la risorsa è formalmente conosciuta;
2. creatore (creator): entità che ha la responsabilità principale della produzione del contenuto della risorsa;
3. soggetto (subject): argomento principale della risorsa; può essere espresso da parole o frasi chiave o da codici di classificazione che descrivono l'argomento della risorsa;
4. descrizione (description): spiegazione del contenuto della risorsa;
5. editore (editor): entità responsabile della pubblicazione della risorsa;
6. autore di contributo subordinato (contributor): entità responsabile della produzione di un contributo al contenuto della risorsa;
7. data (date): data associata ad un evento del ciclo di vita della risorsa;

8. tipo(*type*): natura o genere del contenuto della risorsa;
9. formato (*format*): manifestazione fisica o digitale della risorsa;
10. identificatore (*identifier*): riferimento univoco alla risorsa nell'ambito di un dato contesto;
11. fonte (*source*): riferimento a una risorsa dalla quale è derivata la risorsa in oggetto. La risorsa in questione potrebbe derivare, in tutto o in parte, da un'altra risorsa fonte;
12. lingua (*language*): lingua del contenuto intellettuale della risorsa;
13. relazione (*relation*): riferimento ad una risorsa correlata;
14. copertura (*coverage*): estensione o scopo del contenuto della risorsa;
15. gestione dei diritti (*rights*): informazione sui diritti esercitati sulla risorsa.

Sviluppi

Questo set di elementi, originariamente concepito per le descrizioni generate da autori di risorse web, ha visto poi focalizzare attorno a se l'interesse di diverse comunità, quali musei, agenzie ed organizzazioni commerciali, stabilendo un consenso necessario per la standardizzazione a tutti i livelli. Il Dublin Core promuove infatti concetti quali quelli di trasversalità ed interoperabilità ai fini di favorire un nuovo approccio integrato alle informazioni, che venga incontro alle esigenze dell'utenza attraverso una descrizione delle risorse che sia standardizzata ed estendibile a più settori. Il successo di Dublin Core è dovuto alla facile comprensione dei suoi elementi, dalla semantica universalmente accettata e alla facilità di applicazione in lingue diversi (è sufficiente tradurre il nome degli elementi).

Interoperabilità

Il Dublin Core permette la descrizione di una grande varietà di risorse in formati diversi; è anche abbastanza generale da includere ogni indicazione di contenuto semantico. Data la sua semplicità, il Dublin Core è correntemente molto utilizzato e praticamente tutti gli standard e i progetti fanno riferimento ad esso e ne specificano una mappatura. Il Dublin Core è tuttavia troppo generale per un'adeguata descrizione di risorse specifiche; spesso i service provider sono costretti a personalizzare lo schema per soddisfare le esigenze particolari, con il risultato certo non desiderabile che, pur utilizzando lo stesso schema, non sono direttamente interoperabili se non mappando i rispettivi profili applicativi. Quello che può apparire un limite, può rappresentare un punto di forza per il Dublin Core: esso non costringe

l'indicizzatore ad un'eccessiva rigidità nella registrazione delle caratteristiche di una risorsa, consentendogli di creare specifiche dettagliate qualora il dato che deve essere descritto ne richieda la necessità.

Norma ISO

Sotto la guida del NISO (National Information Standards Organization), lo standard di metadati Dublin Core è stato recepito come norma ISO 15836:2003. L'emissione della norma ISO costituisce un riconoscimento ufficiale per l'uso del set Dublin Core che, sin dalla prima conferenza di Dublin nel marzo del 1995, è stato tradotto in oltre 20 lingue ed utilizzato in tutto il mondo per integrare diverse tipologie di informazioni.

4.2.2 UNIMARC

L'evoluzione del formato

UNIMARC (Universal MARC Format) ha come scopo, come tutti i formati della famiglia MARC, quello di favorire lo scambio delle registrazioni bibliografiche in formato elettronico. L'idea della creazione di un formato MARC internazionale fu sviluppata dall'IFLA (International Federation of Library Associations) nella seconda metà degli anni '70, in coincidenza con la diffusione dei differenti formati MARC e con lo scopo di creare un formato sovranazionale che potesse essere utilizzato come formato per lo scambio di registrazioni in formati MARC differenti. UNIMARC ha il compito di specificare gli indicatori di contenuti (tags, indicatori e codici di sotto campi) che devono essere assegnati ai record bibliografici, in formato leggibile dalle macchine, e devono specificare il formato logico e fisico dei record. Esso include monografie, grafici, materiali video, libri rari e risorse elettroniche. Non stipula la forma, il contenuto o la struttura del record dei dati all'interno di diversi sistemi. UNIMARC fornisce le raccomandazioni sulla forma e sul contenuto dei dati quando questi devono essere scambiati. Sono ormai numerose le utilizzazioni di UNIMARC anche per bibliografie nazionali. Dal 1985 la BNI ha adottato UNIMARC, abbandonando l'ANNAMARC; anche in Francia UNIMARC si è andato assai diffondendo e ha pressochè rimpiazzato INTERMARC. Inoltre il formato è stato adottato dal 1986 dalla Biblioteca Nazionale Portoghese, e anche dalla Biblioteca Nazionale della Croazia e anche dalla Biblioteca Nazionale Ceca. A UNIMARC si ispirano inoltre IBERMARC, sviluppato dalla Biblioteca Nazionale Spagnola, e SA-MARC, la versione sudafricana del formato.

Elementi costitutivi

Nella preparazione di UNIMARC si volle creare un formato, per quanto concerne la descrizione bibliografica, compatibile a ISBD (International Stan-

ard Bibliographic Description for electronic resources), che infatti fu in esso incorporato. Inoltre, fu effettuato un notevole sforzo di razionalizzazione che si riflette nella suddivisione e nel raggruppamento dei campi in blocchi secondo una logica precisa, secondo il seguente schema:

BLOCCHI	CAMPI
0 - Blocco di identificazione	da 001 a 009
1 - Blocco di informazioni codificate	da 100 a 199
2 - Blocco di Informazioni descrittive	da 200 a 299
3 - Blocco delle note	da 300 a 399
4 - Blocco dei legami	da 400 a 499
5 - Blocco di titoli in relazione	da 500 a 599
6 - Blocco analisi semantica	da 600 a 699
7 - Blocco responsabilità intellettuale	da 700 a 799
8 - Blocco della fonte	da 800 a 899
9 - Blocco d'uso locale	da 900 a 999

I blocchi hanno le seguenti specificità:

- *blocchi 0 e 1*: dedicati alle informazioni di identificazione e codificate;
- *blocco 2*: dedicato alla descrizione, suddiviso secondo le aree ISBD;
- *blocco 3*: dedicato alle note, suddivise sulla base della funzione e della tipologia;
- *blocco 4*: dedicato ai legami fra registrazioni;
- *blocco 5*: dedicato alle varie forme di titoli;
- *blocco 6*: dedicato all'analisi semantica (soggetti, classificazioni, termini di thesaurus);
- *blocco 7*: dedicato alla responsabilità intellettuale;
- *blocco 8*: dedicato alla fonte bibliografica;
- *blocco 9*: non definito dallo standard e utilizzabile per inserire dati locali.

In ogni registrazione UNIMARC devono essere necessariamente presenti i seguenti campi:

- 001 Numero record
- 100 Dati per l'elaborazione
- 101 Lingua dell'opera (quando applicabile)
- 200 Titolo e indicazione di responsabilità (solo il primo è obbligatorio)
- 801 Fonte di origine del record

A questi campi si posso aggiungere, per materiali specifici:

- 120 Dati codificati generali per il materiale cartografico
- 123 Dati codificati scala e coordinate per il materiale cartografico
- 206 Dati matematici per il materiale cartografico

Un record UNIMARC può essere lungo al massimo 99.999 caratteri.

Un'ulteriore e importante caratteristica di UNIMARC è la possibilità delle gestione dei legami bibliografici fra record differenti, ai quali è dedicato l'intero blocco 4: è questa un'innovazione importante e qualificante del formato, rispetto ad altri MARC nazionali. La creazione del legame avviene utilizzando una tecnica, detta di incapsulamento (embedding), che consente di inserire all'interno del campo rinvii formali a dati appartenenti ad altri blocchi. Nel 1991 è stato pubblicato anche il formato dedicato alle registrazioni di autorità.

Sviluppi

Nel corso degli ultimi anni, il Gruppo italiano UNIMARC sezione Musica, presso l'ICCU, ha avanzato delle proposte di adeguamento del formato per introdurre alcuni elementi specifici della catalogazione musicale, ampliando campi esistenti o creandone di nuovi. Le proposte sono state esaminate e approvate dal Permanent Unimarc Committee nella quasi totalità.

Si tratta di nuovi campi per l'incipit musicale e il mezzo di esecuzione e dell'ampliamento dei campi per a) le forme musicali e la tonalità/il modo, b) la presentazione della musica (con annesso glossario multilingue), c) la data di rappresentazione, registrazione o rimasterizzazione, d) la citazione bibliografica; sono stati infine inseriti nuovi codici per identificare i libretti e i libri liturgici. La IAML (International Association of Music Libraries) è stata delegata al mantenimento dei codici. Sono state invece rinviate le proposte sui campi codificati riguardanti i manoscritti musicali, perchè si è ritenuto più opportuno inglobarle in una più ampia serie di adeguamenti del formato per i manoscritti. Il gruppo italiano è stato anche incaricato di redigere in tempi brevi le linee guida per il trattamento del materiale musicale.

Normativa

Come tutti i formati MARC (e non solo quelli), UNIMARC utilizza come standard di codifica del dato l'ISO 2709, creato appositamente per l'interscambio di dati bibliografici su supporti magnetici.

4.2.3 MAG (Metadati Amministrativi-Gestionali)

L'evoluzione del formato

L'importanza dei metadati amministrativi-gestionali diventa sempre più netta via via che emergono le problematiche poste sia dalla gestione quotidiana degli oggetti digitali (acquisizione ed amministrazione di archivi di file digitali) sia dalla necessità di garantire la cosiddetta conservazione digitale (la quale implica anche l'accesso a lungo termine degli oggetti in questione). Di minore importanza è, a prima vista, la funzione di questo tipo di metadati ai fini della ricerca e del recupero della risorsa. Tuttavia, è insito in questa tipologia di metadati anche un aspetto di organizzazione degli oggetti digitali parte di una collezione, che diventa indispensabile per una loro corretta fruizione. In sostanza, i metadati amministrativi-gestionali permettono la documentazione di processi tecnici associati alla conservazione permanente, consentono di fornire informazioni sulle condizioni e sui diritti di accesso agli oggetti digitali, garantiscono la certificazione dell'autenticità e dell'integrità del contenuto e, infine, consentono di documentare la catena di custodia degli oggetti identificandoli in maniera univoca.

Elementi costitutivi

Nel lavoro di definizione di un primo set standard di metadati amministrativi e gestionali, si è tenuto conto delle principali iniziative e dei più significativi progetti realizzati in ambito internazionale. Sulla base di tali esperienze si è ritenuto di mettere in evidenza quanto segue:

- più che per particolari tipologie di documenti fonte (es. periodici, musica a stampa o manoscritta, carte geografiche etc.), il set di metadati di base viene definito per tipologie di oggetti digitali.

Una prima distinzione può riguardare:

- immagini statiche,
- testi prodotti con tecnologia OCR,
- suono,
- immagini in movimento e oggetti multimediali,
- *born digital*;

- i metadati ed il loro sistema di gestione devono essere completamente indipendenti da specifiche piattaforme hardware e software, al fine di favorirne un impiego generalizzato;
- i metadati ed il loro sistema di gestione devono invece essere coerenti con le funzioni previste nel modello logico-funzionale standard dell'archivio degli oggetti digitali cui si fa riferimento (ad esempio: le funzioni di immissione, archiviazione, gestione, accesso, amministrazione, pianificazione della conservazione, nel modello OAIS).

Lo schema generale è composto dalle sezioni:

- *gen*: informazioni generali sul progetto e sul tipo di digitalizzazione;
- *bib*: metadati descrittivi sull'oggetto digitalizzato;
- *stru*: metadati strutturali;
- *img*: metadati specifici relativi alle immagini fisse;
- *ocr*: metadati specifici relativi al riconoscimento ottico del testo;
- *doc*: sezione utilizzata per descrivere, ad esempio, un file in formato .pdf o .rtf.

È previsto l'inserimento nello schema delle sezioni relative al suono ed alle immagini in movimento ed un'evoluzione per il trattamento dei metadati nel caso di oggetti digitali nativi.

MAG, e gli schemi di metadati in genere, sono componenti necessarie di un archivio digitale che rispetti lo standard OAIS (Open Archival Information System). OAIS è uno standard ISO del 2002, che si autodefinisce come “un archivio, consistente in un'organizzazione di persone e sistemi, che ha accettato la responsabilità della conservazione dell'informazione e del renderla disponibile per una determinata comunità”. A tale scopo, esso individua termini e concetti rilevanti per l'archiviazione di documenti digitali e identifica le componenti ed i processi chiave comuni alla maggior parte delle attività di conservazione digitale. Propone, inoltre, un modello logico di riferimento per gli oggetti digitali e i metadati loro associati, il quale comprende la creazione e l'uso dei metadati utili a gestire il materiale elettronico, dalla fase di acquisizione a quella della conservazione, fino alla fase di accesso. Di fatto, è lo standard emergente per la conservazione delle risorse digitali. OAIS si pone però come modello logico di riferimento di valore generale ed astratto e prescinde da una specifica implementazione.

Sviluppi

L'ambito naturale di applicazione di questo lavoro è la Biblioteca Digitale Italiana intesa come insieme dei progetti coerenti con determinate impostazioni. Nonostante la frammentarietà delle esperienze di Biblioteca Digitale in Italia, non si può ignorare la presenza di episodi validi che hanno generato, all'interno delle strutture dove si sono realizzate, una serie di specifiche competenze che devono costituire il punto di partenza per qualsiasi futuro sviluppo nel campo.

Il Comitato MAG si è posto, fin dall'inizio, i seguenti obiettivi:

- diffusione dello Schema MAG;
- completamento dello schema per le fasi AIP (Archival information package) e DIP (Dissemination information package) dello schema MAG;
- mantenimento ed evoluzione dello schema;
- produzione di manuali e linee guida;
- assistenza agli implementatori;
- formazione e promozione;
- rapporti con altri progetti e agenzie (Progetti europei, Dublin Core, etc.);
- sviluppo di schemi e profili di applicazione di metadati descrittivi nei diversi settori della cultura.

I gruppi di lavoro hanno di conseguenza avviato le seguenti attività:

- elaborazione del manuale per l'utilizzo dello Schema MAG;
- produzione di documentazione di supporto sia a carattere generale che specifico per le diverse tipologie di materiale sottoforma di raccomandazioni e linee guida;
- produzione di un registro di link a schemi, modelli e reference tecnici.

Normativa

MAG è un "application profile" perfettamente conforme agli standard internazionali. In pratica MAG permette l'uso di metadati mantenuti e definiti da altri schemi (Dublin Core e NISO) in combinazione con metadati specifici definiti per una particolare applicazione (si è fatto ricorso a questa seconda opzione solo dove non è stato possibile trovare una corrispondenza consolidata con schemi esistenti).

4.3 La Rete della Musica Italiana (ReMI)

La Direzione Generale per i Beni Librari ed il Comitato Guida della Biblioteca Digitale Italiana, nel mettere a punto i contenuti della linea tematica “Musica” del programma BDI/NTC, hanno individuato nell’architettura, negli standard e nel modello dei servizi ADM (di cui si è parlato nel secondo Capitolo) il punto di partenza per costruire la Rete della Musica Italiana (progetto ReMI).

Il passaggio da ADM a ReMI apre uno scenario vasto e complesso: ai fini del riutilizzo in ReMI delle soluzioni ADM è necessario, per quanto riguarda il contesto tecnologico, verificare eventuali esigenze di adeguamento a tecnologie hardware/software che nel frattempo possono essere evolute. ADM ha infatti avuto una genesi piuttosto lunga, facendo il più possibile ricorso a componenti di ampia diffusione. Occorre verificare se queste restano commercialmente supportate o se il mercato suggerisce o impone adeguamenti o scelte diverse. È necessario verificare anche le componenti software ADM che possono essere influenzate dalle dimensioni della rete, dal numero delle transazioni e dal numero dei partner. ADM è stato dimensionato su 3 partner; occorre capire cosa va adeguato o modificato nel caso in cui i partner passino da 3 a 30 o ad n.

Alcuni punti fermi e alcune differenze fra le due reti sono qui di seguito elencate:

- conferma dell’architettura distribuita, aperta, e della sostanziale “assenza di un centro”, propria di ADM;
- conferma dell’assoluta centralità degli standard (formato dei dati, protocolli di comunicazione, etc);
- differenziazione dei materiali: oltre alla musica manoscritta, oggetto principale delle attività ADM, la rete ReMI prevede l’accesso a documenti digitalizzati diversi, come ad esempio: musica a stampa, trattati di teoria musicale, bozzetti di costumi o scenografie, fotografie, libretti, documenti d’archivio di teatri o altre istituzioni;
- forte accentuazione della presenza e del ruolo delle collezioni audiovisive, sonore e multimediali, con particolare riferimento all’esperienza della Discoteca di Stato, attuale coordinatore del progetto ADM;
- esigenza di prevedere livelli differenziati di partecipazione a ReMI (le collezioni musicali italiane sono estremamente frammentate, i loro titolari possono avere caratteristiche e capacità molto diverse).

4.3.1 L'architettura di ReMI

Tenendo presenti i punti sopra citati, è possibile osservare come la topologia prevista per la rete ReMI presenti delle differenze rispetto alla rete ADM. Essa è a quest'ultima, in qualche modo, sovra ordinata, nel senso che ADM può presentarsi come una delle reti accessibili da ReMI.

Alle risorse (metadati descrittivi e oggetti digitali) presenti sulla rete si accede tramite un portale che interroga nodi di tipo diverso: nodo autonomo, nodo supportato, nodo rappresentato, nodo concentratore. Non c'è un indice, il portale raccoglie e presenta metadati, le teche digitali forniscono direttamente gli oggetti digitali.

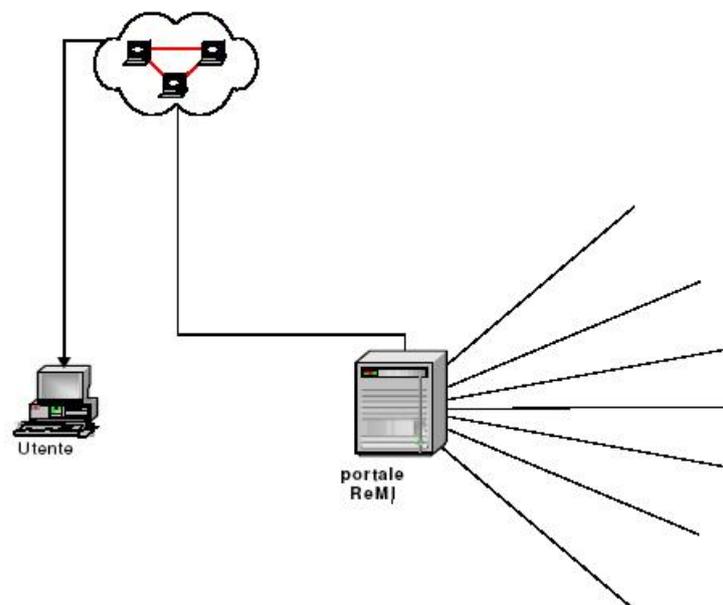
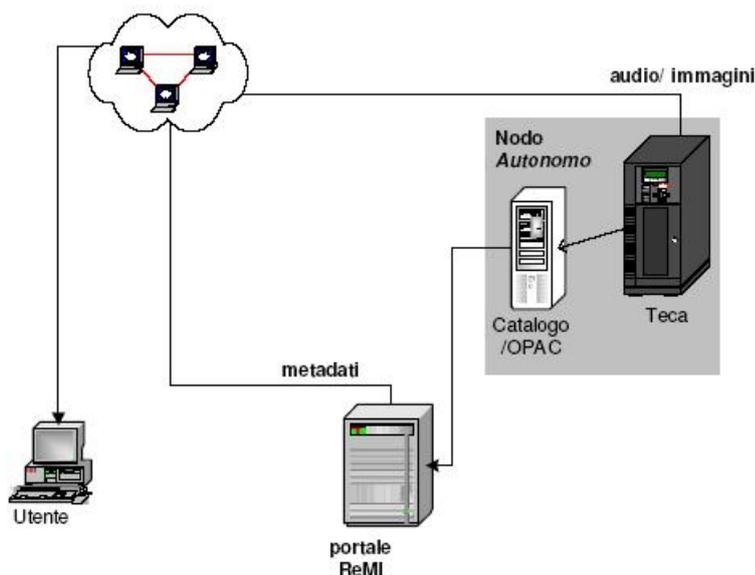


Figura 4.1: Portale ReMI.

Per quanto riguarda il portale ReMI, esso consiste in una macchina dotata di un software che consente all'utente di interrogare in maniera integrata i nodi presenti sulla rete e ottenere la prospettazione dei risultati. Si serve di un webservice e di un application server di comune utilizzo. È accessibile tramite un normale browser web e ha funzioni di interfaccia, sul modello di quelle dell'OPAC ADM. Non è un catalogo, non è un indice e non ospita record catalogafici. Dal punto di vista dell'utente, è l'interfaccia OPAC di un sistema distribuito.

I nodi di rete interrogabili attraverso il portale possono essere di tipo diverso:

- nodo autonomo:



Esso viene interrogato direttamente dal portale; dispone di un catalogo (applicativo gestionale per la catalogazione) accessibile on-line, in grado di fornire dinamicamente dei dati in un formato tale che possano essere processati dal portale, il quale si farà carico della visualizzazione OPAC; dispone di una teca digitale accessibile da remoto (repository degli oggetti digitali e dei relativi metadati amministrativi e gestionali, secondo lo standard MAG);

- nodo supportato:

è un nodo dotato di teca digitale accessibile da remoto che implementa un tracciato del record catalografico o dei protocolli che non soddisfano pienamente gli standard ReMI. In altri casi, appartiene ad una rete che offre servizi a valore aggiunto, come appunto la rete ADM; il nodo supportato viene interrogato dal portale tramite un nodo concentratore;

- nodo rappresentato:

dispone solo di un catalogo; esporta, quindi, i propri record catalografici nell'OPAC del nodo concentratore ed i propri oggetti digitali nella teca digitale del nodo concentratore;

Comitato dei Ministri per la Società dell'Informazione. L'iniziativa è stata messa a punto dalla Direzione generale per i Beni Librari e gli Istituti Culturali (DGBL) e dall'Istituto Centrale per il Catalogo Unico delle biblioteche italiane e le informazioni bibliografiche (ICCU).

Il portale Internet Culturale [14] propone un sistema di accesso integrato alle risorse digitali e tradizionali di biblioteche, archivi ed altre istituzioni culturali italiane, promuovendo e valorizzando la conoscenza e la fruibilità del patrimonio turistico culturale, sia a livello nazionale che internazionale.

Questa nuova piattaforma orizzontale, omogenea ed integrata, della Biblioteca Digitale Italiana consente una visione unitaria da parte dei cittadini di ben 2.300 biblioteche, con la possibilità di consultare in rete tutti i loro titoli. Si verifica un passaggio dalla babele delle biblioteche ad una interconnessione di 54 poli territoriali. Il progetto intende, di pari passo con il suo consolidamento, attuare un nuovo modello di business che preveda anche un ritorno economico per l'acquisto on-line di immagini digitalizzate ad elevata qualità, con pagamento attraverso carta di credito. Con lo stesso standard della biblioteca digitale, si sta inoltre promuovendo il progetto "biblioteche nelle scuole" con l'obiettivo di offrire accesso via Internet a tutti i 350.000 testi delle 500 biblioteche scolastiche. Il nuovo portale, inaugurato ufficialmente il 22 Marzo 2005, permetterà di localizzare un libro o un documento, sapere in tempo reale se è disponibile per la lettura ed il prestito, scoprire se ne esiste una copia digitale, avere informazioni sulla biblioteca che lo possiede e sui servizi che offre e nel caso, attivare il prestito.

Il progetto si prefigge da un lato, lo scopo di tutelare il patrimonio culturale e librario e dall'altro, quello di consentire una sua maggiore fruibilità, da parte del pubblico, in virtù del fatto che l'accesso alle copie digitali sarà alla portata di tutti e non più per una ristretta cerchia di studiosi accreditati. Internet culturale è stato finanziato con 37,3 milioni di euro, di cui 7,1 milioni stanziati dal Comitato dei Ministri per la Società dell'Informazione e 30,2 dal Ministro per i Beni e le Attività Culturali.

L'iniziativa progettuale è stata valutata di preminente interesse nazionale in quanto volta alla creazione delle basi tecnologiche, metodologiche ed organizzative necessarie a garantire l'evoluzione funzionale del sistema informativo delle biblioteche italiane, indispensabile per dotare il paese di un'infrastruttura innovativa in grado di competere sul mercato globale dell'informazione e del turismo culturale.

4.4.1 Lo sviluppo del progetto

Quando il progetto ha preso ufficialmente il via, nel 2001, il cuore del sistema informativo delle biblioteche italiane è stato trasformato da centro di servizi catalografici a sistema di accesso integrato a documenti digitalizzati, oltre che tradizionali, con l'obiettivo di sfruttare al meglio l'enorme patrimonio del Servizio Bibliotecario Nazionale. Ciò a cui si mirava era una rete in grado di collegare biblioteche, centri di eccellenza tematica, scuole e università, in un sistema integrato di catalogazione, digitalizzazione, produzione e vendita di contenuti e servizi per cittadini, studenti, ricercatori, docenti. Nel 2001 è stato istituito un Comitato Guida, sotto la presidenza del professor Tullio Gregory, composto da esperti del mondo bibliotecario e museale, dell'università e della ricerca, con il compito di definire le linee guida in raccordo con il contesto comunitario e internazionale. Nel 2004 a RTI Finsiel, IBM e SSB è stata affidata la realizzazione del primo nucleo del progetto, il cui obiettivo primario era quello di realizzare e sperimentare i principali elementi di innovazione di SBN, anche in considerazione di successive evoluzioni sia a livello centrale che a livello locale, nell'ambito della seconda tranche dei finanziamenti e-government.

Il progetto si articola nelle seguenti componenti:

- governo amministrativo: inteso come sviluppo di strumenti per la gestione del copyright, il pagamento dei diritti, la ripartizione delle entrate, la gestione della sicurezza delle transazioni e dei rapporti economici con gli utenti;
- gestione della conoscenza: intesa come individuazione di metodiche di trattamento cognitivo dell'informazione e conseguente sviluppo di strumenti per l'individuazione e l'accesso alle fonti informative, sia di tipo catalografico che in formato testo/immagine, tramite strumenti di knowledge management;
- realizzazione del "Portale della Biblioteca Digitale Italiana" inteso come nuovo servizio integrato di accesso alle risorse digitali e tradizionali che realizzi una ridefinizione complessiva della Digital Image dei vari siti di accesso attualmente esistenti, al fine di offrire una struttura di interfaccia di facile utilizzo sia per l'utente generico che per l'utente disabile;
- definizione delle metodiche di creazione degli oggetti digitali costitutivi della BDI, di progettazione integrata e coordinata dei loro contenitori e di definizione degli standard di nomenclatura e di fruizione (metadati).

In linea con iniziative di altri paesi europei, il primo vero e proprio progetto che è stato avviato, nell'ambito della BDI, è stato la scansione in

formato immagine dei cataloghi delle biblioteche pubbliche italiane. L'ICCU, in questo contesto, ha avuto il duplice ruolo di ente digitalizzatore e di responsabile del monitoraggio del programma. Successivamente, sono stati finanziati progetti riguardanti il patrimonio musicale, conservato presso biblioteche pubbliche e conservatori di musica, e la digitalizzazione delle pubblicazioni periodiche di particolare valore storico e culturale. Le biblioteche, consapevoli dell'utilità per l'utenza, hanno accolto con grande entusiasmo l'avvento del digitale, dando il via a vari progetti di digitalizzazione. L'ICCU, su indicazione del Comitato Guida, è stato incaricato di realizzare un censimento dei progetti di digitalizzazione conclusi, o in fase di elaborazione, sia per evidenziare le attività svolte sia per evitare duplicazioni.

Il passo successivo, in linea con altre nazioni europee, è stato quello di realizzare un Network a carattere Turistico Culturale che potesse offrire all'utenza, in modo strutturato, tutte le risorse digitali e quelle bibliografiche tradizionali di maggior spessore.

Lo sviluppo del progetto prevede il consolidamento dell'infrastruttura tecnologica, l'ingegnerizzazione degli strumenti sperimentati, la definizione di ulteriori percorsi culturali e aree tematiche, l'inserimento di nuovi contenuti digitali.

Sono messe in programma, fin da subito:

- 1.000.000 nuove digitalizzazioni, nelle aree della cultura musicale, letteraria e scientifica;
- 200.000 digitalizzazioni del patrimonio della Discoteca di Stato di Roma;
- il recupero di 10 repository di varie istituzioni culturali.

Lo sviluppo del progetto è stato realizzato dalla Direzione Generale per i Beni Librari e gli Istituti Culturali che si è servita, per gli aspetti metodologici e scientifici, delle competenze dell'ICCU, con la collaborazione di Regioni, enti locali, istituti culturali e centri di eccellenza, i quali hanno messo a disposizione competenze, progettualità e finanziamenti, secondo le modalità di accordo e collaborazione già sperimentate nella prima fase delle attività.

4.4.2 Il portale Internet Culturale

I servizi e contenuti del primo nucleo del portale sono i seguenti:

- *ricerca bibliografica*: tramite l'OPAC SBN (Servizio Bibliotecario Nazionale) è possibile effettuare una ricerca bibliografica di opere dall'inizio della stampa fino ai giorni nostri, possedute negli archivi di 2300

biblioteche italiane ed appartenenti ad istituzioni pubbliche o private operanti in diversi settori disciplinari. Altra ricerca è fattibile nei *cataloghi storici* che offrono la digitalizzazione di: cataloghi a volume e a scheda di 35 biblioteche italiane, cataloghi di materiali speciali (manoscritti, carte geografiche...) e cataloghi di singoli fondi o raccolte. Inoltre è consentito all'utente l'accesso ai *cataloghi speciali*, in particolare alle basi dati Bibman e Manus per i manoscritti, e a Edit16 che, relativamente alle edizioni XVI sec., consente l'accesso ad archivi di autori, titoli, editori, marche tipografiche e il collegamento alle relative immagini digitalizzate (frontespizi, colophon, marche);

- *accesso alle collezioni digitali*: il repertorio delle collezioni digitali descrive i fondi digitalizzati o in corso di digitalizzazione disponibili presso biblioteche, archivi e altre istituzioni culturali in Italia. L'obiettivo è quello di consentire agli utenti di identificare e localizzare le collezioni digitali disponibili e, ove possibile, consultarle direttamente on-line. La ricerca di immagini e materiale digitale è effettuata mediante l'utilizzo di strumenti all'avanguardia in questo settore, come il sistema internazionale standard SDX per l'accesso alle schede delle collezioni digitali;
- *percorsi culturali*: questa sezione propone suggestivi percorsi di conoscenza che, offerti in più lingue, permettono la diffusione di fonti anche inedite, configurandosi come una vera e propria mappa geografico-culturale. L'offerta si articola in quattro moduli: Mostre, Viaggi nel testo, Itinerari turistico-culturali e Percorsi 3D.
 - *Mostre*: sono attive venti mostre in formato digitale complete di testi ed immagini, rassegne specifiche su alcuni degli autori più importanti del panorama letterario italiano e accurate mostre fotografiche capaci di restituire atmosfere ed immagini ormai perse nel tempo. Sono inoltre presenti esposizioni che riassumono gli importanti progetti di promozione e valorizzazione del patrimonio culturale realizzati dalle biblioteche pubbliche statali nel corso degli anni;
 - *Viaggi nel testo*: è disponibile un'articolata proposta di percorsi monografici strutturati per ipertesti sulla vita e l'opera di importanti personalità della musica e della letteratura italiana (Verdi, Petrarca, ecc.). In quest'area è possibile accedere, tra l'altro, a "Rinascimento Virtuale", percorso realizzato dalla Biblioteca Medicea Laurenziana che illustra i risultati acquisiti e gli strumenti impiegati per la riproduzione e la decifrazione degli antichi palinsesti greci. È possibile, inoltre, scaricare opere di letteratura italiana tradotte in spagnolo nell'ambito del progetto "Un

Mare di Sogni” che si propone di diffondere la cultura italiana in America Latina;

- *Itinerari turistico-culturali*: è presente una serie di itinerari che coniugano contenuti culturali ad informazioni di carattere scientifico su istituzioni, documenti, oggetti. I primi tre percorsi tematici illustrano: i luoghi di maggiore rilevanza scientifica e tecnologica del territorio toscano, le testimonianze della storia e della cultura del Piemonte e gli itinerari toscani dei viaggiatori francesi e inglesi dalla fine del XVII secolo ai primi del XIX secolo;
- *Percorsi 3D*: si può accedere a veri e propri viaggi tridimensionali che offrono all’utente la possibilità di muoversi in ambienti virtuali, ricostruiti grazie all’ausilio di tecnologie all’avanguardia. Il visitatore può aggirarsi liberamente per la sala della Tribuna di Galileo, dove un’esposizione degli strumenti scientifici e degli affreschi restituisce le atmosfere proprie di un periodo cruciale nella storia del pensiero scientifico, oppure può perdersi tra le musiche e la storia dell’Opera di Parma, riassunte attraverso le immagini dei compositori, i libretti d’opera e le locandine d’epoca.
- *un sistema di gestione della conoscenza*: la classificazione dei dati informativi è stata realizzata mediante un sistema KM (Knowledge Management), ovvero di gestione della conoscenza. Pertanto sono stati analizzati i contenuti informativi tratti da un elevato numero di siti web, dal catalogo OPAC SBN e da documenti digitalizzati, dotati di record MAG; quindi sono state approntate 3 aree tematiche (letteraria, scientifica e musicale) con la collaborazione di singoli Centri di Eccellenza: il Dipartimento di Italianistica dell’Università La Sapienza di Roma, l’Istituto e Museo di Storia della Scienza di Firenze, la Casa della Musica di Parma, e con l’utilizzo del software Verity 5.5, attraverso un’applicazione appositamente sviluppata per il progetto;
- *una rete di servizi relativi al patrimonio culturale nazionale*: con informazioni su eventi, novità, attività, istituzioni pubbliche e private pertinenti al mondo degli archivi, delle biblioteche e della cultura in generale.

In merito alla questione della ricerca bibliografica, nello specifico, il portale offre tre nuove interfacce di ricerca in SBN: Semplice, Base e Avanzata.

La *ricerca semplice* offre un unico campo; le parole immesse verranno ricercate nei canali titolo, autore e soggetto. Queste parole sono trattate in OR, ma in realtà le ricerche sono eseguite in AND. Il carattere jolly è l’asterisco e il testo di help comprende una lista di stop word. Il testo non definisce cosa siano le stop word, che di solito sono articoli e altre parole

molto generiche che vengono ignorate dalla ricerca.

La *ricerca base* presenta tre campi distinti per titolo, autore e soggetto, con la possibilità di consultare le tre liste corrispondenti, e fornisce in più due campi per delimitare il periodo di pubblicazione, da anno ad anno. Il testo di help non specifica nulla, né se i campi vengono trattati in AND o in OR, né come vengono trattati maiuscole e caratteri accentati e neppure se si possono inserire più parole chiave in uno stesso campo. Per maiuscole e lettere accentate, il sistema è molto elastico. È possibile inserire più termini in ciascun campo. Tutte le parole chiave inserite vengono trattate in AND, sia nello stesso campo sia tra campi diversi.

Risulta molto più interessante la *Ricerca avanzata* con interrogazioni su titolo, autore, soggetto e altri undici campi, dal luogo di pubblicazione al codice identificativo del record. All'interno di un singolo campo si può operare una selezione di tutte le parole, di almeno una parola o della frase esatta e tra i tre diversi campi delle tre possibilità in E, O ed E NON. A questo si aggiungono campi per selezionare l'anno di pubblicazione, il livello bibliografico, il tipo di documento, il paese, la lingua e la data di aggiornamento, con una maschera che nel complesso è comprensibile e completa, risultando l'unica che può essere usata senza ambiguità anche per ricerche elaborate. Con questo form, tuttavia, non sono disponibili le liste, che vengono presentate solo in Ricerca base. Tutte le liste sono però disponibili dal bottone Indici, sempre presente nei menu in alto sulla schermata.

Sotto il menu Ricerche tematiche si trova la possibilità di scegliere la base dati e la tipologia del documento con interfacce di ricerca analoghe a quella della Ricerca avanzata e abbastanza chiare.

Tali tipologie di ricerche si possono ritrovare in qualsiasi archivio accessibile da tecnologie web, non presentano nulla di innovativo, a differenza di quanto abbiamo visto per il progetto DAM dove anche per una semplice ricerca venivano forniti dati che in questo caso sarebbero ottenibili solo con più ricerche.

4.4.3 Cultura, qualità e costi

Con tali iniziative, in Italia, la cultura ufficiale esce finalmente dai musei e si apre alle nuove tecnologie: attraverso il portale, presentato nella sua veste rinnovata il 15 marzo 2006 a Roma, per la prima volta una pubblica amministrazione (nello specifico la Direzione Generale per i Beni Librari e gli Istituti Culturali) introduce tra i suoi servizi l'e-commerce, per offrire agli utenti web la possibilità di acquistare in Rete immagini esclusive ad alta definizione di manoscritti, partiture musicali, foto storiche.



Figura 4.4: Il portale Internet culturale (homepage)

Il passaggio all'audiovisivo è importante per mettere a disposizione di una platea molto più vasta il patrimonio librario, musicale e iconografico con metodologie che aiutino ad ordinare e rintracciare opere e contenuti altrimenti introvabili. Al tempo stesso, però, non bisogna perdere la cultura del libro, separandola dalle biblioteche e dal rapporto diretto con il mondo che le circonda: Internet è e sarà tanto più efficace se rappresenterà il luogo del dialogo tra la modernità e la tradizione. La digitalizzazione è considerata dall'Unione Europea uno dei fattori essenziali per sostenere e promuovere la diversità culturale, oltre a permettere anche, in ambito bibliografico ed archivistico, di ridurre la manipolazione e l'usura dei beni consultati. L'Italia, tramite il programma di Biblioteca Digitale Italiana, è stata uno dei primi paesi ad offrire un proprio contributo specifico al processo di valorizzazione e diffusione dei contenuti culturali e scientifici europei.

La tempestività di risposta del portale Internet Culturale a una domanda crescente in ambito internazionale è stata indiscutibilmente premiata: dal marzo 2005 a oggi, il sito ha potuto contare 16 milioni di ricerche bibliografiche effettuate ed oltre 270 mila visitatori da cento nazioni (al punto che sono già in cantiere la versione francese e spagnola del servizio attualmente bilingue, in italiano e inglese). In conformità al progressivo aumento del numero di consumatori che preferiscono acquistare via Internet, invece

di utilizzare i canali tradizionali, Internet Culturale ha deciso di offrire ai suoi utenti un nuovo servizio di e-commerce che consentirà loro di acquistare on-line i contenuti culturali proposti gratuitamente a bassa risoluzione o di acquistarli a pagamento a un livello di qualità superiore. La prima fase sperimentale servirà a valutare l'impatto dell'iniziativa sui visitatori, ma anche a fugare i dubbi che riguardano la tutela dei diritti e la sicurezza dei pagamenti on-line, tramite lo sviluppo di strumenti di Digital Right Management (DRM), finalizzati a controllare l'accesso all'opera digitale piuttosto che a cercare di evitarne la duplicazione illegale. Diverse le istituzioni culturali coinvolte nel progetto, tra cui la Biblioteca Nazionale Centrale e il Museo di Storia della Scienza di Firenze, la Casa della Musica di Parma, il Conservatorio napoletano di S. Pietro a Majella.

La questione del commercio elettronico permette di affrontare un tema molto caro agli utenti della rete e, in questo caso, della cultura. Il fatto che il portale Internet Culturale renda disponibili gratuitamente all'utenza una serie di contenuti digitali di bassa qualità, preservando il resto alla vendita, sottolinea l'opinione, legittimata dal Patto di Sanremo, secondo il quale debba esistere un modello legale di fruizione dei contenuti. Ciò va a modificare le aspettative degli utenti, i quali sono ancora legati all'idea che tutto ciò che transita sulla rete debba essere gratuito. Internet quindi deve essere considerata una sorta di area in cui tutte le iniziative e le misure legislative devono essere a prova della Rete, tenendo conto delle sue peculiari caratteristiche tecnologiche e della necessità di far crescere la competitività del mercato. Con il Patto si vuole incentivare ed incoraggiare il raggiungimento di specifici accordi tra tutte le parti interessate, favorendo la collaborazione degli operatori del settore e degli utenti in cui le Istituzioni giochino il ruolo di "garanti" e "promotori".

La vendita di contenuti via web, a seconda del livello di qualità che questi offrono, solleva una questione importante. Facendo un confronto con il progetto DAM della Scala di Milano, di cui si è parlato nel secondo capitolo, è possibile osservare come la preservazione dei contenuti avvenga in maniera differente. Infatti, tale progetto rende fruibili, già in partenza, una serie di contenuti selezionati: quello che il Teatro decide di rendere disponibile e, oltretutto, all'interno di una rete Intranet. In aggiunta, tale progetto prevede l'impiego di profili di accesso che filtrino, anch'essi in partenza, il tipo di utenza che può accostarsi a determinati materiali. Internet Culturale, invece, si è posto fin dal principio l'obiettivo di raggiungere un pubblico non necessariamente specialistico, ma anche e semplicemente appassionato. L'eterogeneità del materiale e l'introduzione dell'audiovisivo on-line deriva anche da questo. Di base, questo portale permette gratuitamente l'accesso a tutti, in maniera indiscriminata, in Internet e a tutto il materiale che i vari enti associati decidono di digitalizzare e mettere in linea. Per quanto

riguarda ciò che ha una qualità di risoluzione adeguata. Se si mira alla fruibilità dello stesso bene, con qualche ambizione e richiesta di qualità in più, si è costretti a pagare.

Capitolo 5

Conclusioni e sviluppi futuri

L'Europa sta preparando un futuro in cui la maggior parte del nostro patrimonio culturale, di libri e di audiovisivi, sarà disponibile su Internet per tutti. Questa è la promessa della Biblioteca Digitale Europea che, entro il 2008, conta di avere due milioni di opere digitali, tra libri, file audio e video, foto e altro. Tali contenuti multimediali saranno resi disponibili attraverso un portale europeo che farà da interfaccia unica per raggiungere le collezioni di opere custodite in forma digitale sui server delle biblioteche nazionali. L'Italia, in materia di digitalizzazione del suo patrimonio, è in anticipo sugli altri Paesi, grazie allo sviluppo della Biblioteca Digitale Italiana concretizzata nel portale Internet Culturale. Oltre a questo progetto, ne abbiamo visti altri quali LaScala DAM, l'Archivio Digitale della Musica Veneta (estesosi fino a poter essere considerato un Archivio Digitale della Musica), il portale ReMI, l'Archivio Storico della Ricordi; tutti molto simili tra loro nelle finalità, ma differenti come implementazione. In un futuro dove si prospetta l'accesso digitale all'intero patrimonio culturale nazionale, questi ultimi progetti citati ne rappresentano solo una piccola e variegata quantità. Sarebbe auspicabile la loro integrazione in un sistema più ampio, quale può essere appunto l'Internet Culturale, nella prospettiva europea di una biblioteca digitale.

Si ritiene che l'impatto sociale di quest'iniziativa sia notevole, soprattutto per studenti e ricercatori. Questa rappresenta un passo avanti per chi vive lontano dai grandi centri, dove ci sono le biblioteche più importanti, ed è svantaggiato nello studio. La produzione cartacea, comunque, non sparirà: la carta continua a conservare alcuni vantaggi rispetto al digitale. Quando devono essere consultati tre o quattro volumi nello stesso momento, ad esempio, è più facile tenerli sulla scrivania piuttosto che visualizzarli sullo schermo. In un contesto di accesso pubblico globale, la produzione dei libri potrebbe subire cambiamenti, grazie allo sviluppo dell'editoria digitale. Per esempio, gli editori potrebbero pubblicare on-line e poi, solo su richiesta,

stampare e rilegare le copie per gli utenti.

L'integrazione dei differenti archivi nazionali porterà ad avere delle enormi quantità di contenuti digitali a cui accederanno sempre più persone. In tale contesto, le strategie di ricerca classiche non saranno più sufficienti a garantire un accesso funzionale ai contenuti. Quindi, accanto alle classiche ed assodate ricerche per metadati, andrà affiancata una nuova tipologia di interrogazione per contenuti. Quest'ultima permetterà di navigare all'interno della base di dati sfruttando i legami semantici che uniscono le diverse opere.

La sfida tecnologica, ma anche sociale, è ostica: bisognerà coordinare il lavoro di tanti Paesi, mettere insieme idee ed esperienze eterogenee e districarsi tra le differenti legislazioni che cambiano da una frontiera all'altra. D'altra parte il progetto è ambizioso: alla fine darà la luce al primo contenitore che sia la summa ideale della produzione culturale europea.

Bibliografia

- [1] Archivio Digitale della Musica Veneta,
<http://193.205.1.21:8080/admv/index.jsp>
- [2] Archivio Storico Ricordi,
<http://www.braidense.it/ricordi.html>
- [3] Biblioteca Marciana di Venezia,
<http://marciana.venezia.sbn.it>
- [4] Biblioteca Nazionale Universitaria di Torino,
<http://www.bnto.librari.beniculturali.it>
- [5] Casanova E., Archivistica, Lazzeri, Siena, 1928.
(Reperibile all'indirizzo: <http://archivi.beniculturali.it/Biblioteca/indexCasanova.html>)
- [6] Discoteca di Stato,
<http://www.dds.it>
- [7] Fondazione Accenture,
<http://www.fondazioneaccenture.it/dam.html>
- [8] Haus G., Ludovico L.A., *The Digital Opera House: an Architecture for Multimedia Databases*, Journal of Cultural Heritage, 2005.
- [9] Laboratorio d'Informatica Musicale,
<http://www.lim.dico.unimi.it>
- [10] Ludovico L.A., *Applicazione DAM Centrale, Manuale Utente*.
- [11] Messina M., *Progetto di Archivio Digitale della Musica Veneta: Stato dell'Arte*,
<http://www.cab.unipd.it/eventi/messina.php3>
- [12] Messina M., *Prospettive per la digitalizzazione dei Documenti Musicali: dall'Archivio Digitale della Musica alla Rete della Musica Italiana*,
<http://eprints.rclis.org/archive/00005718/01/ReMI031107FirenzeAtti.pdf>
- [13] OPAC ReMI,
<http://193.206.197.3:8080/opac/index.jsp>

- [14] Portale Internet Culturale,
<http://www.internetculturale.it>
- [15] Standard Dublin Core,
<http://dublincore.org>
- [16] Standard MAG,
http://www.iccu.sbn.it/MAG/MAG_2.0/MAG_sito_reference/Reference.html
- [17] Standard MAG,
<http://www.iccu.sbn.it/comimag.htm>
- [18] Standard UNIMARC,
<http://www.ifla.org/VI/3/p1996-1/sec-uni.htm>