

Francesco Baldi

Il progetto “Archivio Digitale” della Discoteca di Stato – Museo dell’Audiovisivo.

La Discoteca di Stato e Museo dell’Audiovisivo ha in corso il progetto di riversamento in digitale del materiale audiovisivo per la “realizzazione della Discoteca Digitale”

Il progetto di digitalizzazione è legato ad uno dei più significativi sviluppi tecnologici del periodo: la convergenza di strumenti che consentono di produrre e allo stesso tempo di accedere in rete ad una varietà enorme di informazioni. L’utilizzazione futura dell’informazione viene così garantita a dispetto della fragilità fisica dei supporti e dei condizionamenti dovuti ai programmi e alle attrezzature per l’uso, che sono in continua evoluzione e quindi soggette ad “obsolescenza tecnologica”.

Il riversamento in digitale risulta operazione strategica per la tutela e valorizzazione del patrimonio della Discoteca di Stato e Museo dell’Audiovisivo, volto a fornire un supporto per il soddisfacimento delle esigenze informative e di servizio espresse dalla propria utenza.

Le attività di riversamento sono iniziate nel corso del 2005 e l’obiettivo finale del progetto è quello di riversare l’intero patrimonio audiovisivo dell’Istituto.

L’archivio della Discoteca di Stato – MAV è costituito da circa 320.000 supporti audiovisivi (dischi a 78 giri, cilindri di cera, lacche e dischi a 33 giri, dischi a 45 giri, nastri magnetici audio, CD audio, cassette VHS, BETACAM, dischi di cartone, musicassette, fili metallici ecc.) e da circa 1.200 fotografie storiche, 150 film in pellicola e circa 190.000 diapositive.

In una prima fase del progetto sono state avviate le attività di riversamento in digitale dei dischi storici a 78 giri, dei CD audio e dei nastri magnetici. Entro la fine dell’anno 2007 si inizieranno i lavori di riversamento dei cilindri di cera e dei dischi a 33 giri.

La digitalizzazione consiste nel riversamento di tutti i supporti analogici in formato digitale sotto forma di file che vengono caricati nella TecaDigitale. Contestualmente alla digitalizzazione, vengono anche generate le versioni compresse (file audio mp3 in media e bassa qualità) del materiale e vengono acquisite le immagini relative ai documenti cartacei allegati (copertine, libretti, fotografie, schede tecniche, ecc.). Il riversamento in digitale avviene utilizzando delle stazioni di trascrizione, i documenti da trascrivere sono individuati sulla base delle informazioni gestionali e descrittive fornite dal catalogo. I file prodotti da questo processo vengono memorizzati nella Teca Digitale, che è costituita da diversi sistemi di gestione e da un sistema robotizzato per l’archiviazione di grandi moli di dati. Il supporto di memorizzazione finale per i dati digitali è costituito da una cassetta dati di tipo LTO2 con capacità individuale di 200 GigaByte.

IL PROCESSO DI “TRASCRIZIONE DIGITALE” DEI SUPPORTI AUDIOVISIVI

Vediamo da vicino, in breve, il ciclo completo di lavorazione in atto sul materiale presente in Discoteca di Stato e Museo dell’Audiovisivo per conseguire la generazione della teca digitale e garantire i servizi di conservazione/tutela e l’erogazione dei servizi di ascolto.

I documenti descritti nel catalogo sono prelevati dal magazzino e resi disponibili per le operazioni di riversamento.

Di ciascun documento viene verificato lo stato di conservazione ed eventualmente pianificata una operazione di pulizia/restauro.

I documenti vengono inviati alle stazioni di trascrizione dove tutti i supporti analogici vengono riversati in formato digitale e caricati nella TecaDigitale. Contestualmente alla digitalizzazione vengono scansionati i documenti cartacei allegati. I documenti riversati vengono ricollocati in magazzino e da questo momento saranno usati esclusivamente in caso di particolari esigenze che prevedano l'utilizzo del supporto originale.

Le tipologie di supporti attualmente in possesso, che si stanno riversando in formato digitale per essere rese disponibili per la consultazione, sono:

- Dischi 78 gg, (musica classica, leggera, jazz, voci storiche)
- Nastri ¼", (musica classica, tradizioni popolari, teatro, voci storiche)
- Compact Disc, (musica classica, leggera, jazz, tradizioni popolari, teatro, voci storiche)

Il flusso di lavorazione: un esempio concreto - i dischi a 78 giri

Il sistema è costituito da una catena di riversamento composta da uno o più apparati analogici in parallelo collegati ad un convertitore analogico/digitale, a sua volta collegato tramite un'interfaccia digitale ad un personal computer, e da un secondo PC collegato ad uno scanner per l'acquisizione delle immagini relative alle etichette e a eventuali materiali allegati.

L'operatore viene guidato, da un software dedicato, in tutte le fasi di lavoro, mentre il sistema gestisce automaticamente e in maniera trasparente l'indirizzamento dei file prodotti al sistema di archiviazione centralizzato.

Per ogni tipo di supporto è stata individuata una particolare procedura al fine di ottimizzare i tempi di lavoro pur garantendo la sicurezza e la qualità del risultato.

Al termine delle attività di riversamento giornaliera, il materiale viene riconsegnato al personale della Discoteca di Stato e Museo dell'Audiovisivo incaricato per il controllo dei supporti originali e la collocazione.

Attività preliminari

- Individuazione dei supporti da trascrivere
Sulla base dei dati contenuti nel catalogo gestionale si individuano i lotti di documenti da trascrivere. In questa fase si effettua anche un controllo sulla coerenza dei dati di inventario inseriti in catalogo e la consistenza dei documenti e si verifica lo stato di conservazione dei documenti.
- Presenza nel catalogo informatizzato
Per una gestione razionale del processo è indispensabile che tutti i documenti da digitalizzare siano descritti nel catalogo gestionale. Infatti è importante che il catalogo raccolga tutti i dati sui documenti che sono in fase di digitalizzazione, sui documenti che sono già stati riversati in formato digitale e infine sia in grado di fornire informazioni sui documenti ancora da digitalizzare. Si producono le liste cartacee dei documenti da digitalizzare, comprensivi di moduli di richiesta per i magazzini con codici a barre di inventario e collocazione.
- Materiale cartaceo da allegare ai documenti da trascrivere:

- moduli di richiesta per il materiale da prelevare al magazzino con i codici a barre contenenti il N. Documento e il N. di Inventario da allegare ai supporti da inviare alle stazioni di trascrizione.
- liste di prelievo (picking list) sulle quali riportare annotazioni sullo stato di conservazione e sull'integrità dei dati (dischi mancanti o rotti, errori nelle collocazioni). Le liste di prelievo vengono compilate a partire dai dati presenti nel catalogo in base ai criteri sopra descritti. Nelle liste viene segnalato l'esito di ogni fase della digitalizzazione (prelievo, scheduling, trascrizione), fino alla ricollocazione nei magazzini.

Attività al magazzino

- **Prelievo del materiale dai magazzini**
Per garantire il livello di produttività richiesto agli operatori della trascrizione è necessario prelevare e ricollocare giornalmente oltre 100 supporti (dischi a 78 giri). La movimentazione di una tale quantità di supporti richiede un adeguato numero di collaboratori. Tale attività viene effettuata dal personale ausiliario coadiuvato dal personale interno.
- **Controllo dello stato di conservazione:** sulla base delle liste di prelievo si controlla lo stato di conservazione e si individuano le copie uniche (escludendo le doppie copie) da trascrivere. Nell'eventualità che un documento non sia idoneo al riversamento deve essere immediatamente escluso dal flusso di lavorazione. Se disponibile si procede al riversamento della copia.
 - Il materiale viene prelevato dagli scaffali, lasciando al suo posto i moduli di richiesta.
 - Etichettatura dei supporti da trascrivere applicando le etichette con codici a barre.
 - Assegnazione e trasporto alla stazione di trascrizione: il personale addetto alla movimentazione giornalmente provvede alla consegna dei supporti da digitalizzare agli operatori e alla raccolta dei supporti già digitalizzati, incluse le liste di prelievo che contengono le informazioni sui supporti riversati e l'indicazione dei supporti che hanno presentato problemi e sono stati esclusi dalla digitalizzazione.
 - Lavaggio dei dischi: i supporti a 78 giri vengono lavati, prima del riversamento, con appositi macchinari.

Stazione Scheduling

La fase di scheduling prevede l'inserimento dei dati sui "task" che verranno inviati alle diverse stazioni di trascrizione. In questa fase si inseriscono i dati relativi a codice di documento, numeri di inventario identificativi del supporto, numero dei supporti che compongono il documento, eventuali metadati descrittivi provenienti dal catalogo, dati sulla digitalizzazione delle immagini allegate (copertine, libretti, etichette dei supporti, ecc.). Il codice documento rappresenta l'identificativo univoco della trascrizione (denominato codice di "Album").

- Il materiale prelevato viene trasportato alla stazione di scheduling.
- Il codice documento e i numeri di inventario si inseriscono utilizzando etichette con codici a barre e pistole a lettura ottica.
- I documenti vengono trasferiti alle diverse stazioni di trascrizione.

Trascrizione

Vengono eseguiti i diversi task previsti; il materiale che presenta problemi viene escluso dalla catena di trascrizione e trattato a parte.

Digitalizzazione e scansione degli allegati

L'operatore di trascrizione riceve i supporti assegnati e l'elenco degli stessi e procede alla digitalizzazione dell'audio e alla scansione degli allegati secondo le istruzioni ricevute e riconsegna i documenti trattati al personale addetto alla movimentazione, allegando la documentazione prescritta.

Processamento batch e archiviazione nella TECA

Gli oggetti digitali generati durante il processo di trascrizione vengono automaticamente processati e trasferiti nella TECA Digitale per l'archiviazione. Durante tale fase vengono prodotte le copie per la consultazione in media e bassa qualità destinate all'utenza locale ed al servizio "WEB" via Internet.

Aggiornamento del catalogo con le informazioni sulle trascrizioni effettuate

I dati delle trascrizioni effettuate vengono riportati nel catalogo in linea della Discoteca di Stato; si riportano anche tutte le segnalazioni derivanti dalle varie fasi dell'intero processo di digitalizzazione.

Ricollocazione

Il materiale viene ricollocato ai magazzini.

Descrizione dell'architettura del sistema

L'architettura del Sistema di Digitalizzazione e Gestione dei Documenti Sonori e Audiovisivi della Discoteca di Stato e Museo dell'Audiovisivo (SDGDS) è costituita da quattro componenti principali:

- CATALOGO GESTIONALE: fornisce informazioni catalografiche e meta-dati tecnici relativi ai documenti archiviati. Permette la documentazione di quanto trascritto.
- Sistema TECA DIGITALE: costituisce l'archivio dei dati (documenti sonori ed allegati) in formato numerico. Supporta inoltre la segmentazione dei documenti sonori.
- Stazioni di Trascrizione: permettono il popolamento dell'archivio numerico, a partire dai supporti originali, sia analogici che digitali.
- OPAC-DDS/MAV: l'On-Line Public Access Catalogue, permette la consultazione ed ascolto dei documenti via Web.

Sistema CATALOGO GESTIONALE

Il catalogo, di proprietà della D.d.S.- MAV, è costituito da un database relazionale contenente le schede relative ai documenti sonori e video di proprietà della Discoteca di Stato-Museo dell'Audiovisivo.

Un software appositamente sviluppato consente la gestione (ricerca, inserimento, cancellazione, variazione) di schede elettroniche, la gestione inventariale dei documenti, il governo di flussi di dati in ingresso e in uscita a partire da formati standard (Unimarc e metadati MAG 2.01).

Il catalogo dei documenti audiovisivi guida inoltre l'intero processo della trascrizione: fornisce informazioni sui supporti da trascrivere, memorizza i dati sulle trascrizioni digitali effettuate e fornisce i metadati descrittivi da collegare agli oggetti digitali.

Sistema TECA DIGITALE

La Teca Digitale, memorizza tutti gli oggetti digitali prodotti nel corso delle attività di riversamento. Questo sistema permette la gestione di grosse moli di dati multimediali rendendoli disponibili alla fruizione e offre servizi di archiviazione e movimentazione di dati multimediali sia su base della rete locale che su Internet.

I dati prodotti dalle varie catene di riversamento in digitale dei documenti audiovisivi vengono gestiti da questo sistema.

Nel seguito, è illustrata la struttura attuale del sistema "Teca Digitale" le cui componenti principali sono:

- Sistema Gestione Archivio
- Area di cache
- Robotica
- Sottosistema MagTeca

Sistema Gestione Archivio

Tutti i dati multimediali in transito da e per il sistema robotizzato di archivio digitale e tutti i file accessori necessari alla Teca Digitale sono gestiti da questo sottosistema. Quest'ultimo controlla non solo il flusso dei dati dentro il sottosistema di archiviazione (comanda cioè i registratori della robotica), ma gestisce anche tutte le linee attraverso cui i dati memorizzati sono distribuiti agli utenti e/o da questi ritornano per essere memorizzati (catene di riversamento dei documenti audiovisivi).

Questo sistema gestisce fisicamente i dati informatici, l'organizzazione e lo spostamento dei file sui vari supporti di memorizzazione di massa.

Le operazioni di controllo sull'integrità dei dati e di riversamento periodico dei supporti di memorizzazione sono gestite in modo automatico e sono programmabili attraverso l'utilizzo di procedure di configurazione del sistema.

Area di cache

L'area di cache costituisce la memoria di transito da e per l'archivio robotizzato su nastro magnetico (LTO2). Nell'area di cache risiedono i brani di più recente acquisizione e/o quelli richiesti con maggiore frequenza.

Il rapporto tra la dimensione dell'area di cache e l'archivio su nastro, dipende dalla tipologia delle richieste in lettura e dal carico di materiale da archiviare giornalmente.

Sistema Robotizzato

L'Armadio Robotizzato costituisce l'archivio su nastro magnetico dei dati multimediali in grado di supportare il traffico in ingresso (nuovi brani da archiviare) ed in uscita (richieste dagli utenti) e crescere in conseguenza della crescita continua dei brani archiviati.

Elemento sostanziale è la scalabilità del sistema. Il sistema di storage selezionato (ADIC Scalar 1000) consiste di un Control Module (ospitante la robotica vera e propria ed un numero limitato di cassette LTO2) più una serie di Expansion Module (dedicati ad ospitare LTO2 ulteriori). La scalabilità consiste nell'affiancare i moduli d'espansione in funzione della dimensione raggiunta dall'archivio digitale.

Sottosistema MagTeca (fruizione interna e in internet dei dati della TECA)

Teca Oracle per la connessione ad "Internet Culturale", conserva i dati multimediali (immagini e audio) in bassa e media qualità, per il servizio di fruizione interno e su rete Internet. E' basata su un modello di archiviazione OAIS (Open Archival Information System) e prevede la consultazione automatica attraverso il protocollo OAI – PMH (Open Archive Initiative, Protocol Metadata Harvesting). Al proprio interno sono memorizzati i metadati in formato standard MAG 2.01 con una sezione Bib contenente gli elementi descrittivi estratti automaticamente dal Catalogo Gestionale.

Teca Oracle compatibile con lo schema MAG 2.01 basata su modello di archiviazione OAIS:

- memorizza i dati digitali in media e bassa qualità per la consultazione
- fornisce il servizio di ascolto su rete Intranet ed Internet
- garantisce le funzionalità di Harvesting (PMH) per la rete NTC (internetculturale).

Stazioni di Trascrizione

Le Stazioni di Trascrizione audio sono anch'esse componenti del sistema Teca Digitale. Sono differenziate per tipologia di supporto. Per i supporti di tipo analogico sono composte dai dispositivi audio e da convertitori analogico digitali. Le stazioni di acquisizione per i supporti digitali prevedono l'acquisizione direttamente in formato digitale.

Le versioni su PC, sono state messe a punto nell'ambito del progetto europeo PRESTO, relativo alla preservazione d'archivi analogici audio, video e film, guidato da BBC, RAI ed INA (Istituto Nazionale dell'Audiovisivo, Francia), con ACS ed ITC/IRST come partner tecnici per la parte audio.

La trascrizione ad alta qualità di tutti i supporti genera file secondo lo standard europeo Broadcast Wave Format (BWF). Il formato di campionamento è generalmente 48 kHz 24 bit stereo salvo casi particolari da definirsi per tutti i supporti analogici (secondo le

necessità per materiale che potrà poi essere sottoposto a restauro sonoro si utilizzerà un formato di campionamento di 96 kHz 24 bit). Il materiale digitale nativo (CD e DAT) viene trascritto mantenendo il formato originale.

Vengono inoltre generate due copie compresse per il browsing: una ad alta qualità ad uso degli utenti presso l'istituto ed una adatta alla consultazione su internet.

La copia mp3 ad alta qualità viene effettuata a 256 kbit/s con codifica MPEG 1 Layer 3, mentre quella a bassa qualità viene effettuata a 32 kbit/s con codifica MPEG 1 Layer 3.

La trascrizione in digitale dei supporti e degli allegati è effettuata adottando le indicazioni dello International Association of Sound and Audiovisual Archives - [I.A.S.A.](http://www.iasa-web.org) (www.iasa-web.org), in conformità allo standard BWF (Broadcast Wave File) prodotto dalla European Broadcasting Union - [EBU](http://www.ebu.ch/pmc_bwf.html) (http://www.ebu.ch/pmc_bwf.html) in collaborazione con la Audio Engineering Society – [AES](http://www.aes.org) (<http://www.aes.org>), accogliendo le indicazioni dello “UNESCO RESOLUTION on DIGITAL PRESERVATION” e le indicazioni derivate dalle attività del progetto europeo [PRESTO](http://prestjoanneum.ac.at/index.asp) Preservation Technology for European Archives (<http://prestjoanneum.ac.at/index.asp>).

OPAC DDS-MAV

Rappresenta l'interfaccia verso l'utente finale per la fruizione dei servizi offerti dall'archivio digitale. Sono disponibili diverse modalità di consultazione (accesso a catalogo sintetico, ricerca specialistica).

- Contiene le informazioni descrittive sui documenti audiovisivi.
- Contiene i collegamenti ai documenti multimediali.
- Permette la ricerca dei contenuti audio e delle descrizioni catalografiche.
- Consente la fruizione dei documenti multimediali attraverso rete Intranet e Internet.

L'alimentazione del sistema OPAC si realizza utilizzando gli standard definiti a livello internazionale:

- *Unimarc* Per il riversamento dei dati di catalogo si utilizza il formato di scambio UNIMARC nella versione definita dal Gruppo Unimarc Italiano per la Musica rilasciata dall'ICCU.
- *Mag 2.01* Per i metadati è stato adottato lo standard MAG emanato dal gruppo italiano di studio sui metadati coordinato dall'ICCU (vers. 2.01 che comprende anche le sezioni audio e video).