

PROSPETTIVE PER LA DIGITALIZZAZIONE DEI DOCUMENTI MUSICALI: dall'Archivio Digitale della Musica alla Rete della Musica Italiana¹

Maurizio Messina – Biblioteca Nazionale Marciana
IV Conferenza Nazionale delle Biblioteche, Firenze, 7 novembre 2003

Questo intervento si integra con quello precedente presentato da Aurelio Aghemo, e intende sviluppare alcuni argomenti relativi alla digitalizzazione di documenti e contenuti musicali e dare conto di un progetto che, a partire dall'esperienza dell'*Archivio Digitale della Musica (Veneta)* (ADMV), si sta concretizzando nell'ambito della linea tematica "musica" del programma Biblioteca Digitale Italiana/Network Turistico Culturale (BDI/NTC), di cui ha parlato Luciano Scala nella sua introduzione. Vorrei premettere qualche breve considerazione di ordine generale.

1. Considerazioni generali

I PROGETTI DI DIGITALIZZAZIONE DEVONO ESSERE FINALIZZATI: Credo che possiamo ritenere definitivamente archiviata l'epoca in cui digitalizzare significava produrre una certa quantità di scansioni da collocare su una serie di CD-ROM, difficili da maneggiare e destinati a permanere a lungo sugli scaffali. Gli obiettivi di queste operazioni, non sempre definiti con chiarezza, variavano dalla necessità di ridurre l'uso degli originali analogici a quella di disporre di copie da cui trarre riproduzioni o pubblicazioni, con un uso simile a quello da sempre fatto dei microfilm. Obiettivi del tutto legittimi, beninteso, ma non sempre tali da giustificare gli investimenti e gli impegni di natura organizzativa necessari per conseguirli.

E' proprio l'esperienza di questi progetti che ci induce oggi a "pretendere" per i progetti di digitalizzazione la definizione di un quadro di riferimento che ci aiuti a CAPIRE "PERCHE'" E "PER CHI" DIGITALIZZIAMO. Occorre in altri termini definire i destinatari del nostro lavoro, il nostro *target* di riferimento, occorre chiarire, come diceva con molta efficacia Annamaria Tammaro in un recente intervento, SE INTENDIAMO OFFRIRE "DATI" O "SERVIZI". Digitalizzare costa, non solo in termini monetari, e ancora di più costerà mantenere accessibile nel tempo ciò che sarà stato digitalizzato; lo sforzo non si giustifica se non all'interno di una programmazione più ampia che abbia una precisa valenza sociale ed economica. In altri paesi è la sfera "educativa", nell'accezione più ampia possibile, a fare da sfondo e ad orientare queste attività. Sotto questo punto di vista il programma di Network Turistico Culturale, previsto dalla legge finanziaria 2003, con le sue modalità operative, penso ad es. all'*e-learning*, costituisce un'opportunità importante, certamente da sfruttare. Fra l'altro il marketing del territorio è un'attività di primario interesse anche per gli Enti locali, e non potrebbe essere diversamente in un paese come il nostro, che cela un tesoro dietro ogni angolo. Ci troviamo quindi su un terreno fertile, in cui le sinergie, ma anche la dialettica, fra amministrazione centrale e amministrazioni locali possono svilupparsi positivamente, e la cooperazione, quella cooperazione che ben conosce chi da molto tempo ormai opera in SBN, è possibile. La qualificazione culturale dell'offerta turistica, che costituisce in sostanza, mi pare, quello che ci viene chiesto, significa agganciare un settore economicamente trainante e servirsene per ASSICURARE LA SOSTENIBILITÀ DEI NOSTRI PROGETTI: un intervento di digitalizzazione, la costruzione di una biblioteca digitale, non finiscono mai, sono l'avvio di un processo continuo per la conservazione degli oggetti digitali creati, per l'adeguamento delle tecnologie, per l'aggiornamento professionale degli operatori, tutte attività che chiedono di poter contare su flussi finanziari certi, e probabilmente su un capitolo ordinario di spesa. ADM, ad esempio, è un

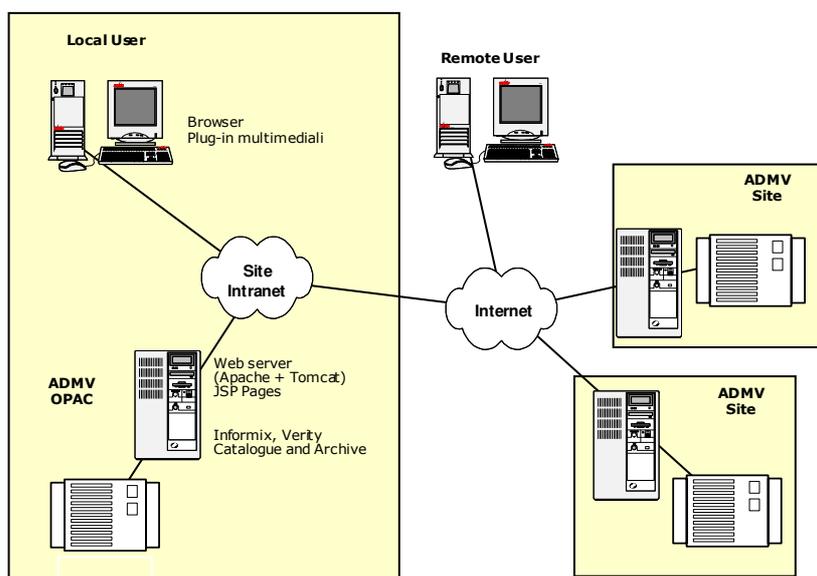
¹ Ringrazio Massimo Sarachino per avermi fornito documentazione essenziale per la redazione di questo intervento. Tutti i link sono stati controllati in data 13.3.2004

progetto sostanzialmente concluso, che ha creato un modello tecnologico, un modello organizzativo e, riteniamo, un modello dei servizi, ma che alla fine produrrà un numero relativamente ridotto di oggetti digitali. Occorre quindi trovare le risorse per andare avanti, tenendo presente che fermarsi significa pregiudicare anche i risultati fin qui ottenuti. In questo contesto l'organismo deputato al governo delle transazioni economiche derivanti dall'erogazione dei servizi nell'ambito del NTC, in collaborazione con il sistema bancario, avrà un ruolo determinante nel rendere accessibile il sistema e favorirne l'autofinanziamento. Certo, occorre mantenersi vigili ed esercitare, ove necessario, la critica. Il ruolo dei bibliotecari, degli archivisti, degli operatori dei musei e delle istituzioni culturali coinvolte nel *network* sarà quello di GARANTIRE LA QUALITÀ DEI CONTENUTI. Al di fuori di questo la nostra attività non avrebbe senso. Dobbiamo cioè continuare a fare il lavoro che abbiamo sempre fatto, tenendo presente che agiremo in un'arena in cui operano competitori che forse hanno valori ed obiettivi diversi dai nostri, e con i quali potrà succedere di entrare in rapporto dialettico, più o meno come capita nella filiera della produzione organizzazione e disseminazione dei contenuti digitali. A questo proposito, e tanto per insistere su un tema che è sotteso a tutto il mio discorso, quello della conservazione digitale, ritengo che ci sia un NESSO MOLTO FORTE FRA QUALITÀ E CONSERVAZIONE. Difficilmente sarà possibile reperire risorse per la conservazione di contenuti digitali, e pensiamo in particolare alla produzione audiovisiva, se questi non saranno contenuti di qualità, contenuti che valga effettivamente la pena di conservare.

2. Dall'Archivio Digitale della Musica (ADM) alla Rete della Musica Italiana (ReMI)

2.1 L'Archivio Digitale della Musica (ADM)

Di ADM si è già parlato a lungo, qui vorrei richiamare pochi aspetti fondamentali: Fin dallo studio di fattibilità abbiamo concepito ADM come un progetto e un sistema fortemente orientato da una parte ad erogare servizi ad un'utenza non di soli specialisti, tramite l'OPAC, e dall'altra alla conservazione di lungo termine degli oggetti digitali creati e della loro accessibilità, tramite l'adozione di standard emergenti per la *digital preservation*. La sua architettura prevede una rete distribuita di basi dati locali di record bibliografico-musicali e di *repositories* digitali di immagini e suoni in grado di interoperare fra loro attraverso l'utilizzo di formati dei dati e di protocolli di comunicazione e trasmissione standard, senza vincolo d'uso di un determinato software applicativo e senza realizzare una struttura indice centralizzata.



1. La rete ADM(V)

Chiunque condivida quegli standard è in grado di partecipare alla rete in modo sostanzialmente paritario. I singoli nodi ADM hanno tutti le medesime componenti logiche, se non fisiche:

- Un applicativo gestionale per la catalogazione dotato di funzionalità di export dei record in formato UNIMARC;
- Una base dati in grado di importare quei record, fruibile tramite un applicativo OPAC dotato anche di funzioni di puntamento ad oggetti digitali;
- Un servizio di digitalizzazione incaricato di produrre file digitali completi dei relativi Metadati Amministrativi e Gestionali (MAG): i MAG sono informazioni tecniche sugli oggetti digitali necessarie per identificarli univocamente, certificarne autenticità ed integrità, e permetterne la gestione, la conservazione permanente e l'uso libero o condizionato da parte degli utenti);²
- Un *repository*, o teca, digitale in grado di caricare e gestire oggetti digitali con i relativi MAG e strutturata per la conservazione e l'accessibilità permanenti agli oggetti digitali medesimi, conforme ad es. al modello OAIS.³

2.2 La Rete della Musica Italiana (ReMI)

La Direzione Generale per i Beni Librari ed il Comitato Guida della BDI, nel mettere a punto i contenuti della linea tematica “Musica” del programma BDI/NTC, hanno individuato nell'architettura, negli standard e nel modello dei servizi ADM il punto di partenza per costruire la Rete della Musica Italiana (progetto ReMI). Si tratta di un progetto le cui specifiche si vengono definendo in questi giorni, e la cui realizzazione sarà oggetto di una specifica procedura di gara. Naturalmente il passaggio da ADM a ReMI apre uno scenario molto più vasto e molto più complesso di quello che si è affrontato finora: ai fini del riutilizzo in ReMI delle soluzioni ADM è necessario, per quanto riguarda il contesto tecnologico, verificare eventuali esigenze di adeguamento a tecnologie hardware/software che nel frattempo possono essere evolute (ADM ha avuto una genesi piuttosto lunga, ha fatto il più possibile ricorso a componenti di ampia diffusione, occorre verificare se tali componenti restano commercialmente supportate o se il mercato suggerisce/impone adeguamenti o scelte diverse); andranno poi verificate e se del caso ingegnerizzate le componenti software ADM che possono essere influenzate dalle dimensioni della rete, dal numero delle transazioni e dal numero dei partner (ADM è stato dimensionato su 3 partner, occorre capire cosa va adeguato o modificato nel caso in cui i partner passino da 3 a 30 o ad n).

Elenco qui di seguito alcuni punti fermi e alcune differenze fra le due reti:

- Conferma dell'architettura distribuita, aperta, e della sostanziale “assenza di un centro” propria di ADM;
- Conferma della assoluta centralità degli standard (formato dei dati, protocolli di comunicazione etc);
- Differenziazione dei materiali: oltre alla musica manoscritta, oggetto principale delle attività ADM, la rete ReMI prevede l'accesso a documenti digitalizzati diversi: musica a stampa, trattati di teoria musicale, bozzetti di costumi o scenografie, fotografie, libretti, documenti d'archivio di teatri o altre istituzioni etc.;
- Forte accentuazione della presenza e del ruolo delle collezioni audiovisive, sonore, e multimediali, con particolare riferimento all'esperienza della Discoteca di Stato, attuale coordinatore del progetto ADM;

² Si veda: <http://www.iccu.sbn.it/schemag.htm>

³ Si veda: <http://ssdoo.gsfc.nasa.gov/nost/isoas/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

- Esigenza di prevedere livelli differenziati di partecipazione a ReMI (le collezioni musicali italiane sono estremamente frammentate, i loro titolari possono avere caratteristiche e capacità molto diverse).

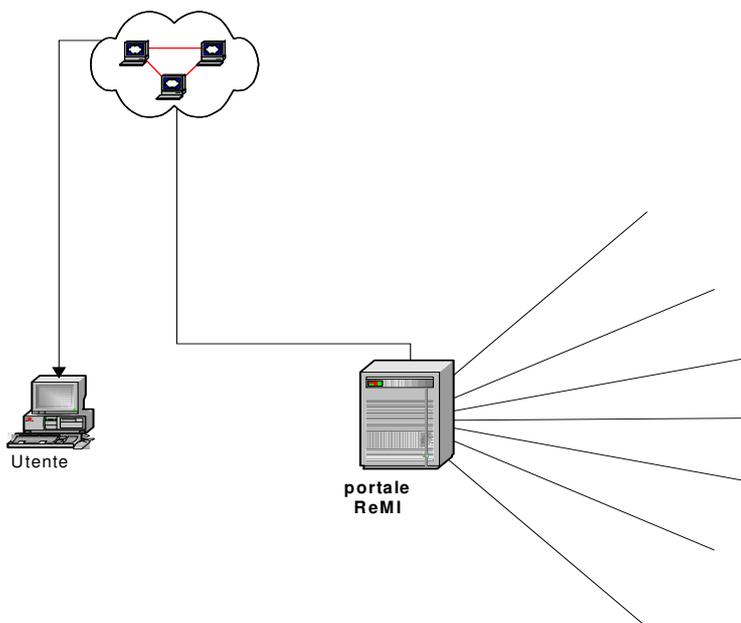
2.3 L'Architettura di ReMI

Tenuti presenti questi punti, e in particolare l'ultimo, la topologia prevista per la rete ReMI presenta delle differenze rispetto alla rete ADM, ed è a quest'ultima in qualche modo sovraordinata, nel senso che ADM può presentarsi come una delle reti accessibili da ReMI.

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

Osserviamo i vari componenti da vicino:

- IL PORTALE ReMI:

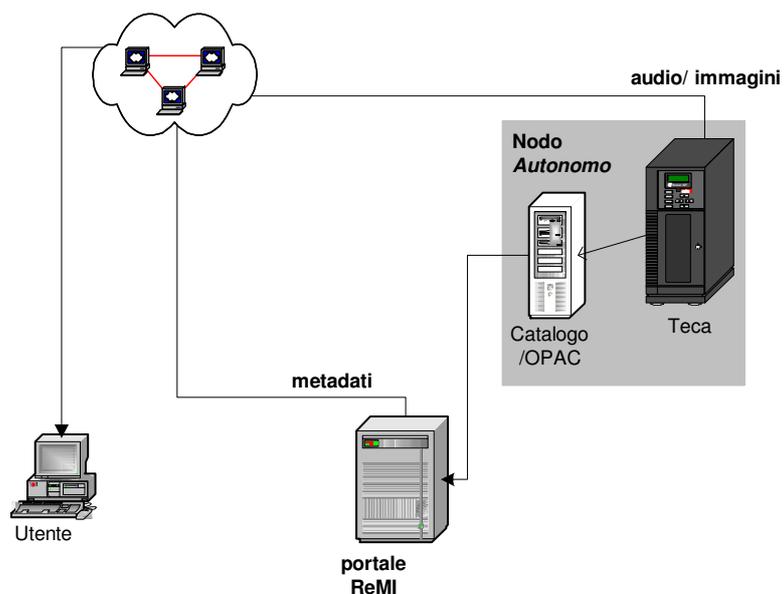


2. Il Portale ReMI

si tratta di una macchina dotata di un software che consente all'utente di interrogare in maniera integrata i nodi presenti sulla rete, e ottenere la prospettazione dei risultati. Si serve di un *web-server* ed un *application server* di comune utilizzo, ed è accessibile tramite un normale *browser web*. Ha quindi funzioni di interfaccia, sul modello di quelle dell'OPAC ADM. Non è un catalogo, non è un indice e non ospita record catalogafici. Dal punto di vista dell'utente è l'interfaccia OPAC di un sistema distribuito.

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

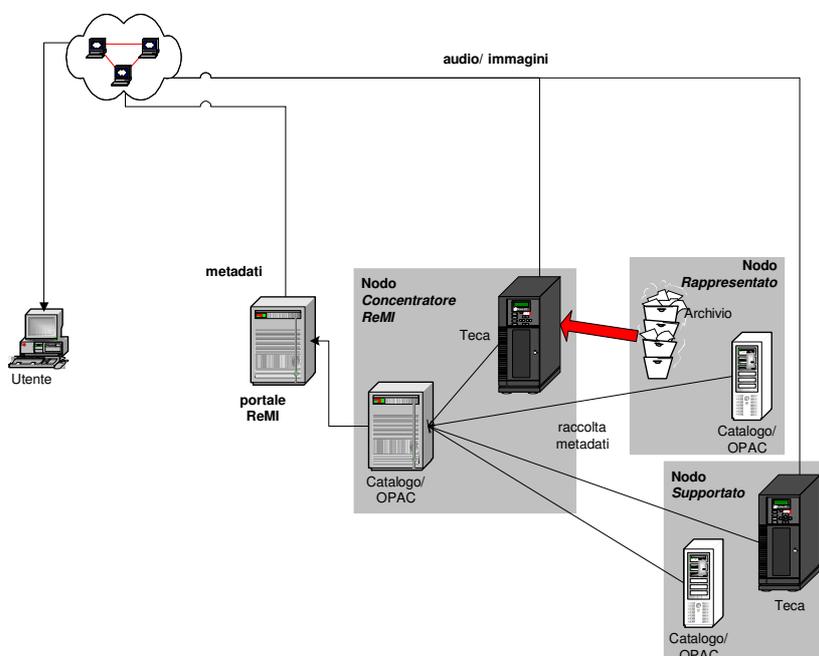
- NODO AUTONOMO:



3. Il Nodo autonomo

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

- **NODO SUPPORTATO:** E' un nodo dotato di teca digitale accessibile da remoto che implementa un tracciato del record catalografico o dei protocolli che non soddisfano pienamente gli standard ReMI, oppure appartiene ad una rete che offre servizi a valore aggiunto, come appunto la rete ADM; il Nodo supportato viene interrogato dal Portale tramite un Nodo concentratore;
- **NODO RAPPRESENTATO:** dispone solo di un catalogo; esporta quindi i propri record catalografici nell'OPAC del Nodo concentratore, e i propri oggetti digitali nella teca digitale del Nodo concentratore;
- **NODO CONCENTRATORE:** funge da intermediario rispetto al Portale ReMI, dispone di un catalogo, di un OPAC che può ospitare, oltre ai propri, record catalografici dei nodi supportati e rappresentati, e di una teca digitale accessibile da remoto che può ospitare, oltre ai propri, oggetti digitali dei nodi rappresentati.



4. il Nodo supportato, il Nodo rappresentato, il Nodo concentratore

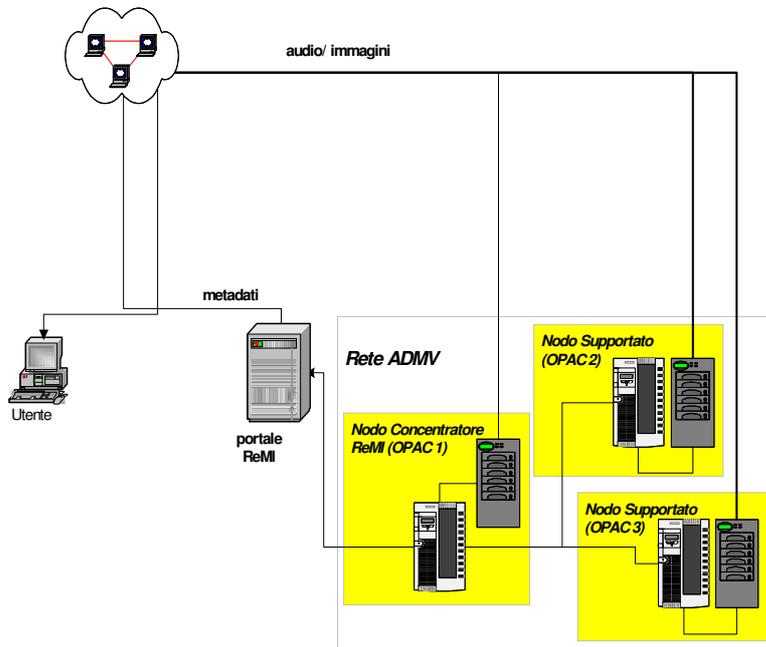
L'interoperabilità dei nodi e la raccolta dei metadati descrittivi sono assicurate dall'utilizzo dei protocolli SOAP⁴, OAI-PMH⁵ o z.39.50, SRW/U⁶, come supportati dai singoli server. In particolare l'interazione fra nodo concentratore e nodi supportati e rappresentati può essere realizzata tanto con export di record catalografici in formato standard, che con protocolli di *harvesting* dei metadati, come appunto OAI. L'OPAC indice SBN sarà egualmente accessibile tramite l'uso dei protocolli adatti.

⁴ Si veda: <http://www.w3.org/TR/2003/REC-soap12-part0-20030624/> ; <http://www.w3.org/TR/2003/REC-soap12-part1-20030624/> ; <http://www.w3.org/TR/2003/REC-soap12-part2-20030624/>

⁵ Si veda: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

⁶ Si veda: <http://lcweb.loc.gov/z3950/agency/zing/srw/>

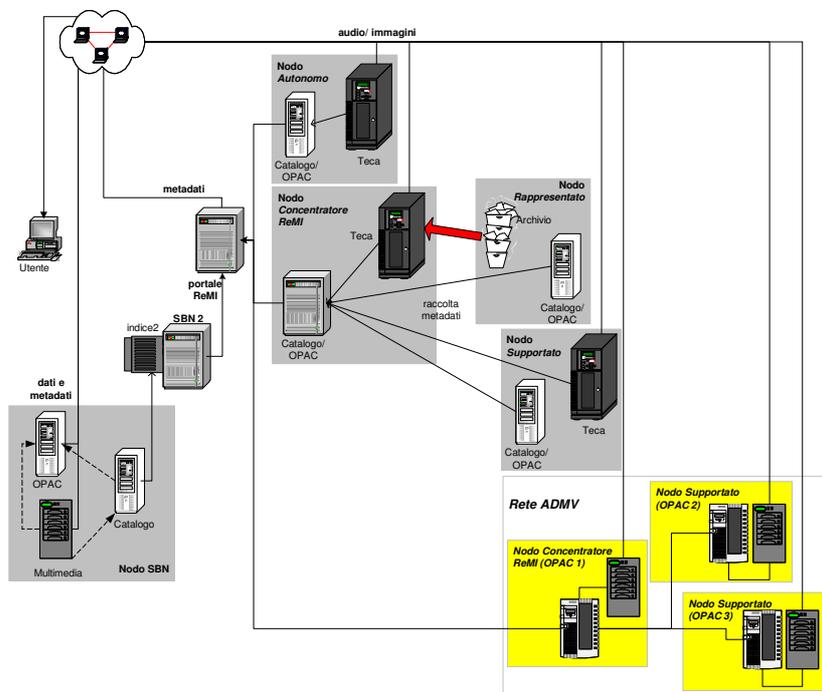
Nella figura 5 è illustrato il caso specifico del rapporto fra il Portale e l'attuale rete ADM. ADM si configura qui appunto come una rete a valore aggiunto, perché realizza già al suo interno un'interrogazione contemporanea di tutti i suoi nodi.



ADM

5. il Portale e la rete

Nella figura 6 è illustrato un grafico complessivo delle rete. Vi è inserita anche la rete SBN: il portale ReMI accede alla rete SBN tramite l'Indice SBN2, relativamente ai metadati descrittivi (record catalografici). La disponibilità degli oggetti digitali dipenderà dalle funzionalità che saranno realizzate all'interno della rete SBN medesima.



6. il Grafico della rete

2.4 I Metadati di ReMI

La differenziazione del livello di partecipazione alla rete è favorita anche dal formato dei metadati descrittivi previsto: ReMI definisce un tracciato record di riferimento, a cui i partner potranno allinearsi in maniera più o meno completa. L'obiettivo è infatti quello di consentire la partecipazione alla rete anche di istituzioni che non dispongono di archivi bibliografici fortemente strutturati, o non appartengono a reti bibliotecarie. Il formato prevede:

- **METADATI OBBLIGATORI:** devono essere presenti e gestiti da tutti i nodi della rete ReMI, corrispondono al Dublin Core Metadata Element Set⁷;
- **METADATI OPZIONALI:** possono o meno essere presenti nei nodi della rete; insieme ai metadati obbligatori costituiscono il tracciato record di riferimento ReMI, interrogabile dal portale, e che si assume uguale al tracciato record UNIMARC come definito da ADM e corrispondente al tracciato UNIMARC musicale gestito da Indice SBN2;
- **METADATI LOCALI:** si tratta di informazioni locali, che non rientrano nel tracciato di riferimento; non sono quindi condivise in ReMI, non sono interrogabili dal portale e, se recuperate insieme ai risultati di una ricerca, vengono prospettate come testo piano alla fine del record, come in un grande campo note.

Per quanto riguarda invece i **METADATI AMMINISTRATIVI E GESTIONALI** è interamente supportata, come in ADM, l'applicazione dello standard MAG.

2.5 Il modello delle teche digitali ReMI

⁷ Si veda: <http://dublincore.org/documents/dces/>

Di fondamentale importanza per la rete ReMI è poi il MODELLO DELLE TECHE DIGITALI che, come si è visto nelle figure precedenti, sono per lo più localizzate e gestite presso i singoli nodi della rete, secondo un modello distribuito. La teca ReMI deve essere innanzitutto orientata alla conservazione permanente degli oggetti digitali inseriti e deve garantire la loro accessibilità nel lungo periodo. Il tema attualissimo della *digital preservation* è stato oggetto di un importante convegno internazionale organizzato meno di un mese fa qui a Firenze dalla Direzione generale per i Beni Librari nell'ambito del semestre italiano di Presidenza dell'Unione Europea, anche a seguito di una Risoluzione del Consiglio dell'Unione Europea del 25 giugno 2002 finalizzata all'avvio di un Piano di azione europeo per la conservazione dei contenuti digitali. Dal convegno è scaturito un documento, la "Firenze Agenda", che definisce alcune azioni praticabili nell'arco di 12 mesi; fra queste la diffusione fra gli Stati Membri di "buone pratiche" basate su esperienze concrete e anche quantitativamente rilevanti. Nella stessa occasione è stato presentato un recentissimo documento UNESCO, la "Carta sulla conservazione dell'eredità digitale", che coniuga la necessità della conservazione permanente con quella dell'accesso, ed invita gli Stati Membri a definire il quadro giuridico ed istituzionale necessario al conseguimento di questi obiettivi. ReMI si pone all'interno di queste linee di lavoro, e intende costituire un terreno di sperimentazione di soluzioni tecniche, funzionali ed organizzative.

La teca digitale, che è sostanzialmente quella messa a punto in ADM, è innanzitutto conforme allo standard ISO 14721 del 2003, che ha definitivamente consolidato l'*Open Archival Information System (OAIS)*⁸, un modello logico di alto livello, cioè indipendente da specifiche applicazioni, per archivi di oggetti digitali destinati alla conservazione ed all'accessibilità permanenti.

Le componenti funzionali di una teca ReMI sono le seguenti:

- L'INTERFACCIA VERSO I FORNITORI DEI CONTENUTI DIGITALI E, DAL LATO OPPOSTO, VERSO GLI UTENTI DEI CONTENUTI DIGITALI

La teca riceve oggetti digitali e metadati MAG prodotti dalle attività di digitalizzazione; sia per gli oggetti digitali immagini che per gli oggetti digitali sonori sono previsti un formato ad alta risoluzione, non compresso, da utilizzare come copia master, e due formati compressi destinati l'uno ad un uso all'interno delle istituzioni (intranet), e l'altro ad un uso libero su rete geografica (internet). Nel caso in cui un'istituzione non disponga di un'applicazione OPAC separata, la teca è in grado di gestire e rendere accessibile un proprio OPAC, basato sui metadati descrittivi Dublin Core presenti nella sezione "bib" del MAG-Schema. Sono previste interfacce di amministrazione e interfacce per la fruizione sia locale che remota, attraverso il Portale, degli oggetti digitali.

- LO STRATO DELLA CONSERVAZIONE PERMANENTE

La teca implementa e supporta i requisiti funzionali previsti dal modello OAIS. Si tratta, come si diceva, di requisiti di ordine generale, che vanno dall'immissione degli oggetti digitali fino alla loro disseminazione, e coinvolgono sia le componenti hardware/software del sistema che le attività "umane" necessarie alla sua gestione, specie, ad esempio, per quanto riguarda la pianificazione e l'organizzazione delle specifiche azioni necessarie alla conservazione permanente, note come *policy* di conservazione. A questi fini, con particolare riferimento agli oggetti digitali sonori, è necessario che il MAG-Schema venga integrato al più presto con i metadati specifici per i file audio, per i quali si può prevedere una sezione apposita⁹. Qui vale la pena di ricordare che, a differenza dei documenti pergamenacei o cartacei che dureranno ancora a lungo, con la nota eccezione dei libri di carta acida pubblicati a partire dalla seconda metà dell'Ottocento, i dischi *shellacs* a 78 giri o i nastri in acetato sono già sempre più spesso irrecuperabili, e la digitalizzazione è l'unica alternativa praticabile, quando è ancora praticabile, per non perderne definitivamente il contenuto intellettuale. In questi casi la copia digitale non sarà una copia, resterà l'unica manifestazione di quell'opera, e andrà conservata come un originale.

⁸ Si veda la nota n. 3

⁹ Nel febbraio del 2004 la Discoteca di Stato ha presentato una proposta di integrazione del MAG-Schema che consente la gestione di file audio, video e degli spogli con record catalografico autonomo.

- LO STRATO DELL'ARCHIVIAZIONE FISICA

La teca ReMI prevede due distinte modalità per l'archiviazione fisica dei dati, a seconda delle dimensioni dell'archivio e della sua tipologia; in ambedue i casi si cerca di contemperare la migliore accessibilità degli archivi con le esigenze di sicurezza:

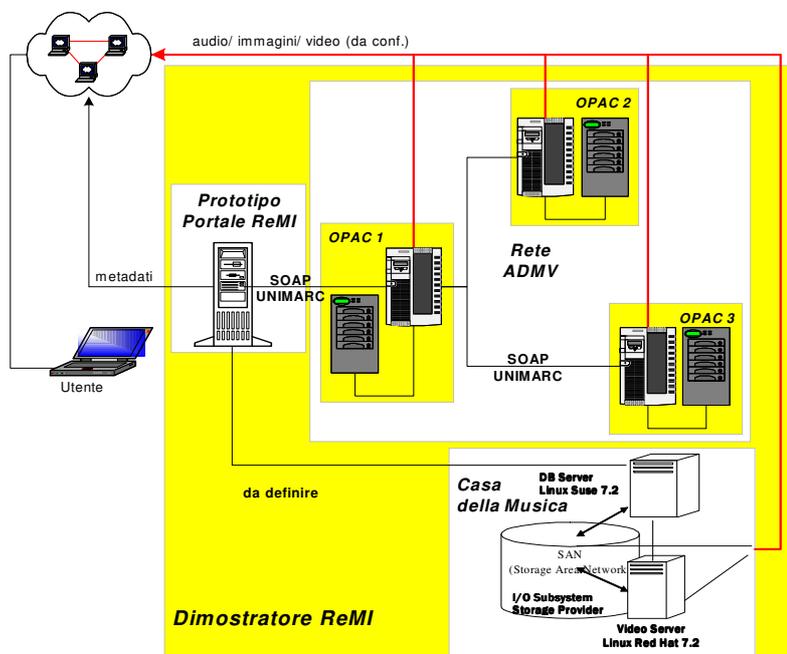
- l'Archivio *on-line*: tutti gli oggetti digitali (compresa la copia ad alta risoluzione) sono immagazzinati in un sistema server NAS collegato alla rete locale dell'istituto e quindi facilmente accessibile. Il rapporto capacità di memoria/prezzo di questi sistemi è sempre più vantaggioso (dai 5 ai 10.000 Euro per Tb), e la soluzione pare adatta soprattutto alla gestione di immagini digitalizzate, come quelle prodotte dalle biblioteche;

- l'Archivio *near-line*: si basa su una libreria robotizzata di nastri (tipicamente nastri LTO da 100 o 200 Gb ciascuno) che vengono fisicamente prelevati da un braccio meccanico e inviati al driver di lettura. Per dare un'idea delle dimensioni, una libreria da 100 nastri (20 Tb) trova posto sotto una scrivania. Anche in questo caso è previsto l'utilizzo di un sistema NAS con funzioni di *cache on-line*, in modo da rendere immediatamente accessibili gli oggetti più frequentemente richiesti. E' la soluzione preferibile per archivi di grandi dimensioni, e in particolare per gli archivi audiovisivi. Un sistema di questo tipo è attualmente utilizzato dalle Teche della RAI, ed è in via di implementazione presso la Discoteca di Stato, la cui esperienza sarà fondamentale per la rete ReMI.

- in ambedue i casi è comunque prevista la produzione di una copia di *back-up* dei dati, su nastri LTO, da conservare *off-line*.

2.6 Il Dimostratore di ReMI

Si prevede di realizzare una prima implementazione di ReMI, che è definita DIMOSTRATORE. Il Dimostratore realizzerebbe il Portale e la connessione alla rete ADM preesistente (BNM, BNT, Discoteca di Stato, altri eventuali partner che si saranno collegati nel frattempo), e ad un nodo autonomo, che è stato individuato nella Casa della Musica di Parma.



ReMI

7. il Dimostratore di