

UNIVERSITÀ DEGLI STUDI DI MILANO

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI

CORSO DI LAUREA TRIENNALE IN SCIENZE E TECNOLOGIE DELLA
COMUNICAZIONE MUSICALE



Standard per la Catalogazione dell'Informazione
Audio nei Database Multimediali

Tesi di Laurea di:
Paolo CERIANI
Matr. n. 653240

Relatore:
Prof. Goffredo HAUS

Correlatore:
Dott. Adriano BARATÈ

Correlatore:
Dott. Luca A. LUDOVICO

ANNO ACCADEMICO 2004-2005

Sommario

Capitolo I - Introduzione al progetto e organizzazione del lavoro	1
Il progetto di database multimediale	2
Partecipanti al progetto: la Discoteca di Stato ed il portale Internetculturale.....	2
La Discoteca di Stato	2
Il portale Internetculturale.....	3
Le tre fasi del progetto.....	4
Prima fase: digitalizzazione dei dati audio	4
Seconda fase: creazione del database locale	5
Terza fase: possibilità di accesso remoto attraverso il portale Internetculturale.....	6
Progettazione e struttura del database	6
Capitolo II - Una prima soluzione: il formato UNIMARC	11
Breve storia dell'UNIMARC	11
Finalità e scopi dell'UNIMARC	12
Mantenimento del formato.....	12
Notazioni convenzionali	13
Struttura del formato.....	13
Etichetta del Record.....	14
Directory	14
Campi a lunghezza variabile	15
Campi obbligatori.....	16
Lunghezza dei record.....	16
Record Linking.....	17
Set di caratteri.....	17
Informazioni ridondanti	17
Trattamento di script differenti.....	18
Ordine dei sotto-campi.....	20
Caratteri utilizzati	21
Forma e contenuto dei dati.....	21
Etichetta del record e campi dati: informazioni generali.....	21

L'Etichetta del Record	21
Elementi di dati a lunghezza fissa	22
Breve descrizione dei campi obbligatori	28
001 – Identificatore del Record	29
100 – Dati Generali di Elaborazione	29
200 – Titolo e dichiarazione di responsabilità	37
801 – Fonte originale	40
Campi di interesse musicale.....	41
013 – Numero Musicale Internazionale Standard (ISMN)	42
021 – Numero di deposito legale.....	43
071 – Numero dell'editore (Musica e registrazioni sonore)	43
125 – Campi dati codificati: registrazioni sonore e musica stampata	45
126 – Campi dati codificati: registrazioni sonore – attributi fisici.....	47
127 – Campi dati codificati: durata delle registrazioni sonore e della musica stampata	55
128 – Campi dati codificati: esecuzioni e partiture musicali	55
323 – Note sugli interpreti (materiale video e registrazioni audio).....	61
Un esempio completo	61
Conclusioni	62
Capitolo III - Un'alternativa all'UNIMARC: il formato MAG.....	64
Introduzione	64
Struttura del formato.....	64
Sezione GEN	66
Sezione BIB.....	68
Sezione AUDIO.....	71
Concludendo: un esempio di record MAG	76
Capitolo IV - Rappresentazione dei metadati	78
Alcune considerazioni.....	80
Conclusioni	81
Riferimenti.....	83
Ringraziamenti	84

Capitolo I - Introduzione al progetto e organizzazione del lavoro

L'idea della digitalizzazione dei beni musicali nasce dalla necessità di salvaguardare le opere di carattere e, più in generale, di argomento musicale, ai fini di una più lunga conservazione non solo dell'oggetto fisico dell'opera (sia essa disco, nastro, partitura, libretto o altro), ma anche e soprattutto del contenuto intellettuale, delle informazioni e delle conoscenze che vi sono contenute. Quando nell'ambito del restauro dei beni musicali si è avvertito il bisogno di uno strumento che superasse l'inevitabile anche se a volte sensibilmente riducibile degrado che colpisce qualsiasi tipo di supporto informativo, si è fatto ricorso alla digitalizzazione, quindi alla fissazione su elaboratore dei contenuti (e, in fondo, dell'“aspetto”) delle opere, il che ha portato innumerevoli benefici, non ultimo quello di poter intervenire nell'eliminazione dei “difetti” causati da tempo, usura o cattiva conservazione senza pregiudicare l'integrità dei supporti. Si è quindi arrivati ad un punto in cui la grande mole di oggetti digitalizzati dei tipi più diversi (fotografie, testi acquisiti tramite scanner, tracce audio e video) andava ordinata, gestita e, cosa più importante, resa fruibile da chiunque. E' in quest'ottica che si inserisce lo strumento del database, più precisamente del database multimediale che, come suggerito dal nome, serve ad accorpare vari tipi di contenuti di natura diversa relativi a una stessa opera, in modo che sia possibile avere accesso a tutte le informazioni disponibili su un determinato oggetto in maniera intuitiva e sintetica, sotto forma di tabella nella quale sono presenti campi che riportano direttamente i dati richiesti per una data opera, (ad esempio i campi testuali), o che rinviano a contenuti multimediali ad essa riferiti¹. Non solo: grazie ad alcuni programmi, detti filtri, è possibile specificare, tramite l'uso di form, quali risultati si vuole siano ritornati all'utente, permettendo così di avere accesso solamente alle informazioni di interesse.

La creazione di basi di dati strutturate per un tale compito non è facile, sia per la diversità dei contenuti che si vogliono accorpare, sia per la lenta espansione dei progetti di digitalizzazione, che include tempi diversi per ciascun medium (i dati testuali già presenti in forma digitale sono molti di più di quelli audio, perché più facili da descrivere e da gestire), sia infine per il problema della descrizione dei vari contenuti informativi presenti in ciascun oggetto. L'utilizzo in questo senso dei metadati, ossia dei “dati che descrivono altri dati”, permette di semplificare notevolmente e di rendere più veloce il processo di

¹ Cfr. [Subrahmanian 1998]

catalogazione e archiviazione (digitale, ma anche tradizionale), e grazie ai formati di interscambio, la circolazione di “blocchi informativi”, quindi di intere collezioni non di opere, ma di dati sulle opere, è ulteriormente favorita.

Questo elaborato si propone di descrivere due dei formati standard più utilizzati per la rappresentazione, la gestione e l’interscambio di metadati, fornendo inoltre una panoramica su un progetto di interscambio di dati bibliografici, musicali e multimediali in genere tra la Discoteca di Stato ed il portale Internetculturale , illustrando una possibile struttura di database multimediale (nel suo contesto d’utilizzo), ed analizzando infine le possibilità di integrazione dei metadati contenuti nel database offerte da ciascun formato.

Il progetto di database multimediale

In questo capitolo verranno innanzitutto presentati i due enti che partecipano al progetto preso in esame, tra i quali dovrà avvenire dunque lo scambio di metadati finalizzato alla messa a disposizione degli utenti del patrimonio di materiale audio della Discoteca di Stato attraverso il portale Internetculturale. Seguirà poi l’illustrazione dei punti chiave del progetto, nella quale verranno inoltre delineate le caratteristiche salienti del database che potrebbe essere realizzato.

Partecipanti al progetto: la Discoteca di Stato ed il portale Internetculturale

Prima di entrare nei dettagli nella descrizione della struttura del database, è opportuno fornire una caratterizzazione, almeno a grandi linee, dei soggetti all’interno dei quali avverrà (almeno all’inizio) lo scambio di dati.

La Discoteca di Stato

La Discoteca di Stato (www.dds.it), fondata nel 1928, è la principale collezione pubblica italiana di documentazione sonora, e garantisce la conservazione e la fruizione del patrimonio sonoro nazionale sia edito che inedito.

La collezione della Discoteca comprende raccolte e collezioni musicali di vari generi (classica, jazz, rock, folkloristica), narrativa orale e discorsi storici, conservati su vari tipi di supporto (la collezione comprende circa 230.000 unità, tra cilindri di cera, fili metallici, dischi, nastri e compact disc). I pezzi catalogati pervengono tramite depositi legali, acquisto, donazioni private e pubbliche e registrazione di manifestazioni culturali.

All'ampliamento ed all'apertura della collezione a nuovi tipi di formato ha contribuito la Legge 237 del 1999, estendendo le prerogative di raccolta, conservazione e fruizione anche al materiale realizzato e prodotto su supporti video, multimediali e su ogni altra nuova tecnologia in futuro disponibile.

Con la Legge 237 del 1999 è stato costituito il Museo dell'Audiovisivo che ha esteso le prerogative di raccolta, conservazione e fruizione anche al materiale realizzato e prodotto su supporti video, multimediali e su ogni altra nuova tecnologia in futuro disponibile.

L'accesso ai documenti conservati nella Discoteca di stato è libero, ed è possibile attraverso la messa a disposizione degli utenti di un catalogo cartaceo a schede e tramite un catalogo in linea; l'intero patrimonio e le informazioni catalografiche sono in corso di digitalizzazione al fine della loro pubblicazione e messa a disposizione su internet nel minor tempo possibile.

Il portale Internetculturale

Molto più recente è senza dubbio la nascita del Portale della Biblioteca Digitale Italiana e Network Turistico Culturale, o più semplicemente Internetculturale, che è comparso in rete nel 2005, con lo scopo di favorire la promozione e la valorizzazione del patrimonio culturale sia nazionale che internazionale, migliorandone la conoscenza e la fruibilità da parte del pubblico. Questo progetto è stato promosso dalla Direzione Generale per i Beni Librari e gli Istituti Culturali (DGBLIC) del Ministero per i Beni e le Attività Culturali e realizzato dall'Istituto Centrale per il Catalogo Unico (ICCU) che ha implementato alcune funzionalità e integrato i servizi del portale con quelli già attivi nel proprio sito.

Attualmente sul sito di Internetculturale sono disponibili agli utenti alcuni servizi informativi di vari generi:

- Ricerca bibliografica: tramite il sistema OPAC SBN è possibile effettuare ricerche di materiale bibliografico su opere a partire dall'inizio della stampa fino ai giorni

nostri, possedute da 2300 biblioteche italiane operanti in diversi ambiti. E' possibile inoltre effettuare una ricerca all'interno dei cataloghi storici, che permettono la consultazione dei cataloghi (a scheda e a volume) digitalizzati di 35 biblioteche italiane;

- Collezioni digitali: in questa sezione vengono descritti i fondi digitalizzati o in corso di digitalizzazione presso archivi, biblioteche e altre istituzioni culturali in Italia. Ciò per consentire l'identificazione e la localizzazione delle collezioni digitalizzate, e dove possibile, la loro consultazione on line;
- Percorsi culturali: sul sito del Portale vengono proposti alcuni percorsi culturali (in più lingue) su diversi argomenti: mostre virtuali multimediali, itinerari scientifico-culturali, percorsi monografici ipertestuali sulle vite e sulle opere di importanti letterati e musicisti, e ricostruzioni 3D di ambienti storici o artistici;
- Gestione della conoscenza: Mediante il sistema informatico di Knowledge Management (KM) è possibile organizzare i contenuti informativi tratti da siti, dal catalogo OPAC SBN e da documenti digitalizzati dotati di record MAG sotto forma di strutture gerarchiche formate da categorie e sottocategorie correlate tra loro e relative ad una stessa area tematica.

Le tre fasi del progetto

Il progetto di digitalizzazione e messa a disposizione del pubblico dei beni audio della Discoteca di Stato si svolgerà in tre fasi principali:

- Digitalizzazione dei dati
- Creazione del database locale della Discoteca di Stato
- Accesso remoto attraverso il portale Internetculturale

Prima fase: digitalizzazione dei dati audio

I supporti audio a disposizione dovranno essere per prima cosa riversati in forma digitale, effettuando una scelta sui formati file da utilizzare e sulla qualità delle tracce audio che poi diventeranno gli oggetti digitali contenuti nel database. A questo proposito si opterà per una soluzione a più livelli, in pratica per una singola registrazione sonora si avranno tre tipi di file audio, ciascuno contraddistinto da una diversa qualità a seconda della destinazione e della fruizione del file. Innanzitutto si avranno copie del file ad *alta qualità*, memorizzate quindi in formato wave, in un formato compresso senza perdita d'informazione (lossless) e/o in formato compresso con perdita (lossy) a massimo bitrate. I file di tipo wave e compresso senza perdita presentano qualitativamente la massima somiglianza ottenibile con la traccia analogica, e quindi verranno conservati negli archivi della Discoteca di Stato a scopo conservativo; le copie in formato compresso con perdita d'informazione a massimo bitrate verranno invece destinate ad enti ed associazioni culturali, e a tutti coloro che richiederanno (previo acquisto di una licenza) una copia dei file di ottima qualità.

Si avranno poi copie del file a *qualità medio-alta*, dove con questo termine ci si riferisce a file contenenti audio in formato compresso lossy con bitrate non inferiore a 128 kbit/s (qualità CD). Questa versione, adatta ad essere scaricata dalla rete, verrà riservata all'acquisto da parte del pubblico.

Si produrrà poi un'altra versione del file, a *bassa qualità*, codificata in formato compresso lossy a basso bitrate. Questo file rappresenterà la tipica "preview" del file, avente durata limitata (inferiore a 30 secondi), destinata alla valutazione del contenuto del file da parte dell'utente prima di procedere all'acquisto della versione a qualità medio-alta.

Seconda fase: creazione del database locale

Il secondo passo nell'attuazione del progetto sarà quella dell'analisi delle richieste del "cliente" (la Discoteca di Stato), della progettazione concettuale e logica della struttura sulla quale realizzare la base di dati, per poi passare alla realizzazione "fisica".

Dato che l'argomento centrale di questo elaborato sono i metadati, la fase in cui essi entrano in gioco (sotto forma di attributi delle entità concettuali), ossia la fase di progettazione logica con la definizione di quali parti degli standard di metadati utilizzare

per definire gli attributi verrà poi ripresa ed approfondita nel capitolo 4; nella prossima sezione verrà invece descritta l'architettura generale del database.

Terza fase: possibilità di accesso remoto attraverso il portale Internetculturale

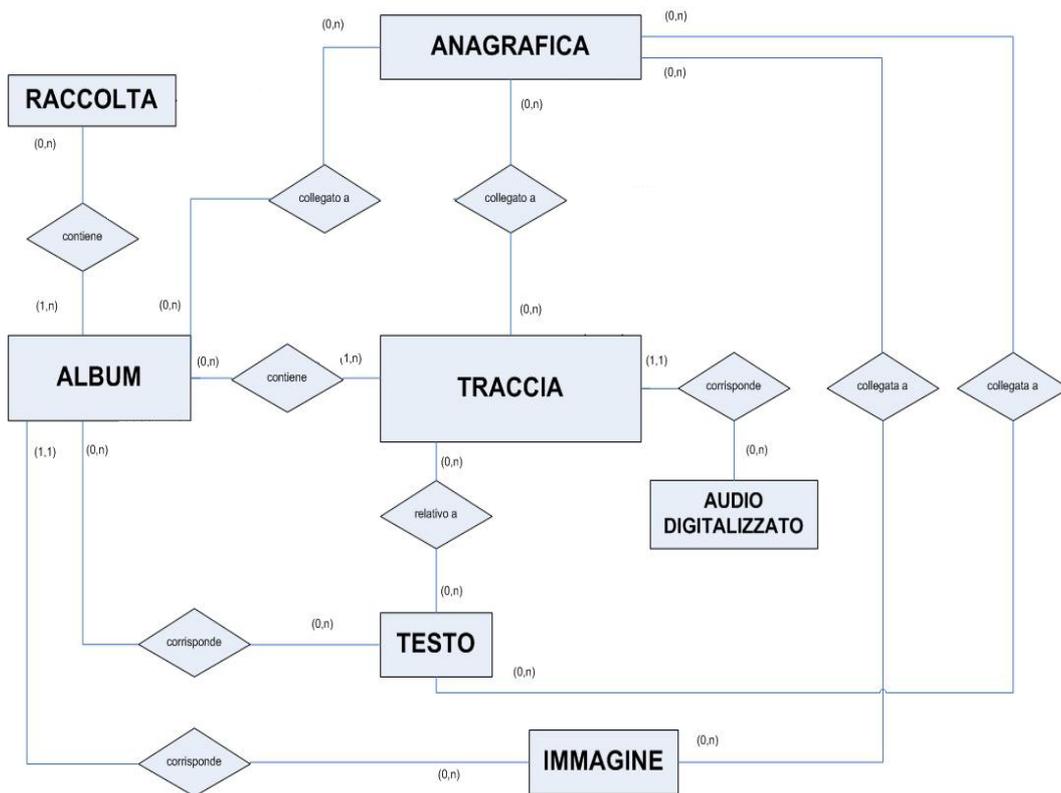
Quando infine il database verrà completato (o almeno, avrà acquisito una certa dimensione ed importanza di contenuti), il passo successivo sarà quello di trasportare le informazioni dagli archivi digitali della Discoteca di Stato a quelli di Internetculturale. Questo perché non è pensabile un accesso remoto in tempo reale al database della Discoteca, in quanto le informazioni complesse contenute in ogni record e che verrebbero ritornate agli utenti in risposta alle loro ricerche dovrebbero essere create dinamicamente; in pratica la ricerca dovrebbe venire inoltrata tramite form presenti sul sito Internetculturale, i dati richiesti dovrebbero a loro volta essere inseriti in una richiesta inoltrata in seguito al database centrale della Discoteca di Stato, che si dovrebbe poi occupare di trovare i contenuti, "filtrarli" per ritornare solo le informazioni richieste, organizzare visivamente (e logicamente) la struttura dei record ed infine inviarli ad Internetculturale, che dovrebbe poi provvedere alla loro presentazione grafica ed al ritorno delle informazioni richieste all'utente. Questa soluzione nel campo dei database multimediali, ed ancor più se legata strettamente all'ambito audio, con file di dimensioni a volte considerevoli e necessità di ricerche a volte complesse, non è praticabile. Invece, si prevede come già detto il trasporto dei dati da un archivio all'altro, utilizzando formati di interscambio di metadati per descrivere il contenuto degli oggetti che vengono scambiati. In questo modo, il portale Internetculturale possiederà una copia del database della Discoteca di Stato, e potrà soddisfare le ricerche del pubblico in maniera autonoma. Naturalmente in futuro questi database potranno poi interagire tra loro per garantire l'aggiornamento dei dati e delle collezioni tra i due archivi.

Progettazione e struttura del database

Prima di procedere alla realizzazione fisica di un database (non solo di tipo multimediale) è necessario innanzitutto effettuare un'analisi dei requisiti che il database dovrà avere per

soddisfare le esigenze e le richieste del “cliente”, o comunque del soggetto dal quale poi verrà utilizzato². Nel nostro caso questo soggetto è rappresentato dalla Discoteca di Stato, e dalla definizione data precedentemente di quest’ente si può evincere che il sistema dovrà supportare non solo la memorizzazione di una grande quantità di dati al fine di una più funzionale conservazione, ma anche la possibilità di una loro esportazione verso altri portali internet attraverso i quali i beni digitalizzati saranno messi a disposizione del pubblico.

Il secondo passo è una progettazione di tipo concettuale, nel quale si individuano e si definiscono i soggetti che parteciperanno alle varie “attività” del database. Questi soggetti vengono chiamati *entità* del database, e sono legati tra di loro attraverso delle *relazioni*, che specificano i legami che uniscono tra loro le entità. Entità e relazioni possono venire visualizzate graficamente tramite uno schema concettuale, e nella figura seguente viene riportato lo schema proposto per il nostro progetto:



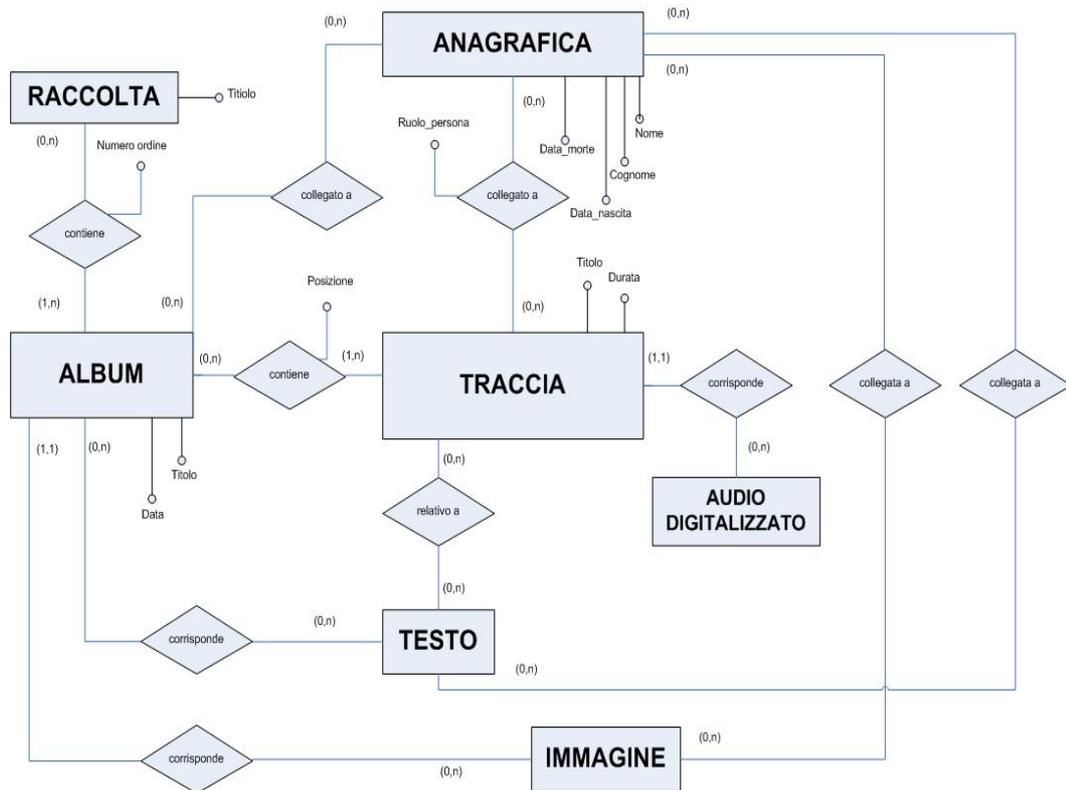
² Vedi anche [Prabhakaran 1997]

Per definizione, le entità sono classi di oggetti del mondo reale di interesse per l'applicazione, collegate tra loro attraverso legami logici o relazioni³. In questo schema si possono individuare 7 entità e 10 relazioni che le uniscono. Le entità sono:

- RACCOLTA (un certo numero di album raggruppati in base ad un criterio comune)
- ALBUM (un singolo album contenente un certo numero di tracce audio)
- TRACCIA (una singola traccia audio)
- AUDIO DIGITALIZZATO (file che contiene effettivamente la registrazione sonora)
- TESTO (materiale cartaceo di accompagnamento all'opera, oppure immagini contenenti materiale testuale, ad esempio scansioni di partiture o libretti)
- IMMAGINE (immagini relative all'opera, per esempio fotografie delle rappresentazioni, delle locandine o dei costumi di scena)
- ANAGRAFICA (scheda anagrafica contenente i dati dei partecipanti all'opera, per esempio l'autore della musica, il librettista, ecc...).

Per ciascuna entità poi vanno definiti degli *attributi*, ovvero delle caratteristiche dell'entità che siano d'interesse per la nostra applicazione. In altre parole, la scelta degli attributi riflette il livello di dettaglio al quale vogliamo rappresentare le informazioni sulle entità o sui legami logici.

³ Vedi [Ramakrishnan – Gehrke 2004]



Quest'immagine mostra lo schema concettuale del database per il quale sono stati definiti gli attributi basilari ai quali poi se ne andranno ad aggiungere altri in base alle esigenze ed al livello di descrizione richiesto. Come si può notare, anche le relazioni possono avere degli attributi che le specificano meglio.

Gli attributi definiti per ciascuna entità o relazione sono:

<i>Entità</i>	<i>Attributi</i>
RACCOLTA	Titolo
ALBUM	Titolo Data
TRACCIA	Titolo Durata
ANAGRAFICA	Nome Cognome Data di nascita Data di morte
<i>Relazione</i>	<i>Attributi</i>
RACCOLTA - ALBUM	Numero d'ordine (inteso come il numero dell'album all'interno nella raccolta)
TRACCIA - ALBUM	Posizione (posizione della traccia all'interno dell'album)
ANAGRAFICA - TRACCIA	Ruolo (ruolo della persona descritta nella scheda anagrafica all'interno della traccia)

All'atto pratico questi attributi rappresentano dei metadati "esterni" all'oggetto digitale, in quanto non sono presenti al momento della sua creazione (metadati "interni", per esempio l'indicazione la risoluzione di un'immagine, o l'autore di un file di testo), ma vengono assegnati in fase di catalogazione, sotto forma di schede catalografiche o di record in un particolare formato di metadati, argomento centrale dei prossimi due capitoli.

Capitolo II - Una prima soluzione: il formato UNIMARC

In questo capitolo verrà presentato il formato di catalogazione Bibliografica UNIMARC, che permette l'interscambio di dati e metadati bibliografici tra diverse aziende, enti od organizzazioni. Il capitolo si compone di sei sezioni: la prima è un'introduzione dove verrà presentato il formato con alcuni cenni storici, si delinearanno le sue finalità ed i suoi scopi e verranno fornite alcune note sulle convenzioni testuali adottate; nella seconda sezione si approfondirà la conoscenza della struttura del formato, presentandone le parti componenti e le regole generali per la costruzione di record UNIMARC; nelle successive tre sezioni verranno descritti in maniera schematica l'etichetta del record ed i suoi campi dati, i campi obbligatori per ogni record UNIMARC ed, in maniera più approfondita, i campi che all'interno del formato hanno una funzione che si adatta meglio all'impiego per la descrizione dei metadati delle informazioni audio contenute in un database multimediale. Nell'ultima sezione, di conclusione, verranno riepilogati brevemente ed analizzati i concetti principali discussi nel capitolo.

Breve storia dell'UNIMARC

Il formato UNIMARC nasce negli anni '60, quando la Library of Congress attua il progetto di convertire il proprio catalogo, a quel tempo ancora conservato in archivi cartacei, in formato digitale su computer, anche per favorire l'interscambio di dati bibliografici con le biblioteche ad essa collegate. A tale scopo viene sviluppato il primo formato MARC (MACHINE READABLE CATALOGUING), da cui nel 1973 deriverà la prima edizione dello standard ISO 2709.

Negli anni '70 in molti paesi verranno sviluppati molti MARC nazionali da parte delle varie agenzie bibliografiche nazionali, ognuno seguente le regole di catalogazione del paese nel quale viene creato; sul finire del decennio, però, le accresciute esigenze di interscambio di dati e metadati bibliografici tra Paesi diversi determinano l'elaborazione della prima versione dell'UNIMARC (UNIVERSAL MACHINE READABLE CATALOGUING). Si può quindi notare come l'UNIMARC in origine non fu pensato come formato di descrizione bibliografica a se stante, ma solo come formato d'interscambio tra MARC di diversi Paesi. Dagli anni '80 in poi, l'UNIMARC comincia ad essere utilizzato, oltre che come formato

di interscambio, anche per la creazione originale di record bibliografici, e per seguire meglio questa evoluzione e promuovere l'utilizzo più ampio del formato (e quindi una sua possibile standardizzazione internazionale) negli anni '90 l'IFLA costituisce il PUC (Permanent Unimarc Committee), che pubblica periodicamente (l'ultimo è del 2000) gli aggiornamenti del manuale e alcune *guidelines*, ossia delle raccomandazioni per la catalogazione di oggetti speciali, come i libri antichi e le risorse elettroniche. Per quanto riguarda l'Italia, già dal 1985 la BNI (Biblioteca Nazionale Italiana) aveva adottato l'UNIMARC come formato di scambio delle registrazioni; in UNIMARC sono anche i dati del catalogo della Biblioteca nazionale centrale di Firenze⁴.

Finalità e scopi dell'UNIMARC

Il principale proposito che si pone l'UNIMARC è quello di facilitare lo scambio internazionale di dati bibliografici in forma digitale tra agenzie bibliografiche nazionali. L'UNIMARC può anche essere utilizzato come modello per lo sviluppo di nuovi formati bibliografici digitali.

Lo scopo dell'UNIMARC è di definire i designatori di contenuto (tag, indicatori e codici di sotto-campi), assegnandoli ai record bibliografici in forma digitale, e di specificare il formato logico e fisico dei record. Comprende monografie, serie, materiali cartografici, musica, registrazioni sonore, grafica, materiali video, libri rari e risorse elettroniche.

L'UNIMARC intende essere un formato di trasporto per ragioni di interscambio. Non definisce la forma, il contenuto o la struttura dei record all'interno di singoli sistemi: l'UNIMARC fornisce suggerimenti sulla forma e sul contenuto dei dati quando vengono scambiati

Mantenimento del formato

⁴ Vedi [Scolari 2000]

Il formato UNIMARC è mantenuto da una commissione dell'IFLA, la Commissione UNIMARC Permanente (PUC⁵). I cambiamenti futuri al formato definiranno principalmente campi, sotto-campi e valori di codici aggiuntivi, in caso di necessità, e verranno effettuati esclusivamente dalla PUC.

Notazioni convenzionali

Vengono riportate di seguito alcune convenzioni notazionali, che verranno utilizzate poi nella descrizione dei campi, degli indicatori e dei sotto-campi:

Identificatore di sotto-campo: il simbolo del dollaro \$ è utilizzato per rappresentare l'IS 1 dell'ISO 646, come primo carattere dell'identificatore di sotto-campo.

Spazio vuoto: il simbolo cancelletto # è stato usato per rappresentare il 2/0 dell'ISO 646, il carattere di spaziatura. Indica quindi un vuoto o uno spazio dove la presenza o la necessità di uno spazio non si può evincere dal contesto.

Separatore di campo: il simbolo "at" o "chiocciola"(@) è usato come carattere separatore di campo. Questo carattere inoltre termina ciascun campo del record UNIMARC.

Marcatore di fine record: il simbolo di percentuale % viene usato come marcatore di fine record.

Struttura del formato

L'UNIMARC è una speciale implementazione dell'ISO 2709⁶, uno standard internazionale che specifica la struttura dei record contenenti dati bibliografici. Specifica che ogni record bibliografico preparato per l'interscambio deve consistere di:

- un'ETICHETTA DEL RECORD, consistente di 24 caratteri;

⁵ Il Comitato si occupa anche di pubblicare le varie edizioni del manuale utenti: cfr. [UNIMARC 2001]

⁶ Cfr. [ISO 1996]

- una DIRECTORY, che consiste di un tag a 3 cifre per ogni campo dati, a seconda della lunghezza del campo e della posizione relativa al primo campo dati;
- CAMPI DATI di lunghezza variabile, ciascuno separato da un separatore di campo.

Con la seguente struttura:

ETICHETTA DEL RECORD	DIRECTORY	CAMPO DATI	R/T
----------------------	-----------	------------	-----

R/T: Terminatore del Record

L'ISO 2709 prevede inoltre che i dati contenuti nei campi possano essere eventualmente preceduti da indicatori e suddivisi in sotto-campi. L'UNIMARC, come implementazione utilizza proprio questa specifica opzione.

Etichetta del Record

L'ISO 2709 prescrive che ciascun record debba iniziare con un'etichetta da 24 caratteri. Questa contiene informazioni relative alla struttura del record e numerosi elementi di dati relativi a questa implementazione del formato ISO 2709. Questi elementi di dati definiti dall'implementazione riguardano il tipo di record, il suo livello bibliografico e la sua posizione in una gerarchia di livelli, il grado di completezza del record e l'uso o meno di regole ISBD o basate sull'ISBD nella preparazione del record stesso. Gli elementi di dati nell'etichetta del record sono richiesti principalmente per processare il record e solo indirettamente possono essere intesi come identificatori dell'oggetto bibliografico stesso.

Directory

Dopo l'etichetta del record troviamo la directory. Ciascuna voce della directory è formata da tre parti: un tag numerico a 3 cifre, un numero di 4 cifre che indica la lunghezza del campo dati ed un numero a 5 cifre che indica la posizione del carattere di partenza. Non sono permessi ulteriori caratteri all'interno di una voce della directory. La struttura della directory è come segue:

Voce 1 della directory			Voce 2	Altre voci	
Tag	Lunghezza campo dati	Posizione di partenza		F/T

F/T: Terminatore del campo

Il secondo segmento delle voci della directory riporta il numero di caratteri in quel campo. Include tutti i caratteri: indicatori, identificatori di sotto-campo, dati testuali o di codice e marcatore di fine campo. La lunghezza del campo è seguita dalla posizione del carattere di partenza del campo, intesa come la posizione di partenza del primo carattere della porzione variabile di campo all'interno del record. Il primo carattere del primo campo a lunghezza variabile rappresenta la posizione di carattere 0. La posizione della posizione di carattere 0 all'interno dell'intero record è data nelle posizioni da 12 a 16 dell'etichetta del record.

Come già detto, il tag ha una lunghezza di 3 caratteri, la "lunghezza del campo dati" occupa 4 caratteri e la "posizione del carattere di partenza" ne occupa 5. Dopo questi 12 caratteri corrispondenti a ciascun campo dati del record, la directory viene terminata dal marcatore di fine campo. Le voci della directory dovrebbero essere ordinate in base al primo carattere del tag, ed è raccomandabile ordinarle in base alla tag completa quando possibile.

Campi a lunghezza variabile

I campi a lunghezza variabile seguono la directory e generalmente contengono dati bibliografici o di elaborazione

Struttura del Campo Dati di controllo (00-)

Dati	F/T
------	-----

Struttura dei campi dati da 01 a 999:

Indicatori		Identificatore di sotto-campo		Altri sotto-campi	
Ind 1	Ind 2	\$a (etc.)	Data	Data F/T

I tag non sono contenuti nei campi dati, ma compaiono solo nella directory, ad eccezione dei tag dei campi incapsulati. I campi con tag di valore 00- (ad esempio 001) sono composti solo dei dati e dal terminatore di campo. Gli altri campi dati consistono di due indicatori seguiti da un numero qualsiasi di sotto-campi. Ciascun sotto-campo comincia con un identificatore, composto da un delimitatore e da un codice (un carattere numerico o alfanumerico) per identificare il sotto-campo. Gli indicatori sono seguiti da dati di codice o testuali di qualsiasi lunghezza tranne se diversamente stabilito nella descrizione del campo. L'ultimo sotto-campo viene terminato con il carattere di fine campo. L'ultimo carattere dei dati nel record è seguito comunemente dal carattere di fine campo seguito da un carattere di fine record.

Campi obbligatori

La seguente lista elenca i campi che devono essere presenti in un record UNIMARC:

- 001* Identificatore del record
- 100* Dati generali di elaborazione
- 101 Linguaggio del lavoro (dove applicabile)
- 120 Campo di dati codificati: materiale cartografico in generale (solo per oggetti cartografici)
- 123 Campo di dati codificati: scale e coordinate del materiale cartografico (solo per oggetti cartografici)
- 200* Titolo e dichiarazione di responsabilità (l'unico sotto-campo obbligatorio è \$a, titolo esatto)
- 206 Area specifica per i materiali: dati matematici per i materiali cartografici (solo per oggetti cartografici)
- 801* Campo fonte originale

I campi contrassegnati con l'asterisco (*) devono essere presenti in ogni record, senza eccezioni.

Lunghezza dei record

La lunghezza dei record, limitata dal formato a 99.999 caratteri è materia di accordi tra le parti in sede di interscambio.

Record Linking

All'atto pratico ci sono situazioni dove può essere desiderabile effettuare un collegamento tra un dato bibliografico ed un altro. Ad esempio, quando un record descrive una traduzione, può essere effettuato un collegamento al record che descrive il materiale originale. Nell'UNIMARC si prevede un metodo per effettuare questi link. Un blocco di campi (il blocco 4--) è riservato a questo proposito. Un campo collegato includerà informazioni descrittive riguardanti l'altro oggetto, con o senza informazioni puntanti ad un record separato che descriverà l'oggetto. Un campo collegato è composto di sotto-campi, ciascuno dei quali contiene un campo UNIMARC a sua volta formato da tag, identificatori e contenuti dei campi, inclusi marcatori di sotto-campo. E' da notare che questi campi incapsulati non sono accessibili tramite la Directory, dato che solo l'intero campo collegato possiede una voce directory. Il tag di questo campo descrive la relazione dell'oggetto identificato al suo interno con l'oggetto per il quale è stato creato il record.

Set di caratteri

Per l'interscambio di dati tramite UNIMARC, dovrebbero essere usati i set standard di caratteri ISO. Le etichette dei record, le directory, gli indicatori, gli identificatori di sotto-campo ed i valori di codice dovrebbero essere inseriti rispettando le funzioni di controllo ed i caratteri grafici dell'ISO 646, considerato il set standard per i record.

Informazioni ridondanti

Ci sono quattro possibili situazioni in cui i dati possono venire ripetuti:

- *I dati compaiono sia in forma di codice che testuale, visualizzati o non visualizzati*: dove possibile, i dati dovrebbero comparire in entrambe le forme, anche se l'informazione è conservata solo una volta nel formato sorgente

- *Il documento contiene le stesse informazioni in una lingua differente*: Lo Standard Internazionale di Descrizione Bibliografica (ISBD) specifica quando e come i dati paralleli debbano essere trascritti. Ciò si applica in UNIMARC attraverso l'uso di sotto-campi differenti o ripetuti.

- *Esiste più di un linguaggio di catalogazione per un'utenza multilingue*: L'uso di più di un linguaggio, per esempio, in un campo destinato alle annotazioni, è utile ed in alcuni casi obbligatorio all'interno di un formato domestico. Ai fini dell'interscambio internazionale questa pratica è meno accettabile: a meno che l'organizzazione ricevente utilizzi gli stessi linguaggi usati nel formato originale, avrà bisogno di eliminare tutti i linguaggi superflui tranne uno. Per questa ragione ogni record in un nastro UNIMARC di interscambio dovrà avere un solo linguaggio di catalogazione, con gli altri linguaggi inseriti in record o addirittura nastri separati.

- *La stessa informazione viene ripetuta con script differenti per provvedere a variazioni o miglioramenti*: Idealmente una voce di catalogo dovrebbe registrare un documento usando lo script del catalogo stesso, ma questo non è sempre possibile. Per questa ragione si dovrebbe poter registrare sia la versione originale che quella tradotta nella stessa voce di catalogo per permettere la selezione dell'opzione migliore da parte delle organizzazioni riceventi.

Trattamento di script differenti

Tutti i campi UNIMARC sono considerati ripetibili per la registrazione di grafiche o script alternativi, che essi siano inclusi nel corpo del testo o meno. I campi elencati come non ripetibili dovrebbero essere utilizzati una sola volta per ogni grafica/script alternativo incluso nel record.

Questa opzione ha lo scopo di fornire un supporto per la registrazione di traduzioni od ortografie e script alternativi eventualmente preparati dalle agenzie di catalogazione in accordo con le tabelle degli standard e le rispettive regole ed indicazioni.

In ciascun campo ripetuto allo scopo di registrare una rappresentazione grafica o uno script alternativo è necessario inserire sia il campo \$6 (Dati di collegamento tra campi) che il campo \$7 (Alfabeto/Script del campo). Si riportano di seguito le istruzioni per l'utilizzo di questi due campi.

\$6 – Dati di collegamento tra campi

Questo sotto-campo contiene informazioni che permettono il collegamento (linking) del campo, per scopi di elaborazione, ad altri campi all'interno del record. Deve inoltre contenere un codice che indica lo scopo del link. I primi due elementi del sotto-campo (posizioni di carattere da 0 a 2) devono essere sempre presenti quando il sotto-campo viene utilizzato; il terzo elemento (posizioni 3-5) è invece opzionale. La lunghezza di questo sotto-campo può quindi essere di 3 o 6 caratteri. Il sotto-campo \$6, se presente all'interno di un campo, dovrebbe essere il primo a comparire, a meno che non venga preceduto da un Authority Record Number \$3, e comunque precede ogni \$7. Non è ripetibile.

La struttura dei dati all'interno di un sotto-campo di tipo \$6 viene rappresentata come segue:

<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>Posizione di carattere</i>
Spiegazione del Link	1	0
Numero del Link	2	1-2
Tag dei campi linkati	3	3-5

\$6/0 – Codice di spiegazione del link (scopo del link)

Questo codice specifica la ragione per la quale viene effettuato il collegamento. Può assumere i seguenti valori: a = rappresentazione grafica/script alternativo

z = collegamento per altra ragione

\$6/1-2 – Numero del link

Questo numero a due cifre è contenuto nel sotto-campo \$6 di ciascuno dei campi collegati tra loro. La sua funzione è permettere l'abbinamento (matching) dei campi collegati, e non quelle di essere un numero progressivo o di identificazione per il link. Questo numero può essere assegnato casualmente in modo tale che i numeri assegnati a ciascun campo nella coppia o nel gruppo di campi collegati siano identici tra loro ma diversi da quelli assegnati a qualunque altra coppia o gruppo all'interno del record.

\$6/3-5 - Tag degli elementi linkati

Questo elemento consiste semplicemente del tag UNIMARC a tre cifre del campo a cui ci si collega. Questo campo è facoltativo: se i tag di entrambi gli elementi sono uguali, in genere li si omette.

\$7 Alfabeto/Script del campo

Questo sotto-campo contiene il codice per l'alfabeto e/o lo script per il contenuto principale del campo. I valori di codice sono quelli definiti per il campo 100 nelle posizioni di carattere 34 e 35 ("Script del Titolo"). Generalmente, il sotto-campo \$7 segue un sotto-campo \$6, a meno che non esistano campi paralleli, e quindi non ci sia un sotto-campo \$6. E' opzionale e non ripetibile.

Ordine dei sotto-campi

I sotto-campi \$0, \$3, \$6 e \$7 precedono tutti gli altri sotto-campi, mentre \$2, \$4 e \$5 seguono sempre tutti gli altri sotto-campi.

I sotto-campi numerici sono i seguenti:

- \$0 Numero di controllo
- \$1 Dati per il linking
- \$2 Codice di sistema
- \$3 Authority Record Number
- \$4 Codice del relatore

- \$5 Istituzione alla quale si applica il campo
- \$6 Dati di collegamento tra campi
- \$7 Alfabeto/Script del campo
- \$8 Non definito
- \$9 Riservato per l'uso locale

Caratteri utilizzati

All'interno dei campi dati UNIMARC è permesso qualsiasi carattere che corrisponda ad un carattere grafico visualizzabile testualmente. All'interno dei sotto-campi contenenti codici e negli indicatori di posizione sono invece validi i caratteri specificati per quel campo/sotto-campo.

Forma e contenuto dei dati

L'UNIMARC è in primo luogo un formato di interscambio, ed il suo primo scopo è quello di "trasportare" dati tra agenzie e sistemi che utilizzano formati di catalogazione diversi tra loro. Siccome questi dati potrebbero essere stati originariamente registrati secondo le disposizioni di codici di catalogazione differenti, non sempre è possibile prevedere esattamente la forma dei dati in un record UNIMARC, o elencare gli elementi di dati che dovrebbero sempre essere presenti. Tuttavia, a causa della loro precisa natura bibliografica, i record prodotti da diverse agenzie tendono comunque ad avere molto in comune, e ci sono vaste aree del formato all'interno delle quali si utilizzano pratiche similari. Una di queste aree è la porzione descrittiva del record, per la quale è stato sviluppato lo Standard Internazionale per la Descrizione Bibliografica (ISBD).

Etichetta del record e campi dati: informazioni generali

L'Etichetta del Record

Quest'area del record contiene informazioni di tipo generale che potrebbero venire richieste durante l'elaborazione del record, ed è costituita seguendo le regole dello standard ISO 2709. Esiste un'etichetta del record all'inizio di ogni record UNIMARC del database. E' obbligatoria e non ripetibile, e non possiede tag, indicatori o identificatori di sotto-campo.

Elementi di dati a lunghezza fissa

Questi elementi di dati sono identificati dalla posizione di carattere all'interno dell'etichetta, la cui dimensione totale è di 24 caratteri. Convenzionalmente, le posizioni hanno una numerazione da 0 a 23.

<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>Posizione di carattere</i>
Lunghezza del record	5	0-4
Stato del Record	1	5
Codici di implementazione	4	6-9
Lunghezza dell'indicatore	1	10
Lunghezza identificatore di sotto-campo	1	11
Indirizzo base dei dati	5	12-16
Definizione addizionale del record	3	17-19
Mappa della directory	4	20-23

0-4 - Lunghezza del record

Cinque cifre decimali che rappresentano il numero di caratteri dell'intero record, inclusa l'etichetta stessa, la directory ed i campi a lunghezza variabile. Questo elemento è solitamente calcolato in maniera automatica quando il record viene assemblato in fase di interscambio.

5 – Stato del record

Un singolo valore, che descrive lo stato di elaborazione del record. Può assumere i seguenti valori:

c = record corretto

Un record a cui sono stati fatti dei cambiamenti per correggere degli errori, che è stato sottoposto a revisione per essere aggiornato, o al quale sono stati cancellati alcuni campi. Tuttavia, se il record esistente era una pre-pubblicazione, ed è stato sostituito da una versione completa dello stesso, andrebbe usato il codice “p”, anziché “c”. Record di tipo “n”, “o” o “p” ai quali sono stati fatte delle correzioni contengono anch’essi un codice “c”.

d = record cancellato

Un record che viene scambiato per indicare che un qualsiasi record avente questo numero di controllo non è più valido. Il record potrebbe a questo punto contenere solo l’etichetta, la directory ed il campo 001 (Numero di controllo del record), o potrebbe anche contenere tutti gli altri campi del record. In entrambi i casi si dovrebbe utilizzare il campo 300 - NOTE GENERALI per indicare la ragione della cancellazione del record.

n = nuovo record

Un record nuovo, inteso anche come pre-pubblicazione. Se viene applicato un codice “o”, compare al posto di “n”.

o = esistenza di un record di livello più alto

indica un record posto in un livello gerarchico sottostante a un altro già esistente.

p = record precedentemente rilasciato come incompleto, o pre-pubblicazione record che sostituisce una pre-pubblicazione.

6-9 – codici di implementazione

I codici di implementazione sono così chiamati perché i codici nelle posizioni di carattere 6-9 non sono definiti nel formato ISO 2709, ma dipendono dalla singola implementazione dello standard. Una di queste implementazioni è appunto l’UNIMARC, e l’uso di questi codici è definito come segue:

6 Tipo di record

Richiesto da alcune agenzie. Il formato domestico assegna funzioni differenti agli stessi campi, sotto-campi, ecc. a seconda del tipo di record.

a = materiali linguistici stampati

b = materiali linguistici manoscritti

c = partiture musicali stampate

d = partiture musicali manoscritte

e = materiali cartografici stampati

f = materiali cartografici manoscritti

g = materiale video e proiettato (filmati, presentazioni, lucidi, registrazioni video)

i = registrazioni sonore, rappresentazioni non musicali

j = registrazioni sonore, rappresentazioni musicali

k = grafica bidimensionale

l = risorse elettroniche

m = multimedia

r = opere tridimensionali

7 – Livello bibliografico

Vengono definiti quattro valori possibili:

a = analitico (parte componente): oggetto bibliografico che è fisicamente contenuto in un altro, in maniera tale che l'ubicazione della parte componente sia dipendente dall'ubicazione dell'oggetto che la contiene. Una parte componente in se stessa può essere monografica o seriale. Alcuni esempi di oggetti bibliografici catalogati con livello "a" sono un articolo in un giornale o un singolo foglio in una serie di documenti di una conferenza.

m = monografico: oggetto bibliografico completo in una singola parte fisica, o che si intende essere completo in una serie finita di parti. Alcuni oggetti contrassegnati da "m"

sono le monografie, una parte separata di un catalogo composto da più parti oppure per citare un esempio più pertinente all'argomento musicale, una singola partitura tratta da un'opera finita.

s = seriale: oggetto bibliografico emesso in parti successive e inteso a continuare indefinitamente. Esempi di materiali codificati con "s" sono i giornali ancora in corso di pubblicazione, le riviste e le serie monografiche.

c = collezione: oggetto bibliografico facente parte di una collezione organizzata.

Esempi di materiali contrassegnati dal codice "c" sono le collezioni di libretti conservate in cofanetti, set di memorabilia di vario tipo organizzati e tenuti insieme da qualcuno come collezione, o tutti i manoscritti di un singolo autore. Questo codice si deve utilizzare solo per le collezioni organizzate.

8 – Codice di livello gerarchico

Questo codice indica le relazioni gerarchiche (se esistenti) tra record e altri record nel file. Vengono usati i seguenti codici:

= relazione gerarchica indefinita

0 = nessuna relazione gerarchica

1 = record di livello più alto

2 = record di livello sotto al primo (tutti i livelli sottostanti al primo)

9 – Non definito

Contiene uno spazio vuoto.

10 – Lunghezza dell'indicatore

Un valore numerico che restituisce la lunghezza degli indicatori. Nell'UNIMARC è invariabilmente 2.

11 – Lunghezza dell'identificatore di sotto-campo

Un valore numerico che indica la lunghezza dell'identificatore di sotto-campo (ad esempio "\$a"). Anch'esso in UNIMARC può avere come valore solamente 2.

12-16 – Indirizzo base dei dati

Cinque cifre numeriche, che indicano la posizione del carattere di partenza del primo campo dati in relazione all'inizio del record. Dato che il primo carattere del record ha numerazione 0 (zero), il numero inserito come indirizzo base dei dati sarà uguale al numero totale dei caratteri dell'etichetta e della directory, inclusi i separatori di campo che terminano la directory. Questo indirizzo è inoltre la base sulla quale vengono calcolate le posizioni di ciascun campo del record, e viene di solito calcolato automaticamente dal computer quando il record finale viene assemblato.

17-19 – Definizione addizionale del record

Tre posizioni di carattere contenenti codici che danno ulteriori dettagli per l'elaborazione del record:

17 – Livello di codifica

Codice a singolo carattere che indica in generale il grado di completezza della registrazione, e se l'oggetto è stato esaminato o meno al momento della creazione del record:

(spazio vuoto) = livello pieno. L'oggetto rappresentato nel record è stato esaminato al momento della preparazione per essere inserito in un database informatizzato.

1 = sottolivello 1. L'oggetto non è stato esaminato al momento della preparazione per l'inserimento nel database.

2 = sottolivello 2. Il record è una pre-pubblicazione. Questi record sono generalmente non completi, cioè il campo riservato alla collazione potrebbe essere assente o incompleto.

3 = sottolivello 3. Il record non contiene una catalogazione completa e potrebbe o meno essere successivamente aggiornato.

18 – Forma della catalogazione descrittiva

Codice di una cifra indicante la forma della catalogazione descrittiva utilizzata nel record. Indica se i campi di descrizione 200-225 sono stati costruiti secondo le regole ISBD. Può avere i seguenti valori:

(spazio vuoto) = il record è in forma ISBD: tutti i campi dati presenti nel record seguono le disposizioni dello standard ISBD.

i = il record è in forma ISBD solo parzialmente, o in maniera incompleta. Alcuni campi, ma non tutti, seguono le disposizioni dello standard ISBD.

n = il record non è in forma ISBD. Nessuno degli elementi di dati presenti nel record è necessariamente in accordo con le disposizioni dell'ISBD.

19 – Non definito

Contiene uno spazio vuoto

20-23 – Mappa della directory

Fornisce dettagli sulla lunghezza e la struttura delle voci della directory per ciascuno dei campi UNIMARC. Le quattro posizioni sono definite come segue:

20 – Lunghezza di “lunghezza del campo”

Cifra decimale che fornisce il numero di caratteri della parte “lunghezza del campo” di ciascuna voce della directory. Questo valore in UNIMARC (come descritto precedentemente) è 4. Questo permette una lunghezza massima per ogni campo di 9.999 caratteri.

21 – Lunghezza di “posizione del carattere di partenza”

Cifra decimale che fornisce il numero di caratteri della parte “posizione del carattere di partenza” di ciascuna voce della directory. Questo valore in UNIMARC è 5.

Questo permette una lunghezza massima del record di approssimativamente 100.000 caratteri.

22 – Lunghezza della porzione definita per implementazione

Cifra decimale che fornisce il numero di caratteri della porzione definita per l'implementazione per ciascuna voce della directory. Dato che le directory UNIMARC non contengono però tale porzione, il valore è 0.

23 – Non definito

Contiene uno spazio vuoto.

Breve descrizione dei campi obbligatori

Prima di cominciare a descrivere i campi di interesse musicale, viene fornita una breve descrizione dei campi obbligatori, ossia i campi che devono essere presenti in tutti i record UNIMARC per garantire la validità degli stessi.

001 – Identificatore del Record

Fa parte dei campi del blocco 0, che in generale contengono numeri che identificano il record o le informazioni registrate all'interno del record.

Il campo 001 contiene caratteri associati unicamente al record, ovvero il numero di identificazione del record assegnato dall'organizzazione che crea il database.

Conformemente a quanto stabilito dall'ISO 2709, non possiede ne' indicatori ne' sotto-campi, in quanto la sua funzione è solo quella di indicare il numero del record; non ci sono inoltre restrizioni alla forma che può assumere (numerica o alfanumerica), ma nel caso sia o contenga parte di una numerazione standard internazionale o un altro tipo di identificativo (per esempio un numero nazionale, o un altro tipo di dati, per esempio il paese di pubblicazione) tale numero va anche inserito nel campo previsto specificamente per tali tipi di informazioni.

Esempio:

001 B7512345

Numero identificativo di un record proveniente dalla *British National Bibliography*, corrispondente al Numero Bibliografico Nazionale (*National Bibliographic Number*)

B7512345. Inoltre, questo numero verrà inserito anche nel campo 020, Numero Bibliografico Nazionale.

100 – Dati Generali di Elaborazione

Fa parte dei campi del blocco 1, nel quale sono contenuti campi contenenti elementi di dati codificati a lunghezza fissa.

Il campo 100, nello specifico, contiene dati codificati a lunghezza fissa applicabili a record relativi a materiali provenienti da qualsiasi tipo di media (quindi anche quelli riguardanti il materiale audio). Come detto, è obbligatorio, e non è ripetibile. Possiede due indicatori (1 e 2), entrambi vuoti (non definiti), e un solo sotto-campo (\$a), che può contenere vari tipi di dati, a seconda della posizione di carattere. Questi dati si possono sommariamente suddividere in tre famiglie: la prima contiene informazioni riguardanti la data, la seconda dei codici che definiscono alcune caratteristiche dell'oggetto e la terza contiene informazioni sugli script ed i set di caratteri impiegati nel record.

\$a – Dati generali di Elaborazione

Tutti i dati inseriti nel sotto-campo \$a vengono identificati dalla posizione di carattere all'interno del sotto-campo stesso. Le posizioni sono convenzionalmente numerate da 0 a 35. Nella tabella sottostante vengono elencati i campi dati del sotto-campo \$a, per i quali viene fornita una breve descrizione.

	<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>Posizione di carattere</i>
Data	Data assegnata al file (obbligatorio)	8	0-7
	Tipo di data di pubblicazione	1	8
	Data di pubblicazione 1	4	9-12
	Data di pubblicazione 2	4	13-16
Codici	Codice di obiettivo di pubblico	3	17-19
	Codice di pubblicazione Governativo	1	20
	Codice del record modificato	1	21
	Linguaggio di catalogazione (obbligatorio)	3	22-24
	Codice di traslitterazione	1	25
Script	Set di caratteri (obbligatorio)	4	26-29
	Set addizionale di caratteri	4	30-33
	Script del titolo	2	34-35

\$a/0-7 – Data assegnata al file

Otto caratteri numerici in formato ISO (ISO 8601- 1988) per le date, secondo lo schema AAAAMMGG (A=anno, M=mese, G=giorno), precedute da zeri se necessario. La data inserita sarà generalmente quella in cui il record è stato creato e messo in formato-macchina.

\$a/8-16 – Tipo di data di pubblicazione, date di pubblicazione 1 e 2

Il tipo di data di pubblicazione è un codice, costituito da un singolo carattere, che indica il tipo di data in base a undici categorie che tengono conto dello “stato di pubblicazione” (*issuing status*) dell’oggetto registrato. Questo codice definisce i tipi di date inserite nei campi “date di pubblicazione” 1 e 2. Questi due campi prevedono l’utilizzo di quattro caratteri ciascuno, con la sola indicazione dell’anno. Riportiamo ora una breve descrizione dei tipi di data definiti per questi campi:

a = serie correntemente in pubblicazione

La data di pubblicazione (d.d.p.)1 contiene l’anno di inizio della pubblicazione o distribuzione, se essa differisce dalla pubblicazione. Se la data è incerta, ogni cifra ignota può essere sostituita da un carattere vuoto #.

La d.d.p. 2 contiene il numero 9999.

b = serie non più in pubblicazione

Per la d.d.p.1 vale quanto detto per il tipo “a”; la d.d.p.2 contiene l’anno in cui la pubblicazione è cessata. In caso di serie la cui fine è accertata ma è incerto l’anno, ogni cifra ignota può essere sostituito da un carattere vuoto.

c = serie in stato sconosciuto

Anche in questo caso, la d.d.p.1 contiene l’anno di pubblicazione/distribuzione, mentre la d.d.p.2 contiene quattro caratteri vuoti (####).

d = monografia completa al momento del rilascio, o rilasciata nel corso di un anno

Monografia pubblicata in un volume/parte o in un numero di parti/volumi recanti tutti la medesima data di pubblicazione. La d.d.p.1 contiene la data di pubblicazione. La d.d.p.2 contiene un carattere vuoto.

e = riproduzione di un documento

l'oggetto catalogato è una riproduzione (ovvero una ristampa, facsimile, ripubblicazione ecc.) e non una nuova edizione. La d.d.p.1 contiene la data di pubblicazione della riproduzione, la d.d.p.2 contiene la data di pubblicazione dell'originale.

f = monografia con data di pubblicazione incerta

La d.d.p.1 contiene la possibile data di pubblicazione più recente.

La d.d.p.2 contiene la possibile data di pubblicazione meno recente.

g = monografia la cui pubblicazione si protrae per più di un anno

la data di pubblicazione 1 contiene l'anno d'inizio della pubblicazione. In caso di incertezza sulla data, ogni valore incerto può venire sostituito da un carattere vuoto.

La data di pubblicazione 2 contiene invece l'anno in cui finisce la pubblicazione, o il valore 9999 se la pubblicazione è ancora in corso.

h = monografia con data effettiva e data del copyright

Utilizzato quando la data di pubblicazione differisce dalla data di inizio del copyright citata sull'oggetto da catalogare.

La d.d.p.1 contiene la data di pubblicazione, la d.d.p.2 contiene invece la data di partenza del copyright.

i = monografia con data di rilascio e data di produzione

Usato per film e opere musicali, nelle quali la data di produzione/registrazione dell'opera spesso differisce dalla data di distribuzione/rilascio.

Data di produzione 1: data di rilascio;

data di produzione 2: data di produzione.

j = documento con data di pubblicazione dettagliata

usato quando viene ritenuto importante registrare mese (e possibilmente giorno) di pubblicazione dell'oggetto.

La d.d.p1 contiene l'anno di pubblicazione, la d.d.p2 contiene la data dettagliata nella forma "MMGG".

u = data/e di pubblicazione sconosciute

usato quando non è oggettivamente possibile assegnare una data all'oggetto. Entrambe le date di pubblicazione sono vuote.

\$a/17-19 - Codice di obiettivo di pubblico

Tre posizioni di carattere sono dedicate all'obiettivo (target) di pubblico, e vanno inserite da sinistra a destra, utilizzando il carattere vuoto per riempire ogni spazio mancante.

a = giovani in generale (da usare quando non è possibile utilizzare i codici b, c o d)

b = fascia pre-primaria , 0-5 anni

c = fascia primaria, 5-10 anni

d = bambini, 9-14 anni

e = adolescenti, 14-20 anni

k = adulti, contenuto serio

m = adulti in generale

u = sconosciuto

\$a/20 – Codice Governativo di pubblicazione

Codice sostituito da un singolo carattere che indica se si tratta di una pubblicazione governativa o meno, ed il livello dell'istituzione che rilascia la pubblicazione. Per pubblicazione governativa si intende una pubblicazione edita o finanziata dal Governo.

Le istituzioni accademiche non sono considerate governative.

a = federale/nazionale

b = statale/provinciale

c = contea/dipartimento/regione

d = locale (municipale, ecc.)

e = multi-locale (interessa più zone)

f = intergovernativo (interessa più nazioni, ex. ONU)

g = governo in esilio o clandestino (ex. Governo francese nel periodo 1941-1944)

h = livello non determinato

u = sconosciuto

y = pubblicazione non governativa (in tutti i casi in cui non si ha una pubblicazione governativa)

z = altro livello governativo

\$a/21 – Codice del record modificato

Questo codice a singolo carattere indica se il set di caratteri disponibile è sufficiente a trascrivere i dati così come si trovano sull'oggetto da catalogare. Infatti in alcuni casi dovuti alla limitazione dei set di caratteri presenti sulle macchine, nella trascrizione delle informazioni possono venire usate delle semplificazioni per rappresentare i dati tramite i caratteri disponibili (ex. caratteri matematici). E' solo di due tipi:

0 = record non modificato

1 = record modificato

\$a/22-24 - Linguaggio di catalogazione (obbligatorio)

Codice di tre caratteri che indica il linguaggio utilizzato per la catalogazione (ex. italiano = ita).

\$a/25 - Codice di traslitterazione

L'utilizzo del formato UNIMARC dà in genere per scontato l'utilizzo di standard internazionali. Tuttavia, nel settore della conversione degli script, gli standard internazionali non sono completamente adeguati, e possono esserne usati degli altri.

Questo codice indica se per la traslitterazione è stato utilizzato uno schema ISO o meno.

a =schema di traslitterazione ISO

b = altro

c = traslitterazione multipla tramite ISO o altri schemi

y = nessuno schema di traslitterazione

\$a/26-29 - Set di caratteri (obbligatorio)

Due coppie di caratteri vengono riservate alla designazione dei principali set di caratteri grafici usati nel record. Si utilizzano le seguenti coppie di caratteri:

01 = ISO 646, IRV version (set Latino base)

02 = ISO Registration # 37 (set Cirillico base)

03 = ISO 5426 (set Latino esteso)

04 = ISO DIS 5427 (set Cirillico esteso)

05 = ISO 5428 (set Greco)

06 = ISO 6438 (set Africano)

07 = ISO 10586 (set Georgiano)

08 = ISO 8957 (set Ebreo) Tabella 1

09 = ISO 8957 (set Ebreo) Tabella 2

10 = [Reserved]

11 = ISO 5426-2 (caratteri latini usati in alcune lingue minori europee e in tipografie obsolete)

50 = ISO 10646 Level 3 (Unicode)

\$a/30-33 – Set addizionale di caratteri

Ulteriori due coppie di caratteri per indicare fino a due set di caratteri addizionali utilizzati nel record.

\$a/34-35 – Script del titolo

Coppia di caratteri indicante l'alfabeto utilizzato nel titolo:

ba = Latino

ca = Cirillico
 da = Giapponese - script non specificato
 db = Giapponese – kanji
 dc = Giapponese – kana
 ea = Cinese
 fa = Arabo
 ga = Greco
 ha = Ebreo
 ia = Thailandese
 ja = Devanagari
 ka = Coreano
 la = Tamil
 ma = Georgiano
 mb = Armeno
 zz = altro

Ecco un esempio completo di campo di tipo 100:

100##\$a19601104a19599999m##c0engy0103#####ba

All’inizio troviamo il codice del campo (100), l’indicazione che gli indicatori del campo sono entrambi vuoti (##) e l’identificazione del sotto-campo (\$a); seguono le 35 posizioni di carattere dei campi dati, che riportiamo qui in versione schematica:

<i>Posizione di carattere</i>	<i>Valore</i>	<i>Note</i>
0-7	19601104	Data riportata sul file
8	a	Serie correntemente in pubblicazione
9-12	1959	Anno di inizio della pubblicazione
13-16	9999	La serie è in fase di continuazione
17-19	m##	Obiettivo di pubblico: adulti in generale
20	c	Documento di tipo regionale
21	0	Record non modificato
22-24	eng	Linguaggio di catalogazione: inglese
25	y	Nessuno schema di traslitterazione utilizzato

26-29	0103	Si utilizzano codici ISO 646 (IRV) e 5426
30-33	####	Nessun set addizionale di caratteri utilizzato
34-35	ba	Alfabeto del titolo: Latino

Il record è stato steso in formato-macchina il 4 Novembre 1960. La serie iniziò ad essere pubblicata nel 1959 ed è ancora in pubblicazione. L'obiettivo di pubblico è "adulti in generale". Il documento è una relazione annuale pubblicata dal Consiglio della Contea dell'Hampshire (quindi un consiglio di tipo regionale); il record non è stato modificato ed è catalogato in lingua inglese senza schemi di traslitterazione.

200 – Titolo e dichiarazione di responsabilità

Questo campo contiene il titolo del record insieme a qualsiasi tipo di informazione sul titolo e/o dichiarazione di responsabilità relativa al titolo stesso. Fa parte dei campi inclusi nel blocco 2, che coprono tutte le informazioni contenute nell'ISBD, ad eccezione delle note e dei numeri standard. Il campo 200 è obbligatorio in ogni record e non è ripetibile. Possiede due indicatori: l'Indicatore 1 (Indicatore della significatività del titolo), che può assumere due valori (0 = titolo non significativo, 1 = titolo significativo), e l'Indicatore 2, che è vuoto.

La tabella qui sotto illustra i sotto-campi disponibili per questo campo, con il nome e la sezione ISBD che ricoprono.

<i>Sotto-campi UNIMARC</i>	<i>Nome del sotto-campo</i>
\$a	Titolo vero e proprio
\$a (ripetuto)	Titoli dello stesso autore
\$b	Definizione generale del materiale
\$c	Titoli di altri autori (utilizzato per le collaborazioni)
\$d	Titolo parallelo (in altre lingue)
\$e	Altre informazioni sul titolo
\$f	Prima dichiarazione di responsabilità
\$g	Successive dichiarazioni di responsabilità
\$h	Numero di parte
\$i	Nome della parte

\$v	Designazione di volume
\$z	Lingua del titolo parallelo
\$5	Istituzione alla quale si applica il campo

\$a – Titolo vero e proprio

Titolo principale dell'opera, include titoli alternativi ma esclude altre informazioni sul titolo (ad esempio i sottotitoli), ed i titoli paralleli (in altre lingue). Obbligatorio in tutti i record UNIMARC.

\$b – Definizione generale del materiale

Contiene un termine che indica la classe di materiale al quale appartiene l'oggetto, riportato nel linguaggio o nello script dell'agenzia che ha preparato il record. Corrisponde all'omonimo campo ISBD. Ripetibile.

\$c – Titoli di altri autori

Titolo vero e proprio di un lavoro di un autore che non è quello responsabile del titolo vero e proprio dell'oggetto catalogato, quando lavori di entrambi gli autori sono presenti nello stesso oggetto e non c'è un titolo collettivo.

\$d – Titolo parallelo

Titolo vero e proprio in un altro linguaggio-script relativo ai titoli che compaiono nei sotto-campi \$a e \$c. Ripetibile per ciascun titolo parallelo.

\$e – Altre informazioni sul titolo

Sottotitoli ed altre informazioni che appaiono subordinate al titolo vero e proprio dell'oggetto.

\$f – Prima dichiarazione di responsabilità

Prima dichiarazione di responsabilità per un titolo che compaia in un sotto-campo \$a, \$c o \$d, o di una parte di opera che compaia nei sotto-campi \$h o \$i.

\$g – Successive dichiarazioni di responsabilità

Dichiarazioni di responsabilità che seguono il primo e sono relativi allo stesso titolo. Ripetibile per ciascuna dichiarazione di responsabilità seguente la prima, anche se riguardante dati paralleli.

\$h – Numero di parte

Numerazione di una sezione o una parte di oggetti identificati da un nome comune e da un nome di sotto-sezione. Ripetibile per ciascuna sotto-sezione o numero di parte parallela.

\$i – Nome della parte

Nome di sezione di una parte di oggetti identificati da un nome comune e da un nome di sotto-sezione. Ripetibile per ciascuna parte, sotto-sezione o parte parallela.

\$v – Designazione di volume

Usato per indicare una parte dell'oggetto che è in relazione ad un altro oggetto. Questo sotto-campo è usato solo quando un campo di tipo 200 viene incapsulato in un campo di tipo 4-- (linking).

\$z – Linguaggio del titolo parallelo

Identificazione del linguaggio parallelo che compare in un sotto-campo \$d. Si consiglia la codifica internazionale USMARC, anche se non esistono ancora codici internazionalmente accettati come standard.

Alcuni esempi di campi di tipo 200:

Es. 1

200 1#\$aThree adventures of Asterix\$iAsterix in Switzerland\$text by Goscinny\$gdrawings by Uderzo\$gtranslated by Anthea Bell and Derek Hockridge.

L'oggetto è uno di tre volumi, con un titolo comune per il set di tre e un titolo individuale assegnato a ciascun volume. La prima dichiarazione di responsabilità è inserita in \$f, e tutte le altre in \$g. Da notare la ripetibilità del sotto-campo \$g.

Es. 2

200 1#\$aFlash and filigree\$and, The Magic Christian\$by Terry Southern

L'oggetto catalogato contiene due lavori singoli dello stesso autore, ciascuno con il proprio titolo, senza un titolo collettivo.

801 – Fonte originale

Fa parte dei campi del blocco 8, contenenti campi accettati in base ad accordi internazionali che non trovano posto nei precedenti blocchi.

Il campo 801 contiene l'indicazione dell'origine del record, inclusa una di queste opzioni: l'agenzia che ha creato i dati, l'agenzia che ha trascritto i dati in formato-macchina, qualsiasi agenzia che abbia modificato i dati originali e l'agenzia che rilascia il presente record.

E' l'ultimo dei campi obbligatori, ed è ripetibile per ogni funzione precedentemente descritta.. Ha due indicatori, 1 e 2: il primo è vuoto, il secondo contiene l'indicatore della funzione:

0 = agenzia originale di catalogazione (agenzia che ha preparato i dati nel record);

1 = agenzia trascrittrice (agenzia che ha messo i dati in formato-macchina);

2 = agenzia modificatrice (agenzia che ha modificato il contenuto intellettuale e/o la struttura del record);

3 = agenzia editrice = agenzia che ha rilasciato il record.

Sotto-campi:

<i>Sotto-campo</i>	<i>Nome del sotto-campo</i>	<i>Descrizione</i>
\$a	Paese	Nazione dell'agenzia che ha rilasciato il record
\$b	Agenzia	Codice USMARC/nome completo/numero nazionale dell'agenzia
\$c	Data di trascrizione	Usato quando si vuole

		indicare la data di creazione o modifica di un record
\$g	Regole di catalogazione	Contiene un'abbreviazione del codice di catalogazione usato per la descrizione bibliografica.
\$2	Codice di sistema	Nome del formato usato per il formato-macchina

Esempio di campi di tipo 801:

801 #0\$aUS\$bDLC\$c19590000\$gAACR1

801 #1\$aUS\$bMH\$c19790506

801 #2\$aUS\$bMH\$c19790506\$gAACR2

801 #3\$aUS\$bDLC\$c19790912

L'oggetto è stato inizialmente catalogato dalla Library of Congress nel 1959. Nel 1979, la Harvard University ha modificato i dati li ha trascritti in formato-macchina. Questo record è stato quindi rilasciato dalla Library of Congress. Per identificare queste due istituzioni sono stati usati i codici provenienti dalla Lista USMARC di codici per le Organizzazioni.

Campi di interesse musicale

All'interno dell'UNIMARC ci sono più di 160 campi, che coprono una grande varietà di aspetti della catalogazione bibliografica. Essendo stato ideato per la catalogazione di libri, esso si adatta efficacemente anche alle partiture musicali, ed al materiale musicale cartaceo in generale. Il discorso si fa più complesso se si prende in considerazione anche l'aspetto audio del materiale musicale (ossia la registrazione sonora vera e propria). In un database che contenga, oltre alle informazioni bibliografiche ed alle partiture, anche delle registrazioni audio degli oggetti descritti, si possono utilizzare alcuni di questi campi, in alcuni casi adattandoli all'uso specifico della catalogazione di materiale audio.

Viene fornita ora una descrizione dei campi di maggiore utilità in questo senso.

013 – Numero Musicale Internazionale Standard (ISMN)

Questo campo contiene un Numero Musicale Internazionale Standard⁷ (*International Standard Musical Number*), e una qualifica che distingue gli ISMN quando all'interno di un record ve ne è contenuto più d'uno. Questo campo corrisponde al Numero Standard ISBD, e può contenere degli indicatori di disponibilità/prezzo, anche se non è presente un numero ISMN. E' opzionale, e ripetibile quando devono essere registrati più numeri ISMN. Possiede due indicatori (1 e 2), entrambi contenenti un carattere vuoto (e quindi non definiti), e quattro sotto-campi:

\$a – Numero (ISMN)

Un numero ISMN completo, inclusi i trattini. I numeri ISMN vengono assegnati dalle apposite agenzie in ciascun paese. Non è ripetibile.

\$b – Qualifica

Un'indicazione dello scopo dell'ISMN nel campo \$a (quando presente). Di solito si tratta del nome dell'editore, un'indicazione della rilegatura dell'oggetto o della relazione dell'oggetto con un particolare volume. Non ripetibile.

\$d – Indicazioni di disponibilità e/o prezzo

Campo che contiene qualsiasi tipo di informazioni e commenti sulla disponibilità dell'oggetto. Non ripetibile.

\$z - ISMN errato

Un ISMN che è stato riconosciuto come applicato erroneamente ad un oggetto ed è quindi invalidato. Ripetibile.

Esempio:

013 ##\$aM-9005202-2-7\$b(bound)

⁷ Il numero ISMN viene utilizzato per identificare le pubblicazioni musicali disponibili per la vendita, il noleggio, a titolo gratuito, o a scopi esclusivamente di copyright, secondo le disposizioni dello standard ISO 10957. Ciascuna parte della pubblicazione disponibile separatamente deve avere il proprio numero ISMN, che consiste nella lettera M seguita da nove cifre. Ogni pubblicazione musicale ha il proprio numero ISMN

013 ##\$aM-9005202-3-4\$b(pbk)

Questa pubblicazione è disponibile sia in edizione rilegata che in versione tascabile.

021 – Numero di deposito legale

Questo campo contiene il numero assegnato all'oggetto catalogato dal deposito legale o dall'ufficio copyright in ciascun paese. E' opzionale e ripetibile. Anche il campo 021 ha due indicatori vuoti, e tre sotto-campi:

\$a - Codice del Paese

Identificativo della nazione dell'agenzia di deposito legale che assegna il numero. Non ripetibile.

\$b – Numero

Numero così come viene assegnato dall'agenzia. Non ripetibile.

\$z – Numero errato

Numero assegnato erroneamente all'oggetto. Ripetibile.

Esempi:

022 ##\$aUS\$bE17.302.W58/91

Pubblicazione del Governo degli Stati Uniti.

022 ##\$aGB\$bECC.56/81

Pubblicazione del Consiglio della contea dell'Essex, Regno Unito.

022 ##\$bPGI/81/WS/22

Pubblicazione dell'UNESCO.

071 – Numero dell'editore (Musica e registrazioni sonore)

Questo campo contiene il numero identificativo dell'editore che viene usato per le registrazioni sonore e musicali. E' opzionale e ripetibile.

Ha due indicatori: l'Indicatore 1 (tipo di numero dell'editore) specifica il tipo di numero utilizzato:

0 = numero di rilascio

Chiamato anche numero di etichetta, viene utilizzato per l'identificazione la registrazione sonora da parte dell'editore. Questo numero di solito compare sull'etichetta e spesso anche sulla confezione delle registrazioni

1 = numero di matrice

Usato per identificare il master dal quale la singola registrazione è stata stampata. Viene usata per l'identificazione della registrazione sonora in assenza del numero di etichetta.

2 = numero di esemplare

Il numero di esemplare è un numero di serie assegnato dall'editore ad ogni singola pubblicazione musicale. E' di solito stampato in fondo a ogni pagina contenente musica, e a volte appare sulla prima pagina.

3 = altri numeri dell'editore

Simili al numero di esemplare, ma non compaiono su ogni pagina. Possono comparire in prima pagina, in copertina e/o sulla prima pagina di musica.

L'Indicatore 2 è un indicatore di note, e specifica se l'agenzia che fornisce il record ha intenzione di utilizzare i dati per inserirli in una nota. Se non sono richieste note (ad esempio perché già inserite nei campi 300 o 301) questo numero viene settato a 0.

0 = non creazione di nota

1 = creazione di nota

Il campo 071 contiene due sotto-campi:

\$a = Numero dell'editore

\$b = Fonte (l'agenzia che ha fornito il numero. Non ripetibile)

Esempi:

071 01\$aSTMA 8007\$bTamla Motown

Classificazione rilasciata per una registrazione sonora pubblicata dalla casa editrice Tamla Motown. Verrà generata una nota, ad esempio Tamla Motown: STMA 8007.

071 21\$aA 880 V\$bArs Viva Verlag

Numero di esemplare di musica stampata. Verrà generata una nota, ad esempio Esempl. No. A880 V.

125 – Campi dati codificati: registrazioni sonore e musica stampata

Questo campo descrive il formato della musica stampata, annota le parti esistenti e codifica il tipo di testo letterario in caso di rappresentazioni non musicali. E' opzionale e non ripetibile, e ha due indicatori vuoti.

Contiene due sotto-campi:

\$a = formato della musica stampata

Costituito da due caratteri e non ripetibile.

Elementi di dati del sotto-campo \$a:

<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>Posizione di carattere</i>
Tipo di partitura	1	0
Indicatore delle parti	1	1

L'elemento di dati \$a/0 (Tipo di partitura) indica il tipo di partitura musicale dell'oggetto catalogato nel record.

a = partitura completa

b = partitura in formato miniatura o formato da studio
c = partitura vocale, accompagnamento ridotto per pianoforte
d = partitura vocale o corale, accompagnamento “dropped”
e = partitura condensata, riduzioni per canto e pianoforte
g = partitura a parti strette (inni)
m = formati multipli
u = sconosciuto
x = nessun tipo applicabile/nessuna partitura
z = altro

\$a/1 Indicatore delle parti

Indicatore a singolo carattere che segnali la presenza di parti per una partitura

a = presenza di parti
u = sconosciuto
x = non applicabile
y = parti non presenti

\$b = Indicatore di testo letterario

Codice alfabetico che indica il tipo di testo letterario in una rappresentazione non musicale. Possono essere usati fino a due singoli codici, con caratteri vuoti per i codici non usati. Non ripetibile.

a = poesia
b = drammatico
c = racconti (favole, storie brevi, ecc...)
d = storia
e = letture, discorsi
f = istruzioni (come fare...)
g = suoni
h = autobiografie
i = biografie

j = saggi
 k = resoconti
 l = memorie
 m = prove
 n = interviste
 o = testi pubblicitari
 p = istruzioni (lingue)
 q = avvenimenti di conferenze
 r = comicità
 s = storie tradizionali
 t = testi sacri
 z = altri tipi di testi letterari

Esempi:

125##\$aaa

Partitura completa di un'opera per la quale sono disponibili le parti.

125##\$ada

Partitura corale per la quale sono disponibili le parti

126 – Campi dati codificati: registrazioni sonore – attributi fisici

Questo campo contiene dati relativi alle caratteristiche fisiche delle registrazioni sonore (ossia dei supporti che le contengono). Non è ripetibile, e contiene due indicatori vuoti. Contiene due sotto-campi (\$a, ripetibile, e \$b, non ripetibile) con diversi campi dati:

\$a = Dati generali della registrazione sonora

Elementi di dati del sotto-campo \$a a lunghezza fissa:

<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>di</i>	<i>Posizione di carattere</i>	<i>di</i>

Forma di edizione	1	0
Velocità	1	1
Tipo di suono	1	2
Larghezza del solco	1	3
Dimensioni del supporto	1	4
Larghezza del nastro	1	5
Configurazione del nastro	1	6
Materiale cartaceo di accompagnamento	6	7-12
Tecnica di registrazione	1	13
Caratteristiche speciali di riproduzione	1	14

Vediamo uno per uno questi campi dati:

\$a/0 - Forma di edizione

Indica la forma di edizione della registrazione sonora

a = disco

b = nastro (bobina)

c = nastro (cassetta)

d = nastro (cartuccia)

e = registrazioni su filo

f = cilindro

g = rulli (per pianoforte o organo automatici)

h = pellicola (pellicola sonora)

z = altro

\$a/1 – velocità

Dischi:

a = 16 2/3 rpm

b = 33 1/3 rpm

c = 45 rpm

d = 78 rpm

e = 8 rpm

Nastri:

k = 1 7/8 in. per second

l = 15/16 in. per second

m = 3 3/4 in. per second

n = 7 1/2 in. per second

o = 15 in. per second

g = 1.4 m. per second (compact discs)

p = 30 in. per second

q = 8/10 in. per second

r = 4/10 in. per second

Cilindri:

h = 1 in. per second (120 rpm)

i = 160 rpm

u = sconosciuto

x = non applicabile

z = altro

$\$a/2$ – Tipo di suono

Indica il metodo di registrazione dell'oggetto

a = monoaurale

b = stereofonico

c = quadrafonico

u = sconosciuto

z = altro

$\$a/3$ Larghezza del solco

Indica la larghezza del solco sul disco

a = grande/standard

b = microsolco/fine

u = sconosciuto

x = non applicabile

z = altro

Grande/ standard: i dischi a 78 giri sono di solito a solco grande. I cilindri rilasciati a 120 giri/minuto (che hanno 100 solchi per pollice) sono di solito standard.

Microsolco/fine: i dischi a 16 2/3, 33 1/3 e 45 giri sono di solito definiti microsolchi. I cilindri a 160 giri/minuto (200 solchi per pollice) sono definiti solchi di tipo fine.

Nel campo “non applicabile” sono inclusi i compact disc, che non possiedono solchi.

\$a/4 – Dimensioni del supporto

Codice a singolo carattere che specifica il diametro di una bobina o di un disco, le dimensioni di una cassetta, cartuccia o cilindro.

a = 3 in.

b = 5 in.

c = 7 in.

d = 10 in.

e = 12 in.

f = 16 in.

g = 14 in.

h = 4³/₄ in. (compact disc)

j = 3 7/8 x 2¹/₂ in. cassetta

o = 5¹/₄ x 3 7/8 in. cartuccia

s = 2³/₄ x 4 in. cilindro

u = sconosciuto

x = non applicabile

z = altro

\$a/5 – Larghezza del nastro

Indica la larghezza del nastro

a = 1/4 in.

b = 1/2 in.

c = 1 in.

d = 1/8 in.

e = 2 in.
f = 1/3 in. (8 mm.)
u = sconosciuto
x = registrazione non su nastro
z = altro

\$a/6 – Configurazione del nastro

Indica la configurazione del nastro

a = full (1) track
b = half (2) track
c = quarter (4) track
d = eight track
e = twelve track
f = sixteen track
g = twenty-four track
h = six track
u = sconosciuto
x = registrazione non su nastro
z = altro

\$a/7-12 – Materiale cartaceo di accompagnamento

Codice alfabetico che indica il contenuto del programma e altro testo d'accompagnamento. Possono venire usati fino a sei di questi codici: se più materiali sono presenti, bisogna effettuare una stima di importanza ed inserire i primi sei.

a = discografia
b = bibliografia
c = indice tematico
d = testo o libretto
e = biografia del compositore
f = biografia dell'esecutore o storia dell'ensemble

g = informazioni tecniche o storiche sugli strumenti

h = informazioni tecniche sulla musica

i = informazioni storiche sulla musica

j = altre notizie storiche

k = informazioni etnologiche

l = biografia dell'arrangiatore o trascrittore

r = materiale contenente istruzioni

s = partitura

z = altro materiale cartaceo di accompagnamento

\$a/13 – Tecnica di registrazione

Indica la tecnica di registrazione utilizzata

a = acustica

b = elettrica

c = digitale

u = sconosciuta

z = altro

\$a/14 – Caratteristiche speciali di riproduzione

Indica le caratteristiche di riproduzione

a = NAB standard

b = CCIR/IEC standard

c = processato DBX

d = digitale (compact disc)

e = codificato Dolby-A

f = codificato Dolby-B

g = codificato Dolby-C

h = codificato CX

x = non applicabile

u = sconosciuto

z = altro

Vediamo ora i campi di dati del sotto-campo \$b (Dati dettagliati della registrazione sonora):

<i>Nome dell'elemento di dati</i>	<i>Numero di caratteri</i>	<i>Posizione di carattere</i>
Tipo di disco, cilindro o nastro	1	0
Materiale	1	1
Tipo di taglio	1	2

\$b/0 – Tipo di disco, cilindro o nastro

Indica il tipo di disco, cilindro o nastro utilizzato

a = istantaneo (unico, “*cut on the spot*”, fatto su misura)

b = prodotto di massa (la maggior parte di dischi o nastri commerciali)

c = nastro master

d = nastro master di duplicazione

e = disco master (negativo)

f = disco madre (positivo)

g = disco stampato (negativo)

h = stampa di prova

u = sconosciuto

x = non applicabile

z = altro

\$b/1 – Materiale

Specifica il materiale per i dischi, cilindri e nastri

u = sconosciuto

x = non applicabile

z = altro

Dischi:

a = materiali laccati (ex. acetato)

b = metallo (ex. alluminio)

c = stampaggio shellac (prodotto di massa)

d = stampaggio plastico (prodotto di massa)

e = metallo e plastica (compact disc)

Cilindri:

g = cera (istantaneo)

h = plastica (prodotto di massa)

nastri:

i = a base cartacea

j = acetato

k = PVC

l = poliestere

\$b/2 – Tipo di taglio

Indica il tipo di taglio. Dato che tutti i cilindri hanno un taglio di tipo “hill and dale”, l’uso di questo codice per i cilindri è opzionale.

a = taglio laterale o combinato

b = taglio verticale (hill and dale)

u = sconosciuto

x = non applicabile

Esempio:

126 ##\$aagbxhxxdes###cd\$bbex

Compact disc contenente musica registrata in digitale e accompagnato dal libretto dell’opera, dalla biografia del compositore e dalla partitura dell’opera.

127 – Campi dati codificati: durata delle registrazioni sonore e della musica stampata

Questo campo contiene uno o più numeri di sei caratteri corrispondenti alla durata di una registrazione sonora, di una porzione di registrazione sonora o della durata stimata di una composizione stando a quanto riportato sulla partitura stampata della composizione.

Non è ripetibile, ha due indicatori vuoti e un solo sotto-campo, \$a, che contiene la durata della registrazione sonora. E' formato da sei cifre divise in tre elementi: due cifre per le ore, due per i minuti e due per i secondi. E' un sotto-campo ripetibile.

Esempi:

127 ##\$a003100\$a001839

Durata di una registrazione sonora con due composizioni, una della durata di 31 minuti e l'altra della durata di 18 minuti e 39 secondi.

127 ##\$a001356\$a002005

300 ##\$aDurations: 13:56; ca. 20:05

Partitura musicale a stampa con due pezzi, il primo lungo 13 minuti e 56 secondi, il secondo lungo approssimativamente 20 minuti e 5 secondi. E' stata inoltre inserita una nota nel campo 300 (Note Generali).

128 – Campi dati codificati: esecuzioni e partiture musicali

Questo campo descrive la forma della composizione e fornisce dettagli sugli strumenti e sulle voci che formano la composizione stessa.

E' ripetibile e contiene due indicatori vuoti.

Contiene tre sotto-campi:

\$a = forma della composizione

Formato da un codice di due caratteri che indica la forma della composizione. Se è presente più di una forma di composizione il sotto-campo viene ripetuto.

ai = arie (airs)

an = inni

ar = arie (arias)

bt = balletti

bd = ballate (non ballate tradizionali)

bg = bluegrass

bl = blues

cb = canti non sacri

c = canti sacri

cl = preludi corali

cg = concerti grossi

ch = corali

cn = canoni

Composizioni che utilizzano procedimenti di imitazione severi. Comprende i canoni.

co = concerti

cp = chanson polifoniche

cr = canti di Natale

cs = *chance compositions*

ct = cantate

cy = musica country

cz = canzona

Musica strumentale designata come tale

df = forme di danza

Musica per danze individuali ad eccezione di mazurke, minuetti, pavane, polonaise e walzer.

dv = divertimenti

Musica strumentale designata come tale; include serenate, cassazioni, e divertimenti del 18mo secolo

fg = fughe

Musica strumentale del 17mo secolo designata come tale; composizioni contrappuntistiche basate su imitazione del soggetto dal 17mo secolo in avanti.

fm = musica folk

Include la musica folk in generale(folk canzoni, ballate, ecc.)

ft = fantasie

Musica strumentale designata come tale;include fantasie e ricercari non fugati del 16mo e 17mo secolo.

gm = musica gospel

hy = inni religiosi (hymns)

in = intermezzi

jz = musica jazz

li = lieder

mc = riviste e commedie musicali

md = madrigali

mi = minuetti

mr = marce

ms = messe

mo = mottetti

mp = musica da film

mu = forme multiple

mz = mazurke

nc = notturni

op = opera e operette

or = oratori

ov = ouverture

pf = preludi e fughe

pg = musica a programma

pm = passioni

po = polonaise

pp = musica popolare (ad eccezione dei casi in cui viene assegnato un codice specifico)

ps = passacaglie

Musica strumentale designata come tale, include tutti i tipi di basso ostinato.

pt = canzoni-parte

pv = pavane

rc = musica rock

rd = rondò (inclusi i rondò strumentali)

rg = ragtime

ri = ricercari (I ricercari non fugati del 16mo e 17mo secolo sono classificati come fantasie)

rp = rapsodie

rq = requiem

sg = canzoni

Ad eccezione dei casi in cui viene assegnato un codice specifico

si = sinfonie

sn = sonate

sp = poemi sinfonici

sq = balli di piazza

st = studi ed esercizi

su = suite

sy = sinfonie

tc = toccate

musica strumentale designata come tale

ts = triosonate

vr = variazioni

wz = walzer

zz = altre forme

\$b = Strumenti o Voci per ensemble

\$c = Strumenti o Voci per solisti

I sotto-campi \$b e \$c utilizzano gli stessi codici. Ciascuno dei due può essere lungo due o quattro caratteri, ed è ripetibile. I primi due caratteri indicano il tipo di voce o strumento specificata; il terzo ed il quarto sono cifre decimali che specificano il numero di una particolare voce o strumento. Quando il numero è minore di zero, la terza cifra corrisponde a zero. Se non si conosce il numero, il sotto-campo sarà lungo solo due caratteri.

Ottoni

Musica elettronica

ba = corno
bb = tromba
bc = cornetta
bd = trombone
be = tuba
bf = tromba baritono
bt = tromba etnica
bu = non specificato
bz = altro

ea = sintetizzatore
eb = nastro
ec = computer
ed = onde Martenot
eu = non specificato
ez = altro

Coro

ca = misto
cb = femminile
cc = maschile
cd = voci bianche
ct = etnico
cu = non specificato

Strumenti a tastiera

ka = pianoforte
kb = organo
kc = clavicembalo
kd = clavicordo
ke = continuo
kf = celesta
ku = non specificato
kt = etnico
kz = altro

Orchestra

oa = orchestra completa
ob = orchestra da camera
oc = orchestra d'archi
od = banda
oe = orchestra da ballo
of = ensemble più ampio – banda di ottoni
(ottoni con raddoppi, con o senza percussioni)
ot = etnica
ou = non specificato
oz = altro

Voci

va = soprano
vb = mezzo-soprano
vc = contralto
vd = tenore
ve = baritono
vf = basso
vg = controtenore
vh = voci – voce alta
vi = voci – voce media
vj = voci – voce bassa
vt = etnica

vu = non specificato

Percussioni

pa = timpani

pb = xilofono

pc = marimba

pd = batteria

pt = etnico

pu = non specificato

pz = altro

Strumenti a corde pizzicate

ta = arpa

tb = famiglia delle chitarre

tc = famiglia dei liuti

td = mandolino

tt = etnico

tu = non specificato

tz = altro

Strumenti ad arco

sa = violino

sb = viola

sc = violoncello

sd = contrabbasso

se = viol

sf = viola d'amore

sg = viola da gamba

st = etnico

su = non specificato

sz = altro

Legni

wa = flauto

wb = oboe

wc = clarinetto

wd = fagotto

we = ottavino

wf = corno inglese

wg = clarinetto basso

wh = flauto dolce

wi = sassofono

wt = etnico

wu = non specificato

wz = altro

Non specificato

uu = non specificato

Esempi:

128 ##\$afg\$b03

200 1#\$aFour fugues for guitar trio

Catalogazione di una fuga per trio di chitarre. La dichiarazione del titolo avviene in un campo di tipo 200 (Titolo e dichiarazione di responsabilità)

128 ##\$avr\$cka01

128 ##\$avr\$bka01\$csa01\$scsc01

128 ##\$avr\$bka01\$csa01\$scsb01\$scsc01

200 1#\$aVariations for solo piano, piano trio and piano quartet

Variazioni per pianoforte solista, trio d'archi e pianoforte e quartetto d'archi e pianoforte

323 – Note sugli interpreti (materiale video e registrazioni audio)

Appartiene ai campi del blocco 3 (Note), ed include i nomi di attori, esecutori, narratori, presentatori, ecc. di un certo materiale proiettato o video, il nome dei partecipanti, esecutori solisti o gruppi di esecutori, ed il medium di esecuzione delle registrazioni sonore.

E' opzionale e ripetibile per ogni categoria di partecipanti. Ha due indicatori, entrambi vuoti, ed un solo sotto-campo, \$a (“Testo della nota”)

Esempio:

323 ##\$aMarshall Moss, violin; Neil Roberts, harpsichord

Un esempio completo

Riportiamo ora un esempio di record UNIMARC completo, così come viene inserito nel nastro magnetico di interscambio.

ETICHETTA DEL RECORD

00796nas0#22002531##450#

DIRECTORY (le virgole separano tag, lunghezze dei campi e posizioni di partenza. Virgole e spazi, così come le indicazioni delle parti del record sono stati aggiunti per chiarezza, e non compaiono nel record vero e proprio)

001,0009,00000 011,0031,00009 020,0017,00040 100,0041,00057 101,0008,00098
102,0007,00106 110,0016,00113 200,0009,00129 207,0051,00138 210,0037,00189
215,0027,00226 326,0011,00253 421,0054,00264 447,0050,00318 447,0039,00368
530,0025,00407 670,0077,00432 676,0012,00509 801,0021,00774@

CAMPI DATI (i numeri dei campi, ad esempio [102] sono stati aggiunti per chiarezza, e non compaiono nel record)

[001]B8132254@[011]##\$a0261-6726\$d£0.20 per issue@[020]##\$aGB\$bB8132254@

[100]##\$a19811005b19811982|||y0engy01#####a@[101]0#\$aeng@[102]##\$aGB@[110]##\$aacaz###0yy0@

[200]1#\$aGirl@[207]#1\$aNo. 1 (14th Feb. 1981) - no. 65 (26th Apr. 1982)@

[210]##\$aLondon\$cIPC Magazines\$d1981-1982@[215]##\$av.\$cchiefly ill.\$d30cm@

[326]##Weekly@

[421]#1\$1001B8211156\$1011##\$a0262-9208\$12001#\$aGirl annual@

[447]#1\$1001B8207648\$1011##\$a0262-1185\$12001#\$aDreamer@

[447]#1\$1001B8233867\$12001#\$aGirl and dreamer@

[530]1#\$aGirl\$b(London, 1981)@

[670]##\$b0225398\$c*z11030*serials in English*igirls*z60030*atexts\$e0000167\$ztexts

[676]##\$a052\$v19@

[801]#3\$aGB\$bUk\$c19821216@%

Conclusioni

Conclusa questa breve panoramica sul formato UNIMARC, possiamo riassumere i concetti principali in tre punti: UNIMARC è

- completo
- standardizzato

- flessibile

Diremo che è un formato *completo* perchè tramite i record UNIMARC è possibile descrivere in maniera dettagliata tutti i metadati degli oggetti catalogati, inserendoli negli appositi campi; è un formato *standardizzato* perchè, essendo un'implementazione dello standard ISO 2709 (quindi riconosciuto in campo internazionale) ha delle regole ben definite, viene controllato e supervisionato da un ente centrale, il PUC, anche per quanto riguarda le estensioni future, ed in generale si presta quindi bene all'interscambio di dati bibliografici e bibliografico-musicali. Infine è un formato *flessibile* perchè, soprattutto tramite la possibilità di utilizzare i campi del blocco 8 e 9 per l'inserimento di dati rispettivamente ad uso internazionale e domestico/locale, ed i campi del blocco 3 (note), permettono l'inserimento anche di quei dati che non trovano posto in un campo specifico, e l'adattamento del formato alle singole esigenze ai fini dell'interscambio di dati.

Per quanto riguarda gli sviluppi futuri dello standard, l'UNIMARC è già ampiamente adottato come formato di interscambio di informazioni bibliografiche riguardanti opere cartacee (anche se il suo utilizzo in campo audio è ancora scarsamente diffuso), e le modifiche che verranno apportate alle successive versioni riguarderanno prevalentemente la definizione di campi speciali per usi particolari, ma non è detto che, se l'utenza lo richiedesse a seguito di un incremento nel trattamento di oggetti audio, non venga magari "potenziato" il formato anche sotto il punto di vista della catalogazione di questi oggetti

Capitolo III - Un'alternativa all'UNIMARC: il formato MAG

Introduzione

Un altro metodo per rappresentare i metadati di un determinato oggetto digitale di carattere musicale è costituito dallo standard MAG.

Il MAG (Metadati Amministrativi e Gestionali) in sé stesso non è altro che uno schema XML pensato per raccogliere i metadati gestionali relativi agli oggetti digitali prodotti in un progetto di digitalizzazione. Le varie caratteristiche ed informazioni relative all'oggetto da catalogare sono quindi rappresentate sotto forma di tag XML⁸.

Questo formato è stato sviluppato dal Comitato omonimo a partire dal luglio 2003 (è dunque uno standard abbastanza recente, e ciò giustifica in parte la carenza di informazioni reperibili al riguardo⁹) per iniziativa dell'ICCU, con lo scopo di proseguire il lavoro fino a quel momento svolto dal Gruppo di Studio sugli Standard e le applicazioni dei metadati nei Beni Culturali nell'ambito dei programmi di digitalizzazione, e di effettuare un'attività di promozione, supporto, gestione ed evoluzione del MAG. Il Comitato MAG inoltre si occupa della produzione di manuali e linee guida¹⁰, dà assistenza agli implementatori, fornisce la formazione e promozione del formato e si occupa dei rapporti con altri progetti del settore e con le agenzie bibliografiche nazionali.

Rispetto al formato UNIMARC descritto nel capitolo precedente, lo standard MAG si colloca in un'ottica più "informatizzata", in quanto non solo si pone l'obiettivo dell'estrazione automatica dei metadati da un oggetto digitalizzato, ma appoggiandosi al formato XML è dunque esplicitamente creato per essere fruito tramite elaboratore.

Struttura del formato

⁸ Maggiori informazioni si possono trovare in [Gilmour 2003]

⁹ Soprattutto di carattere audio. Esempi commentati MAG riferiti alle immagini si trovano su [Bergamin]

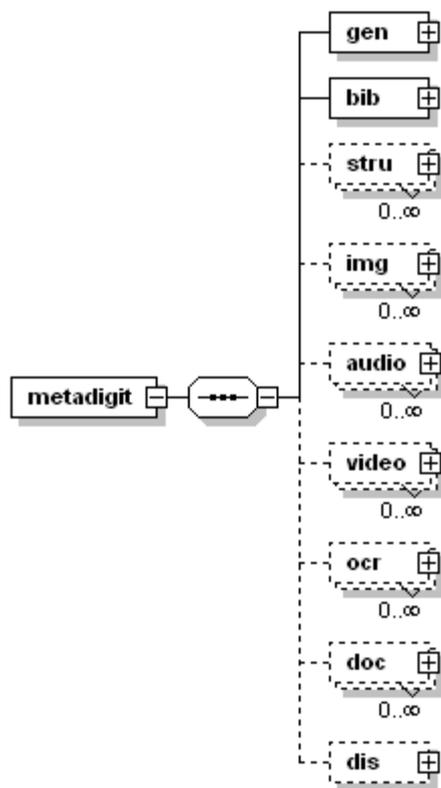
¹⁰ E' recentissima (7 aprile 2006) la pubblicazione sul sito dell'ICCU (<http://www.iccu.sbn.it/>) del manuale della versione 2.0.1 del formato, che aggiorna la precedente versione con una forma più scorrevole e l'introduzione di esempi commentati, anche se nessuno di essi è riferito all'ambito audio. [MAG 2006]

Lo schema generale del Mag è composto da 9 sezioni, ognuna riguardante un tipo specifico di metadati:

- **gen**: informazioni generali sul progetto e sul tipo di digitalizzazione
- **bib**: metadati descrittivi sull'oggetto digitalizzato
- **stru**: metadati strutturali
- **img**: metadati relativi alle immagini fisse
- **ocr**: metadati relativi al riconoscimento ottico del testo
- **doc**: sezione utilizzata per descrivere file di testo (ex. pdf o rtf)
- **audio**: metadati relativi agli oggetti audio digitalizzati
- **video**: metadati relativi alle immagini in movimento
- **dis**: contiene informazioni sulla fruibilità dell'oggetto digitale.

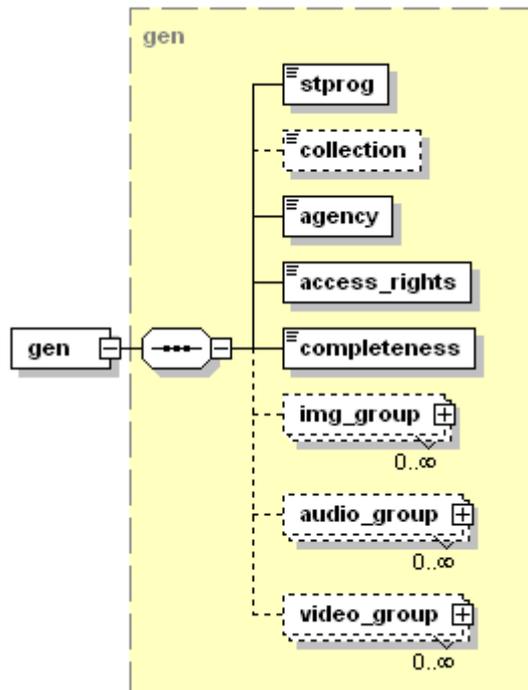
Le prime due sezioni sono obbligatorie per ogni record MAG, mentre le altre sono opzionali.

A monte esiste però un tag "radice", dal quale si diramano poi le sezioni: l'elemento root metadigit. Nella figura seguente viene mostrata sotto forma di schema ad albero la struttura principale di un record MAG.



Nella nostra trattazione verranno prima descritte in maniera sintetica le sezioni obbligatorie GEN e BIB, per poi passare all'analisi più approfondita della sezione AUDIO e dei sottoelementi di maggior interesse nell'ambito delle informazioni audio contenute nei database multimediali.

Sezione GEN



La sezione gen raccoglie informazioni generali sul progetto e sul tipo di digitalizzazione effettuata. Contiene al suo interno 8 sottosezioni (4 obbligatorie e 4 opzionali):

stprog = riferimento a una risorsa di rete che descrive gli standard di progetto. Si tratta, in pratica, di un link ad una pagina web

collection = riferimento alla collezione di cui farà parte la risorsa digitale

agency = agenzia responsabile del processo di digitalizzazione. Per questa sezione si raccomanda l'utilizzo della sintassi UNIMARC definita nel campo 801

access_rights = riguarda le condizioni di accesso all'oggetto analogico originale, e se quindi è destinato all'uso interno dell'istituzione o se è ad uso pubblico

completeness = livello di completezza della digitalizzazione dell'oggetto (parziale o completa)

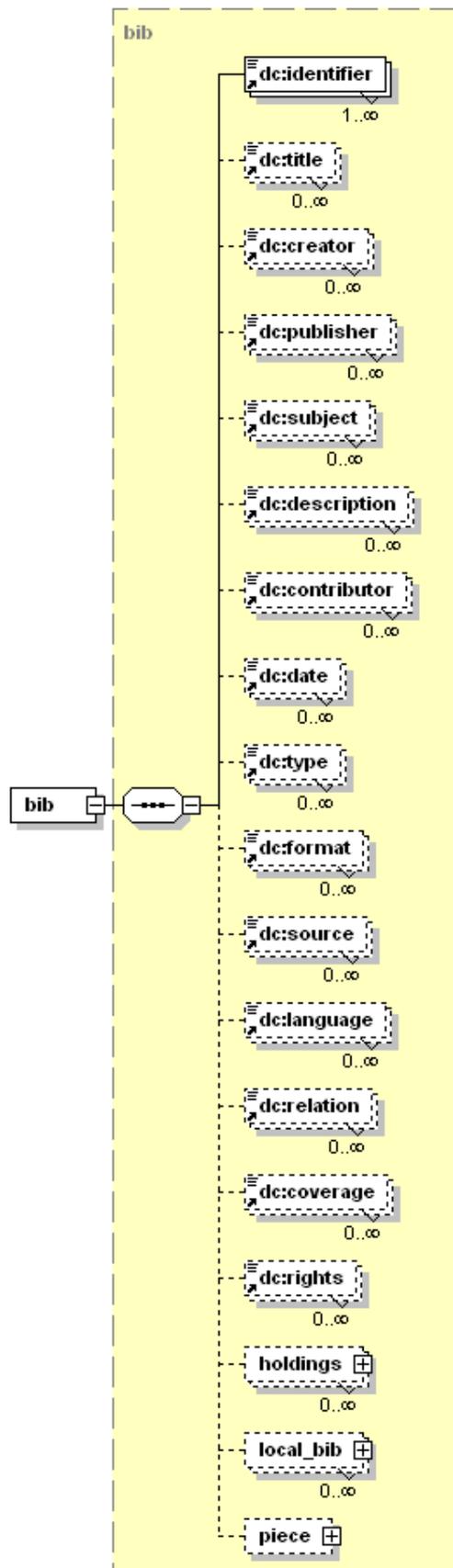
img_group = caratteristiche comuni a gruppi omogenei di immagini (immagini con le stesse caratteristiche tecniche, ex. risoluzione)

audio_group = caratteristiche comuni a gruppi omogenei di file audio (si usa quando molti file audio condividono le medesime caratteristiche tecniche; nel caso esistano più gruppi di file audio con caratteristiche tecniche comuni l'elemento può essere ripetuto)

video_group = caratteristiche comuni a gruppi omogenei di file video

Sezione BIB

La sezione bib contiene i metadati descrittivi relativi all'oggetto analogico digitalizzato o comunque all'oggetto fonte



Questa sezione contiene 18 sottosezioni, delle quali solo la prima è obbligatoria per ogni record MAG.

dc:identifier = identificatore univoco della risorsa analogica. Può essere un numero progressivo o di codice, o qualunque tipo di identificazione

dc:title = nome dato all'oggetto digitalizzato (titolo)

dc:creator = entità che ha la responsabilità principale della produzione del contenuto della risorsa analogica. Può contenere il nome del compositore e/o dell'autore di un'opera musicale. Se ripetuto, può contenere il nome della persona responsabile della digitalizzazione o del progetto

dc:publisher = entità responsabile della produzione della risorsa analogica, disponibile nella sua forma presente. Può contenere ad esempio il nome dell'editore di una partitura o della casa discografica di una registrazione sonora.

dc:subject = argomento della risorsa analogica

dc:description = spiegazione del contenuto di una risorsa analogica. Una descrizione può includere (ma non è limitata solo a): un riassunto analitico, un indice, un riferimento al contenuto di una rappresentazione grafica o un testo libero del contenuto

dc:contributor = entità responsabile della produzione di un contributo al contenuto della risorsa analogica

dc:date = data associata a un evento del ciclo di vita della risorsa analogica

dc:type = natura o genere del contenuto della risorsa analogica

dc:format = caratteristiche fisiche della risorsa analogica. In questa sottosezione possono essere incluse le dimensioni fisiche dell'oggetto

dc:source = riferimento a una risorsa analogica dalla quale è derivata la risorsa in oggetto

dc:language = lingua del contenuto intellettuale della risorsa analogica

dc:relation = riferimento ad una risorsa analogica correlata

dc:coverage = estensione o scopo del contenuto della risorsa analogica

dc:rights = informazioni sui diritti esercitati sulla risorsa analogica. Spesso contiene informazioni sullo stato e sulla presenza del copyright sull'opera.

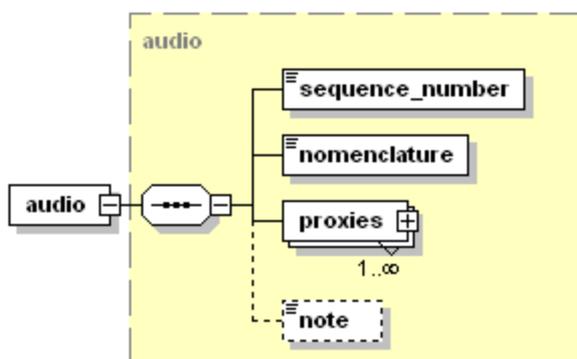
holdings = numero di inventario e collocazione della risorsa analogica digitalizzata

local_bib = contiene sistemi di catalogazione specializzati relativi alla singola biblioteca

piece = unità fisica componente di un superiore (es.: fascicolo di un seriale, parte di una unità bibliografica). Contiene le informazioni (data di inizio/fine della serie, numero di parti, identificazione, ecc...) relative alle serie bibliografiche

Sezione AUDIO

La sezione audio contiene i metadati relativi ad una traccia audio, o meglio, al file che la contiene, che rappresenta l'oggetto digitale incluso nel record.



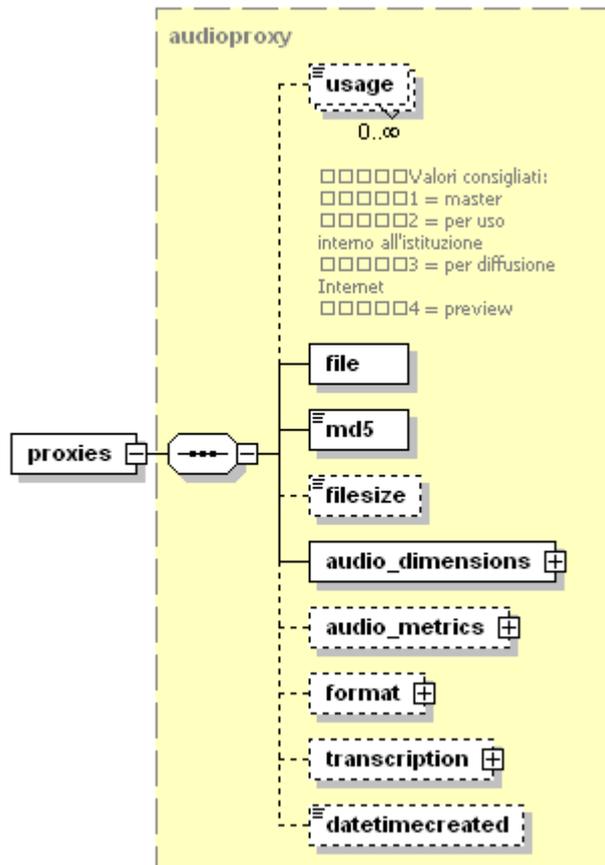
La sezione audio contiene tre sottosezioni più una dedicata alle note.

sequence_number = numero progressivo che identifica un contenuto audio all'interno del file XML. A questa sottosezione è imposto un vincolo di unicità, per garantire l'identificazione del singolo oggetto digitale.

nomenclature = descrizione della traccia audio (“ouverture”, “preludio”, ecc.)

proxies = descrizione della traccia audio digitale che riproduce l’oggetto fisico individuato da `sequence_number`.

A sua volta l’elemento proxies ha uno schema di questo tipo:



usage = modalità d'uso consigliata per la traccia in funzione delle applicazioni per le quali ne è prevista la fruizione. Può assumere i seguenti valori:

- 1: master
- 2: alta risoluzione
- 3: bassa risoluzione
- 4: anteprima
- a: il depositario dei diritti non usufruisce del copyright
- b: il depositario dei diritti usufruisce del copyright

file = link al file multimediale

md5 = impronta del file

filesize = dimensione del file espressa in bytes

audio_dimensions = dimensione della traccia audio. Contiene il sottoelemento duration, che indica la durata della traccia audio.

audio_metrics = descrive le caratteristiche tecniche della traccia audio. Contiene tre sottoelementi:

samplingfrequency = frequenza di campionamento espressa in kHz. Possibili valori: 8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 96.

bitpersample = numero di bit per campione. Possibili valori: 8, 16, 24

bitrate = kbit per secondo del campione. Possibili valori: 24, 32, 48, 56, 64, 96, 128, 160, 192, 256, 320, 384.

format = tipologia della traccia audio digitale e modalità di compressione. Contiene i sottoelementi:

name = formato della traccia audio (WAV, MP3, ecc...)

mime = tipologia MIME del file audio

compression = algoritmo di compressione del file audio. Possibili valori: Uncompressed, Linear PCM, MPEG-1 Layer 1, MPEG-1 Layer 2, MPEG-1, Layer 3, AC3, Dolby, DTS.

channel_configuration = definisce lo schema di configurazione dei canali audio. Possibili valori: Mono, Dual mono, Joint stereo, Stereo, 2 ch., 4 ch., 5.1 ch., 6.1 ch.

transcription = modalità di trascrizione digitale della traccia audio. Contiene i sottoelementi:

`sourcetype` = informazione sul tipo di sorgente dei dati audio. Possibili valori (esempi): registrazione audio, disco, nastro (open reel), nastro (cassetta), cilindro, rullo, disco 33 giri, disco 45 giri, disco 78 giri, altro.

`transcriptionagency` = nome dell'istituzione che ha effettuato la digitalizzazione

`transcriptiondate` = data della digitalizzazione

`devicesource` = informazioni generali sugli apparecchi utilizzati per la digitalizzazione (microfono, registratore, ecc...)

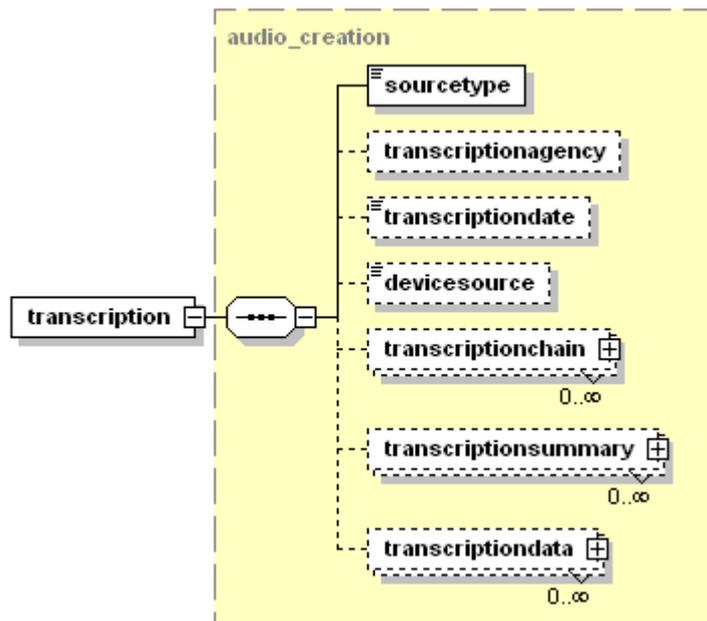
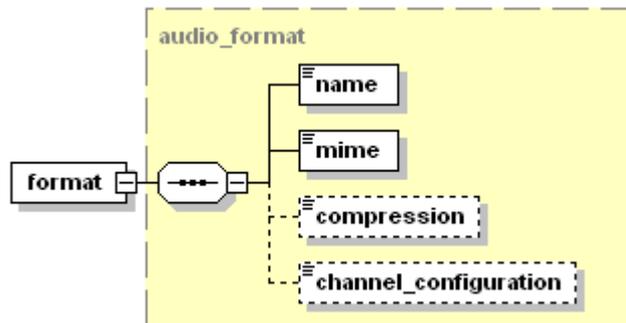
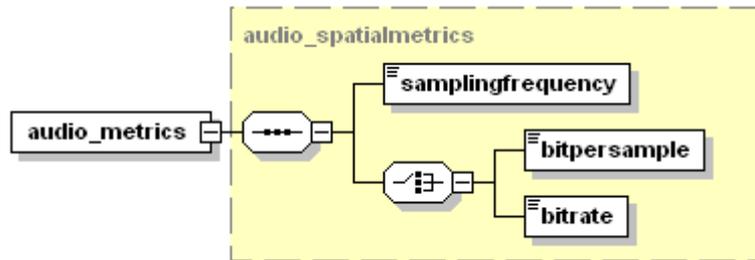
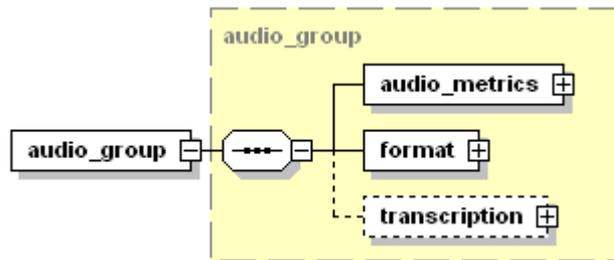
`transcriptionchain` = descrive la catena di trascrizione, ed include altre informazioni sulla digitalizzazione, come tipo di dispositivo utilizzato, tipo di software, ecc...

`transcriptionsummary` = raccoglie i dati misurati durante il processo di digitalizzazione per i valori di sintesi. Permette di definire nomi, tipi e valori delle grandezze fisiche misurate, consentendone una nidificazione gerarchica (ad esempio valori per canale)

`transcriptiondata` = raccoglie le sequenze di dati misurate durante il processo di digitalizzazione. Permette di definire nomi, tipi e valori delle grandezze fisiche misurate, consentendone una nidificazione gerarchica (ad esempio valori per canale).

`datetimecreated` = data di realizzazione della digitalizzazione

Nell'elemento `audio_group` della sezione GEN sono presenti elementi simili a quelli qui descritti, con la differenza che le informazioni contenute si applicano ad un gruppo di oggetti digitalizzati invece che ad un singolo oggetto. Nelle figure seguenti sono rappresentati gli alberi dell'elemento `gen/audio_group` ed i suoi sottoelementi principali. I sottoelementi `audio_metrics`, `format` e `transcription` hanno lo stesso schema di quando vengono utilizzati in funzione dell'elemento AUDIO.



Concludendo: un esempio di record MAG

Qui di seguito viene riportato un esempio di record MAG, che descrive le caratteristiche di un oggetto digitale contenente la registrazione audio del Primo Concerto Brandeburghese di J. S. Bach. La registrazione è in formato MP3 a 128 Kbps.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <metadigit xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:niso="http://www.niso.org/pdfs/DataDict.pdf"
xmlns:xlink="http://www.w3.org/TR/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns="http://www.iccu.sbn.it/metaAG1.pdf"
xsi:schemaLocation="http://www.iccu.sbn.it/metaAG1.pdf metadigit.xsd"
version="1.0">
- <gen>
  <stprog>http://www.dds.it </stprog>
  <agency>IT:DDS</agency>
  <access_rights>0</access_rights>
  <completeness>0</completeness>
</gen>

- <bib level="c">
  <dc:identifier xsi:type="DDS">SBL0285585</dc:identifier>
  <dc:title>Concerti Brandeburghesi</dc:title>
  <dc:creator> Bach, Johann Sebastian </dc:creator>
  <dc:type>File audio</dc:type>
  <dc:format>Registrazione audio</dc:format>
</bib>

- <audio>
  <nomenclature>Primo Concerto Brandeburghese </nomenclature>
  <proxies>
    <usage>2</usage>
    <file>./174805/BACH/B12547.WAV</file>
    <md5>956a4b885c1279218cf978e7a015f846e</md5>
    <filesize >20047000</filesize >
    <audio_dimensions>
      <duration>00:20:12</duration>
    </audio_dimensions>
    <audio_metrics>
      <samplingfrequency>44.1</samplingfrequency >
      <bitpersample >16</bitpersample>
      <bitrate >128</bitrate >
    </audio_metrics>
    <format >
      <name>MP3</format >
      <mime >audio/mp3<mime >
      <compression >Mpeg-1 Layer 3</compression >
      <channel_configuration >Stereo</channel_configuration >
    </format >
    <datetimecreated >19072006</datetimecreated >
```

</proxies>

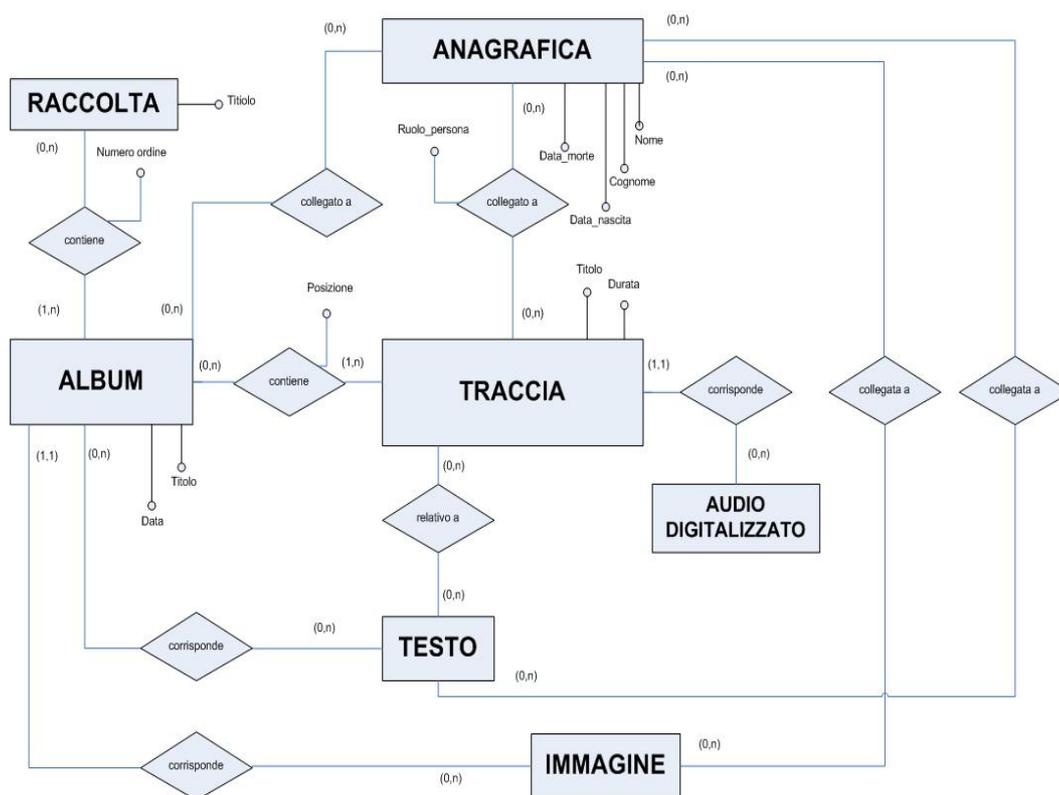
</audio>

...

Come si può notare, il formato MAG è molto più “esplicito” nella sua struttura rispetto all’UNIMARC, e come detto precedentemente, più specificatamente indirizzato verso la descrizione di metadati di oggetti multimediali, piuttosto che verso la trascrizione di schede catalografiche in formato cartaceo. Inoltre il prevedibile sviluppo di tecniche di creazione, individuazione ed estrazione automatica di record MAG direttamente dall’oggetto digitale fanno di questo formato (anche se ancora scarsamente utilizzato per descrivere oggetti digitali di tipo audio) un’interessante alternativa all’UNIMARC soprattutto in prospettiva futura, in quanto la sua evoluzione si profila molto più rapida, come si conviene ad un formato nato per la fruizione su computer. I manuali per gli utenti e le *guidelines* sono aggiornati relativamente spesso, e daranno conto di eventuali modifiche alla struttura dello schema.

Capitolo IV - Rappresentazione dei metadati

Ora che si sono introdotti i due formati per la descrizione dei metadati, e si conoscono gli attributi delle entità concettuali, che in pratica rappresentano proprio i metadati degli oggetti digitali contenuti all'interno del nostro database multimediale, si può tentare di “incrociare” questi elementi, cercando di individuare quali campi UNIMARC e quali elementi dello schema MAG possono rappresentarli in maniera più completa ed efficace.



Dallo schema concettuale del database riportato nella figura, il primo attributo ad essere definito è il titolo della raccolta (che chiameremo **raccolta_titolo**, convenzione che verrà usata anche per gli altri metadati): questo metadato può essere incluso in un campo UNIMARC 200\$a (Titolo vero e proprio), o in un elemento <dc:title> della sezione BIB del MAG. Tuttavia l'utilizzo dell'elemento <dc:title> è piuttosto forzato in quanto, come detto precedentemente, il MAG si presta meglio a descrivere dati intrinseci degli oggetti digitali, piuttosto che le loro informazioni catalografiche¹¹. In alternativa, il titolo della raccolta di provenienza della registrazione potrebbe essere inserita in un elemento <note>

¹¹ Infatti, l'elemento <dc:title> andrebbe utilizzato per inserire il titolo delle singole registrazioni audio.

della sezione AUDIO. **Album_titolo** può essere però inserito in un campo UNIMARC 200\$a se si tratta di un album a sé stante, o in un campo 200\$i (Nome della parte), se facente parte di una raccolta: parallelamente in MAG possono essere utilizzati l'elemento <dc:title> e il sottoelemento <part_name> contenuto nell'elemento <piece> della sezione BIB.

Essendo un album un'opera solitamente coperta da copyright, l'attributo **album_data**, ossia la data di uscita di un album, può venire descritto in UNIMARC dal sotto-campo 100\$a/8-16 utilizzando il codice h (monografia con data effettiva e data del copyright), ed inserendo la stessa data come data di pubblicazione e data di partenza del copyright, adottando questo utilizzo come convenzione interna all'ente di catalogazione¹²; la rappresentazione in MAG si potrebbe ottenere tramite l'utilizzo dell'elemento <dc:date>del BIB. L'attributo **traccia_titolo** può agevolmente essere inserito in un campo UNIMARC 200\$a se si tratta di una traccia isolata (caso piuttosto raro), oppure in un campo 200\$i, intendendo la traccia come parte di un album (che di per sé è una raccolta di tracce); in MAG il nome della traccia viene descritto nell'elemento <dc:title>, oppure nell'elemento <nomenclature> della sezione AUDIO.

La durata della traccia audio (**traccia_durata**) è di semplice inserimento in entrambi i formati, avendo in UNIMARC il campo 127\$a (durata delle registrazioni sonore) ed in MAG l'elemento <duration> all'interno di AUDIO/proxies/audio_dimensions.

L'organico dalla quale è interpretata l'opera (**traccia_organico**) viene descritto, a seconda dei casi, dai campi 128\$b e/o 128\$c (Strumenti per voci o ensemble e Strumenti per solisti). Per quanto riguarda il MAG si dovrà inserire l'informazione sotto forma di nota alla traccia audio, non essendo previsti campi specifici a questo scopo.

Riguardo all'entità ANAGRAFICA, per inserire nome e cognome di una persona che ha partecipato alla creazione dell'opera (**anagrafica_nome** e **anagrafica_cognome**) si possono usare rispettivamente i campi UNIMARC 200\$f e 200\$g (Dichiarazioni di responsabilità), e in MAG gli elementi di BIB <dc:creator> e <dc:contributor> (creatore e contributo dell'opera). Per quanto riguarda le date di nascita e di morte (**anagrafica_data_nascita** e **anagrafica_data_morte**), non sono previsti campi UNIMARC a questo scopo, quindi queste informazioni andrebbero inserite sotto forma di note (campo 300\$a, note generali), e così si dovrebbe fare anche in MAG.

¹² Teoricamente, e sempre sotto forma di convenzione interna, si potrebbe utilizzare il sotto-campo 100\$a/0-7 (Data assegnata al file), inserendo, al posto della data di creazione del record, la data di uscita dell'album. La data di creazione del record potrebbe venire inserita successivamente in un campo di tipo 801\$c (Data di trascrizione).

Passando agli attributi delle relazioni, l'attributo **collegato a ruolo persona** trova posto in UNIMARC nei campi 200\$f o 323\$a (note sugli interpreti), a seconda se la persona alla quale si riferisce sia un autore/scrittore o sia invece in esecutore/strumentista, e nel campo note del MAG.

Gli attributi **contiene_posizione** e **contiene_numero_ordine** possono essere trattati analogamente nel campo 200 \$h di UNIMARC (Numero della parte, intendendo le tracce come parti dell'album) e nel sottoelemento <part_number> all'interno di BIB/piece nel MAG.

Alcune considerazioni

La tabella sottostante riassume schematicamente quanto detto:

<i>Attributo</i>	<i>Rappresentazione UNIMARC</i>	<i>Rappresentazione MAG</i>
raccolta titolo	200\$a	<dc:title>
album titolo	200\$a 200\$i	<dc:title> <part_name>
album data	100\$a/8-16	<dc:date>
traccia titolo	200\$a 200\$i	<dc:title> <nomenclature>
traccia durata	127\$a	<duration>
traccia organico	128\$b 128\$c	<note>
anagrafica nome	200\$f 200\$g	<dc:creator> <dc:contributor>
anagrafica data nascita	300\$a	<note>
anagrafica data morte	300\$a	<note>
collegato a ruolo persona	200\$f 323\$a	<note>
contiene posizione	200\$h	<part_number>
contiene numero ordine	200\$h	<part_number>

Da questo confronto si può immediatamente notare il grande utilizzo degli elementi appartenenti alla sezione BIB dello standard MAG. Come già detto, il BIB è la sezione che descrive i metadati descrittivi relativi agli oggetti digitali, e tenendo presente che l'UNIMARC è nato proprio per rappresentare questo tipo di metadati (anche se non relativi agli oggetti digitali), si può stabilire una certa corrispondenza tra alcuni campi di questi

formati, anche se in alcuni casi è necessario adottare delle interpretazioni “convenzionali” (ad esempio considerando le tracce come “parte” dell’album, utilizzare le note come campo di descrizione dei metadati che non trovano posto in campi specifici (traccia_organico, anagrafica_data_nascita, anagrafica_data_morte, collegato a _ruolo_persona) o come nel caso della rappresentazione UNIMARC dell’ attributo album_title), ed accettare una certa “forzatura” nell’uso di certi elementi (come nell’utilizzo di <dc:title> per la descrizione di album_titolo).

Si può inoltre evincere che, nonostante i due formati possano essere considerati alternativi, la soluzione migliore ai fini della completezza e dell’efficacia della rappresentazione dei metadati sarebbe quella di adottare un approccio di tipo misto, utilizzando l’UNIMARC per i metadati descrittivi (ossia quelli delle entità previste nel nostro schema), ed il MAG per i metadati relativi agli oggetti digitali incorporati all’interno dei record del database. In questo modo sarebbe possibile sfruttare sia la completezza descrittiva offerta dall’UNIMARC che la flessibilità caratteristica del MAG per quanto riguarda la descrizione delle caratteristiche tecniche degli oggetti digitali.

Conclusioni

Le problematiche da affrontare in sede di descrizione delle informazioni catalografiche degli oggetti audio digitali all’interno di database multimediali sono senza dubbio molte, e ne sono state illustrate solo una parte: infatti, il vero banco di prova del buon funzionamento di un linguaggio di descrizione di metadati si realizza nel momento dell’interscambio, quando i record creati da un’agenzia devono (o sarebbe meglio dire, dovranno) essere inviati ad un’altra. Solo allora si potrà avere il riscontro di quante e quali informazioni potrebbero venire perse o corrotte, aprendo così la strada alla fase di risoluzione dei problemi. Sarebbe inoltre auspicabile un’interazione tra i vari progetti di digitalizzazione di materiale audio e non: se infatti si riuscisse ad avere un buon numero di enti utilizzanti tutti lo stesso formato (o gli stessi formati), facilitando l’interscambio di record multimediali, si potrebbe arrivare a qualcosa di molto simile ad un grande “database multimediale centrale”, contenente una quantità notevolissima di dati e contenuti multimediali provenienti da varie agenzie, e che rappresenterebbe senza dubbio un grande

strumento per la ricerca bibliografica. Sappiamo tuttavia che questo grande database comporterebbe una progettazione (inclusa tutta la fase di stipulazione di accordi tra le agenzie) molto lunga e complessa, così come lunghe e complesse sarebbero la sua realizzazione e la sua gestione.

Ma il campo che offre più interessanti prospettive¹³ è quello dello sviluppo di sistemi sempre migliori di estrazione, gestione, catalogazione ed esportazione automatiche dei metadati degli oggetti digitali, che andranno anche ad influire sulla sopracitata fase di correzione degli errori, in quanto non si sarà (o dovrebbe essere) più dipendenti dalla trascrizione manuale dei record, che può dare luogo ad errori da parte dell'utente che la effettua. Più in generale, la collezione della Discoteca di Stato è di dimensioni considerevoli, e quindi è considerevole la mole di informazioni che il database (una volta trasferito sul portale Internetculturale) dovrebbe gestire, tenendo conto anche del grande numero di utenti che si troverebbero a voler effettuare ricerche di varia complessità al suo interno, e ciò pone anche un problema di natura fisica, ossia di ottimizzazione del database che si va a programmare, fase della quale non ci si è occupati nel nostro approccio al progetto e in questa trattazione. In questo elaborato sono stati illustrate le basi sulle quali si potrà poi procedere alla costruzione di un importante strumento di conservazione, diffusione e valorizzazione dei beni culturali a carattere musicale.

¹³ Vedi anche [Subrahmanian – Jajodia 1996]

Riferimenti

[UNIMARC 2001] *UNIMARC manual: authorities format* / International Federation of Library Associations and Institutions, IFLA Universal Bibliographic Control and International MARC Programme. IFLA 2001 www.ifla.org

[ISO 1996] Format for Information Interchange. Geneva, ISO, 1996 (ISO 2709-1996)

[Scolari 2000] Antonio Scolari, *Unimarc*. Ed. AIB, 2000

[Subrahmanian 1998] S. Subrahmanian, *Principles of multimedia database systems*. San Francisco, Calif. : Morgan Kaufmann Publishers, 1998

[Prabhakaran 1997] B. Prabhakaran, *Multimedia database management systems*. Boston : Kluwer Academic Publ., 1997

[Subrahmanian – Jajodia 1996] V.S. Subrahmanian, Sushil Jajodia, *Multimedia database systems : issues and research directions*. Berlin ; New York : Springer-Verlag, 1996

[MAG 2006] Comitato MAG / *MAG Metadati Amministrativi e Gestionali: Manuale Utente*, a cura di Elena Pierazzo. Versione: 2.0.1 aggiornamento all'8 marzo 2006
www.iccu.sbn.it

[Bergamin] Giovanni Bergamin, *Introduzione ai metadati*,
<http://www.unicam.it/ateneo/organizzazione/biblioteche/documenti/Bergamin.pdf>

[Gilmour 2003] Ron Gilmour, *XML : a guide for librarians*. Chicago : American Library Association, 2003

[Ramakrishnan – Gehrke 2004] R. Ramakrishnan, J. Gehrke. - *Sistemi di basi di dati*. Ed. McGraw-Hill

www.dds.it

www.internetculturale.it

<http://ifla.org>

www.iccu.sbn.it

www.w3.org

www.bncf.firenze.sbn.it

www.wikipedia.org

Ringraziamenti

Un grazie particolare:

- Al prof. Haus, e a tutti i professori;
- ad Adriano Baratè e Luca Ludovico, per la pazienza e la collaborazione dimostratemi;
- a mamma, papà, Federica e tutta la mia famiglia, per aver reso possibile tutto questo;
- a Marianna, per avermi accompagnato in questi quattro anni e averli resi ancora più straordinari...grazie di tutto!
- a Fabio, un amico vero (thanks, bro!)

Grazie anche:

- a Dante Tanzi e tutto il LIM;
 - a Mattia, Ale, Costanza, Cristina e tutto STCM;
 - a Gianluca, Marco, Paolo B., Simone, Luca, e alla mia band;
 - Fabio M. e gli amici del Corotrecime;
- ...e a tutti coloro che mi hanno portato fin qui.