



Descrizione della piattaforma



Riferimenti

Riferimenti	
Oggetto	XDAMS
Categoria	Descrizione della piattaforma
Destinatario	
Referenti	
Nome File	XDAMS_2005_descrizione_01a.doc
Versione	2.1 – pagine: 28
Data di creazione	10 gennaio 2005
Ultimo aggiornamento	15 ottobre 2007
Autore	Donato Lanzellotto (dlanzellotto@regesta.com) – Ilaria Barbanti (ibarbanti@regesta.com)
Approvazione	Giovanni Bruno (gbruno@regesta.com)

XDAMS è un prodotto sviluppato da **regesta.exe** srl
(www.regesta.com)



X Dams rientra
nel progetto
Dams (Digital
Archives and Memory
Storage)
approvato e
sostenuto
nell'ambito
del programma
TEN-telecom dalla
Commissione
Europea

Extraway® Information Retrieval è un marchio registrato
della **3D Informatica** srl (www.3di.it)

Indice dei contenuti

Descrizione della piattaforma	1
Riferimenti.....	2
Indice dei contenuti.....	3
LA PIATTAFORMA XDAMS.....	4
Introduzione.....	4
XML.....	4
La descrizione archivistica.....	5
La descrizione e la gestione dei documenti multimediali.....	5
EAD (Encoded Archival Description).....	5
Gli elementi del modello dati EAD.....	6
Mapping ISAD(G) – EAD 2002.....	12
La soluzione XDAMS per gli archivi multimediali.....	13
Il materiale fotografico: la scheda F.....	15
Le norme FIAF per la descrizione dell'archivio audiovisivo.....	17
La gestione e l'accesso: MAG.....	17
LA SOLUZIONE TECNICA.....	19
Descrizione delle funzionalità.....	19
Architettura software.....	21
Software di base.....	24
Software applicativo.....	24
Infrastruttura.....	25
Architettura software di dettaglio.....	25
La piattaforma XML di riferimento: Extraway® Information Retrieval.....	27

La piattaforma XDAMS

Introduzione

La soluzione che proponiamo si basa sull'utilizzo di specifici moduli applicativi del sistema XDAMS.

- È un sistema integrato di gestione e accesso agli archivi.
- Si tratta di un applicativo interamente web based, che utilizza il protocollo TCP/IP per l'accesso ai dati e l'erogazione dei servizi.
- Utilizza XML come standard di comunicazione e di conservazione fisica dei dati e la piattaforma documentale Extraway™ per la creazione degli indici e l'accesso ai dati: Extraway™.
- Per la descrizione archivistica e la codifica delle risorse archivistiche adotta lo standard EAD2002 (Encoded Archival Description), pienamente compatibile con ISAD(G)

Ogni archivio ha la necessità di gestire al meglio il proprio patrimonio documentario e di renderlo accessibile alla ricerca e alla consultazione.

Gestire e accedere alle risorse documentarie mediante procedure automatizzate significa scegliere tra le metodologie più opportune nel rispetto dei contenuti. Se da un lato i metodi di descrizione dei dati d'archivio trovano sostegno in una consolidata tradizione, non sempre gli applicativi, a causa della loro diversità e della loro particolarità, rendono facilmente accessibili i dati.

XDAMS risolve i problemi di interoperabilità e condivisione delle risorse. La sua architettura è infatti sviluppata sulla piattaforma web e costruita in base alle direttive degli standard archivistici per la descrizione delle risorse documentarie in formato elettronico.

XDAMS rientra nel progetto DAMS (Digital Archives and Memory Storage) approvato e sostenuto nell'ambito del programma TEN-Telecom dalla Commissione Europea. (www.damsolutions.org).

XML

XML (eXtensible Markup Language) è un standard emanato dal W3C nel 1998¹. Si tratta di un linguaggio di codifica testuale derivato dallo standard SGML (Standard Generalized Markup Language – ISO 8879) e pensato per elaborare file di testo senza avere la necessità di dipendere, per la loro codifica, da un supporto e/o da una piattaforma applicativa locale. XML

- fornisce una sintassi per definire ed esprimere la struttura semantica e le relazioni tra le sue componenti (**DTD, Document Type Definition**)
- identifica le parti di un documento e le sue relazioni logiche, esplicitate in una **struttura di tipo gerarchico** così come definite nella DTD

L'utilizzo di questo tipo di tecnologia risulta vantaggiosa poiché

- la sintassi definita (DTD) costituisce una **struttura normalizzata** per la codifica di fonti appartenenti ad una stessa tipologia (nel nostro caso i documenti d'archivio)
- il formato testuale dei dati è **web compatibile**, così da renderli indipendenti da qualsiasi piattaforma hardware e software e da garantirne la ricerca, il recupero, la visualizzazione e la navigazione dei dati on line (interoperabilità per dati e formati)

¹ Tutti i materiali relativi allo standard, documenti ufficiali, informazioni e aggiornamenti, sono pubblicati sul sito del consorzio all'indirizzo <http://www.w3.org/XML>.

- consente di effettuare **ricerche relative al contenuto**. Attualmente i motori di ricerca sono sostanzialmente sistemi di ricerca *full-text*, che non tengono conto della struttura del documento, ma ricercano e restituiscono documenti interi grazie all'individuazione delle c.d. "meta – informazioni", parole chiave inserite arbitrariamente nel testo codificato in HTML dal produttore della pagina. In XML, le informazioni veicolate dai tag possono essere utilizzate da un sistema di *information retrieval* per individuare tutte le volte che un certa stringa compare al loro interno.

La descrizione archivistica

XDAMS consente di inserire, modificare, gestire esclusivamente in rete Intranet/Internet i dati descrittivi sulla documentazione conservata. Per archivi già schedati in sistemi automatizzati adottati in precedenza, sarà possibile importare all'interno di una piattaforma condivisa tutto il contenuto descrittivo.

Le caratteristiche di interoperabilità e condivisione delle risorse sono garantite dall'utilizzo della tecnologia XML.

Sulla scorta degli ultimi progressi nel campo della descrizione delle risorse archivistiche, **XDAMS** adotta lo standard **EAD** (*Encoded Archival Description*) una DTD/XML per la codifica delle risorse archivistiche, come verrà approfondito più avanti.

La descrizione e la gestione dei documenti multimediali

XDAMS estende il proprio servizio anche al trattamento digitale di materiale di diverso tipo: fotografie, oggetti d'arte, materiali audiovisivi, garantendo

- l'**acquisizione delle immagini** mediante la scelta dei macchinari più adatti al grado di definizione dell'immagine voluto e meno invasivi per gli oggetti fisici
- la **descrizione** secondo procedure standard sia nel caso di materiale multimediale considerato come insieme a sé, sia per la descrizione di risorse archivistiche a cui venga associato materiale originale digitalizzato
- le funzioni di **search and retrieval** nell'insieme degli archivi digitali acquisiti

EAD (Encoded Archival Description)

EAD è una DTD/XML standard per la codifica di strumenti di corredo archivistici tramite SGML. La versione 1.0, elaborata nel 1998 dalla SAA (*Society of American Archivists*) in collaborazione la *Library of Congress*, è stata recentemente aggiornata nella nuova EAD version 2002 (<http://www.loc.gov/ead/>).

Obiettivo di EAD è fornire le istituzioni archivistiche di strumenti di ricerca elettronici. Per raggiungere tale obiettivo, EAD, come standard, deve tenere necessariamente conto delle diversità esistenti tra le pratiche descrittive adottate da ogni comunità scientifica, e si propone quale formato di comunicazione che consente agli istituti archivistici di mettere a disposizione per utenti locali o remoti i propri strumenti di corredo, principalmente via Internet.

Le principali caratteristiche di EAD sono le seguenti²:

1. è un sistema per la gestione e la condivisione del patrimonio archivistico, e per la conservazione, accessibilità, utilizzo, conservazione fisica o trattamento del materiale
2. utilizza la tecnologia **SGML/XML** per la conservazione e la comunicazione fisica dei dati. In alcuni ambienti la descrizione archivistica può essere creata e mantenuta adottando

² Dal sito <http://sun3.lib.uci.edu/~blandis/eadwg/>, gennaio 2001.

tecnologie informatiche come database relazionali o *object-oriented*, e EAD verrà utilizzata principalmente quale meccanismo di importazione, anche per la **generazione e pubblicazione di inventari** a partire dalle schede descrittive presenti nel sistema. In altre situazioni, gli archivi possono gestire i dati descrittivi direttamente mediante sistemi basati su SGML/XML

3. La **tecnologia standard** di cui si serve (SGML/XML) rende lo standard indipendente da specifiche piattaforme hardware e software, per cui la continuità della struttura e del contenuto garantisce l'accettabilità e la validità nel tempo di ogni applicazione che lo utilizzi
4. permette il **recupero retrospettivo** di strumenti di ricerca archivistici (guide, inventari, ecc.) in formato cartaceo e la loro **pubblicazione** elettronica
5. I **nomi degli elementi** e degli **attributi** di EAD devono essere il più possibile universali dal punto di vista della lingua e dell'applicazione, al fine di favorire l'interscambiabilità e la portabilità. Nello stesso tempo, è importante approntare meccanismi capaci di accogliere il linguaggio specifico e i differenti output dei media utilizzati
6. EAD è aperta a proposte di modifica, da effettuare nel modo più trasparente possibile, e preferendo l'aggiunta piuttosto che la modifica di elementi, facendo in modo che le versioni successive della DTD EAD siano compatibili con quelle precedenti.

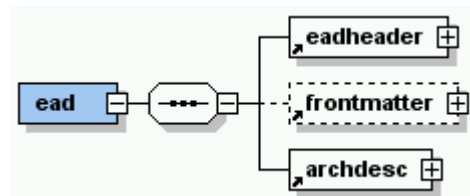
Sebbene **EAD** abbia cominciato il suo sviluppo negli USA basandosi sulle pratiche descrittive locali, le sue evoluzioni sono state osservate con attenzione dalla comunità archivistica internazionale e sono state costantemente sottoposte a test di funzionalità ed applicabilità a livello internazionale, particolarmente attraverso il confronto con lo standard per la descrizione dei dati **ISAD(G)**.

A differenza di quest'ultimo, EAD non è uno standard per il contenuto dei dati, né fornisce indicazioni su come esprimerli, ma sulla struttura in cui inserirli. In quanto tale risulta così quale completamento dello standard maggiormente condiviso a livello internazionale, nei confronti del quale è risultato perfettamente compatibile.

Gli elementi del modello dati EAD³

Gli elementi che compongono la DTD EAD sono circa 140, la maggior parte dei quali è opzionale. Tale caratteristica ha dato la possibilità di selezionare, nei limiti del rispetto delle più comuni norme per la descrizione archivistica, in particolare ISAD, quegli elementi dello standard ritenuti più idonei sia per effettuare il mapping dei dati già presenti nel formato di origine, sia per la schedatura ex novo.

La figura che segue riproduce l'albero del modello dati della DTD EAD (per un prospetto visivo degli ulteriori elementi tra quelli più significativi che compongono lo standard si veda la sezione **La visualizzazione del modello dati EAD**).

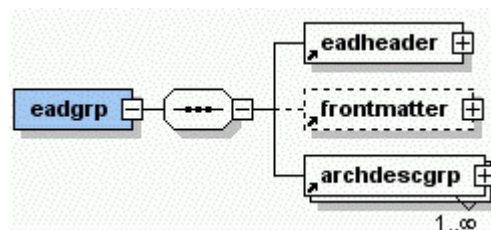


³ Per una trattazione esauriente di tutti gli elementi che compongono il modello dati si rimanda alle Tag Library, <http://www.loc.gov/ead/tglib/index.html>, e alle Application Guidelines relative alla DTD EAD v. 1.0, <http://www.loc.gov/ead/ag/aghomes.html>, entrambe redatte a cura della SAA e della LC.

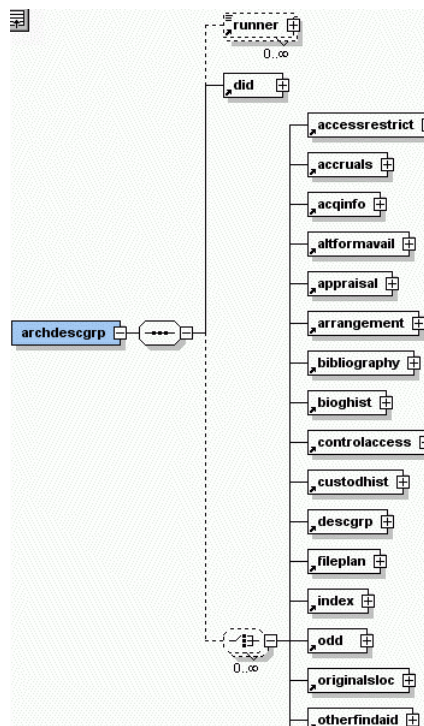
L'elemento contenitore più esterno, **ead**, non contiene direttamente informazioni ma altri elementi, che sono:

1. **eadheader** (obbligatorio): fornisce informazioni sul *finding aid* (titolo, compilatore, data di compilazione, etc.)
2. **frontmatter** : permette di strutturare gli elementi di un eventuale "frontespizio" personalizzato
3. **archdesc** (obbligatorio): contiene le informazioni e le descrizioni relative ad un corpus (collezione, fondo, serie, etc.) di materiale archivistico

La nuova versione della DTD permette anche di descrivere in un documento unico i dati descrittivi relativi a più archivi o fondi. In tal modo sarà possibile codificare mediante tale modello dati le guide, per loro natura strumenti di ricerca che raccolgono le informazioni descrittive relative a più archivi, unite per motivi diversi (appartenenza allo stesso territorio, tipologia, ecc.). In questo caso l'elemento contenitore è **eadgrp**, e presenta una struttura gerarchica di questo tipo:



A sua volta l'elemento <archdescgrp> apre ad una serie di elementi in cui sarà possibile fornire indicazioni generiche sullo strumento di ricerca o sull'insieme dei finding aids qui codificati, sino ad arrivare a descrivere specificamente ognuno:



Ad un livello alto dunque è possibile dare sul finding aid quelle informazioni che servono a descriverne la natura e la struttura generali, mentre sarà possibile indicare le informazioni specifiche relative a ciascun fondo scendendo al livello corrispondente.

Dei tre elementi principali quello che raccoglie i dati descrittivi è il terzo, **archdesc** (Archival Description). Esso costituisce l'elemento contenitore dell'intera informazione riguardante il contenuto, il contesto e la consistenza della documentazione archivistica, come pure molte delle informazioni supplementari che possono facilitarne l'accesso e l'utilizzo da parte degli utenti. In altre parole, tale sezione viene dedicata alla descrizione archivistica del complesso documentario per il quale è stato redatto il *finding aid*.

I due elementi <archdescgrp> e <archdesc> presentano un attributo obbligatorio level che permette di specificare a quale livello descrittivo ci si trova in quel punto. Il valore dell'attributo level deve essere scelto all'interno di un set di valori predefiniti previsti dallo standard⁴:

collection
fonds
class
recordgrp
series
subfonds
subgrp
subseries
file
item
otherlevel

⁴ Nel caso in cui non sia possibile scegliere tra alcuno dei valori previsti, è possibile indicare un valore "otherlevel" e l'effettivo livello gerarchico dell'unità di descrizione.

All'interno di <archdesc> si trova informazione organizzata gerarchicamente descrivente un *corpus* archivistico a livello più elevato di complesso documentario, per poi procedere, attraverso l'introduzione del sottoelemento Description of Subordinate Components <dsc>, a descrizioni sempre più dettagliate delle parti componenti.

Seguendo i principi della descrizione multilivello, gli elementi disponibili a livello di <archdesc> si ripetono inalterati ai livelli descrittivi inferiori, e l'informazione viene ad essere ereditata dal livello gerarchico più ampio a quello subordinato, conformemente anche a quanto indicato nelle ISAD(G).

Tra i sottoelementi di <archdesc> va notata la presenza, oltre che dell'elemento **runner**, opzionale, con informazioni sulle note che si trovano nel testo, dell'elemento obbligatorio Descriptive Identification <did>.

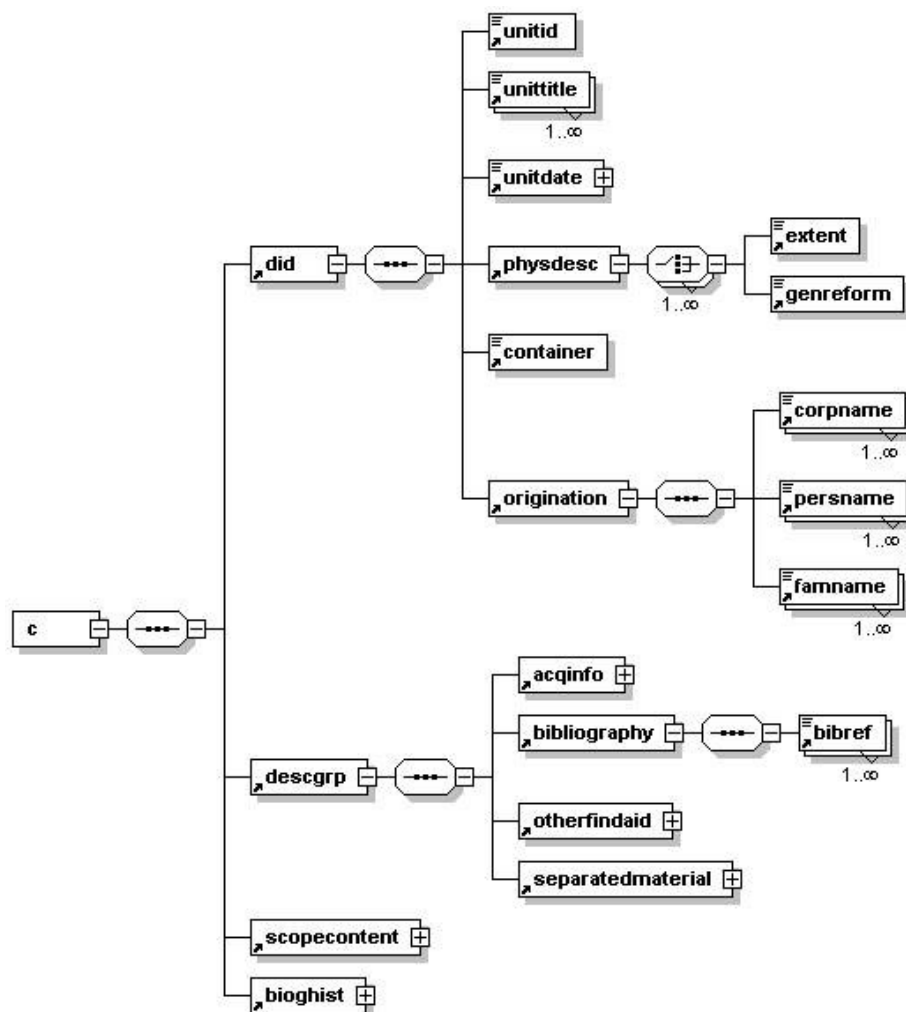
- **did** contiene elementi informativi della descrizione archivistica del complesso documentario essenziali per garantire l'efficace descrizione di un'unità archivistica a qualsiasi livello: mantiene infatti le stesse caratteristiche sia a livello **archdesc** che nei livelli inferiori. Può contenere i seguenti elementi:
 - **abstract** : sommario del materiale. Descrizione che si trova in forma più diffusa negli alti elementi di archdesc. Qui serve per indicare all'utente in maniera veloce ciò che è contenuto nel complesso.
 - **physdesc** (physical description) identifica la consistenza e tutte le informazioni sulle caratteristiche fisiche del materiale descritto. L'informazione può essere data come testo libero (PCDATA) o suddivisa negli elementi *dimension*, *extent* (= consistenza, quantità del materiale descritto), *genreform* (= genere e caratteristiche del formato fisico, stile e tecnica del contenuto, sia funzionale, sia intellettuale, *physfacet* (= dà informazione sull'aspetto esteriore/fisico del materiale es: colore, materia, sostanza, tecniche di creazione).
 - **note**, campo note
 - **repository** : istituzione o ente responsabile della conservazione e dell'accesso intellettuale
 - **origination** : informazioni sugli individui responsabili della creazione o dell'accumulo del materiale descritto prima che venisse depositato in un istituto di conservazione.
 - **unitdate** : estremi cronologici del complesso. Testuale o numerico. Data singola o range.
 - **unitid** : identificazione dell'unità di descrizione. Codifica ogni stringa di testo alfanumerico che identifica il materiale descritto. Attributi: *country code* e *repository code*
 - **unititle** : titolo dell'unità di descrizione
 - **container** : numero di unità di condizionamento (busta, scatola ecc.). Con attributo *type* per specificare il tipo di unità (busta, faldone ecc.). Fornisce informazioni sull'*extent* = consistenza del complesso documentario.
 - **physloc** (physical location) luogo della custodia/conservazione fisica del materiale descritto
 - **dao** e **daogrp** servono per i link a immagini digitali o altro

Una volta fornita attraverso <did> una descrizione basilare della documentazione, gli altri sottoelementi di <archdesc> hanno il compito di consentire la codifica di tutte quelle informazioni tipicamente presenti in un tradizionale strumento di ricerca archivistico, riguardanti il contenuto, il contesto, la consistenza, i criteri di ordinamento, e la gestione amministrativa in tutti i suoi aspetti, dall'acquisizione alle condizioni di accesso, del materiale descritto. Ecco di seguito quelli maggiormente significativi:

- **Accessrestrict** (restriction on access)/**userrestrict** (restriction on use): danno informazioni sulle condizioni di accesso e utilizzo del materiale descritto;
- **Acqinfo** : informazioni su donazioni, trasferimenti, acquisto e deposito, in generale sulle fonti da cui abbiamo avuto il materiale;
- **Altformavail** (Alternative Form Available): informazioni sulle copie dei documenti presenti su altri formati
- **Appraisal** : informazioni relative alla selezione del materiale e allo scarto, con descrizione di decisioni e metodologie;
- **Accruals** : informazioni sugli incrementi anche futuri del materiale;
- **Arrangement** : fascicolazione del materiale cioè se in ordine cronologico o alfabetico
- **Bibliography** , interessanti per interpretare il materiale, o basati sul stesso
- **Bioghist** : storia biografica. Può avere un titolo (biografia o storia amministrativa). Elemento ricorsivo. L'annidamento viene considerato come uno strumento per ottenere titolazioni per parti specifiche dell'elemento considerato, cosa che si rivela vantaggiosa per le funzioni di recupero dell'informazione.
- **Controlaccess** : elemento in cui vengono raggruppati tutti i nomi, soggetti, argomenti che illustrano al meglio la natura, l'ambito e la rilevanza della documentazione descritta. Chiavi d'accesso in cui trovano posto gli authority file.
- **Custodhist** : informazioni sulla catena dei passaggi di proprietà e di custodia della documentazione, significativi in termini di provenienza, integrità e interpretazione delle fonti;
- **Descgrp** : tale elemento raccoglie tutti gli elementi sopra indicati eccetto <dsc>, <dao> e <daogrp>. A differenza di <did>, e come gli altri elementi del suo stesso livello, non può prevedere direttamente testo, ma è vincolato dalla presenza di sottoelementi.
- **Fileplan** , schemi di classificazione per il recupero del materiale (titolario)
- **Index** , liste di parole chiave. Trova nomi già presenti (e codificati) nel testo, e prevede anche che ci sia un puntamento per i termini non descritti o non fisicamente presenti nell'inventario ma presenti nelle carte
- **Otherfindaid** , riferimenti ad altri strumenti di ricerca (elenchi di versamento e di consistenza).
- **Prefercite** : indica la forma normalizzata di citazione;
- **Processinfo** : informazioni sull'accesso, l'ordinamento, la descrizione, la conservazione, la disposizione, o ogni altra operazione che rende disponibile all'utente il materiale. Ha come sottoelementi address, blockquote, chronlist, head, list, note, p, table.
- **Odd** (Other Descriptive Data): fornisce informazioni sui materiali, contenute nei *legacy data* (= strumenti di ricerca precedenti al modello EAD), che non trovano posto negli elementi identificati in <archdesc> o nei livelli inferiori, o ancora permette di aggiungere informazione di varia natura, che risulta difficile scorporare nei vari elementi. In genere è preferibile ricorrere il meno possibile a tale elemento: le Application Guidelines raccomandano infatti di gestire i dati contenuti nei *legacy data* anche spostando le informazioni nei campi appropriati.
- **Organization** : struttura della sedimentazione della carte o schema delle serie (livelli inferiori)
- **Scopecontent** (ambito e contenuto). Corrisponde alla nota archivistica, e viene usato, ad esempio, quando la documentazione viene versata in momenti successivi. Ha come sottoelementi *arrangement* e *organization*.
- **Related materials** , descrive materiali ritenuti utili per comprendere quello descritto, fisicamente e logicamente separati, non facenti parte del corpus documentario (altri enti produttori)
- **Separated materials** materiali associati per provenienza dallo stesso ente produttore ma conservati altrove o separatamente

- **Dsc** (Description of Subordinate Components), una volta terminata la descrizione generale del *finding aid* sul complesso documentario, si giunge alla descrizione dei raggruppamenti gerarchici veri e propri del materiale descritto.

I livelli descrittivi di **dsc** si trovano nell'elemento **C** (Componente). Esso presenta la stessa potenzialità informativa disponibile in **archdesc**, ma in questo caso si accede alla vera e propria descrizione nel rispetto della gerarchia in cui si articola la documentazione, dunque dalle serie si arriva all'unità archivistica di base, o all'unità documentaria (item).



Come si può notare dalla figura precedente, <c> contiene a sua volta did, di cui naturalmente alcuni elementi non saranno presi in considerazione, poiché già descritti a monte nella DID di archdesc. Questo è il principio di ereditarietà, secondo il quale tutto ciò che viene inserito a livello superiore, viene poi ereditato nei livelli gerarchicamente inferiori.

Mapping ISAD(G) – EAD 2002

Uno schema di mapping ISAD(G) – EAD v. 2002 è disponibile anche al sito
http://www.loc.gov/ead/tglib/appendix_a.html

ISAD(G)	EAD 2002
---------	----------

AREA DELL'IDENTIFICAZIONE	
3.1.1 Segnatura/e o codice/i identificativi	<unitid> con gli attributi COUNTRYCODE e REPOSITORYCODE
3.1.2 Denominazione o Titolo	<unittitle>
3.1.3 Data/e estreme	<unitdate>
3.1.4 Livello di descrizione	<archdesc> e <c> con l' attributo LEVEL
3.1.5 Consistenza e supporto dell'unità di descrizione (quantità, volume, dimensione fisica)	<physdesc>, e i sottoelementi <extent>, <dimension>, <genreform>, <physfacet>

AREA DELLE INFORMAZIONI SUL CONTESTO	
3.2.1 Denominazione del/dei soggetto/i produttore/i	<origination>
3.2.2 Storia istituzionale/amministrativa, nota biografica	<bioghist>
3.2.3 Storia della custodia (passaggi di responsabilità giuridica)	<custodhist>
3.2.4 Modalità di acquisizione o versamento	<acqinfo>

AREA DELLE INFORMAZIONI RELATIVE AL CONTESTO E ALLA STRUTTURA	
3.3.1 Ambiti e contenuto	<scopecontent>
3.3.2 Procedure, tempi e criteri di valutazione e scarto	<appraisal>
3.3.3 Incrementi previsti	<accruals>
3.3.4 Criteri di ordinamento	<arrangement>

AREA DELLE INFORMAZIONI RELATIVE ALLE CONDIZIONI DI ACCESSO E UTILIZZAZIONE	
3.4.1 Condizioni che regolano l'accesso	<accessrestrict>
3.4.2 Condizioni di riproduzione e copyright	<userrestrict>
3.4.3 Lingua/scrittura della documentazione	<langmaterial>
3.4.4 Caratteristiche materiali e requisiti tecnici	<phystech>
3.4.5 Altri strumenti di ricerca	<otherfindaid>

AREA DELLE INFORMAZIONI RELATIVE A DOCUMENTAZIONE COLLEGATA	
3.5.1 Esistenza e localizzazione degli originali	<originalsloc>
3.5.2 Esistenza e localizzazione di copie	<altformavail>
3.5.3 Unità di descrizione collegate; materiale documentario complementare	<relatedmaterial> <separatedmaterial>
3.5.4 Bibliografia	<bibliography>

AREA DELLE NOTE	
3.6.1 Note	<odd>

AREA DI CONTROLLO DELLA DESCRIZIONE	
3.7.1 Nota sulle fonti consultate per la descrizione e sulla redazione dell'inventario	<processinfo>
3.7.2 Norme o convenzioni	<descrules>
3.7.3 Data/e di redazione/modifica dell'inventario	<processinfo><p><date>

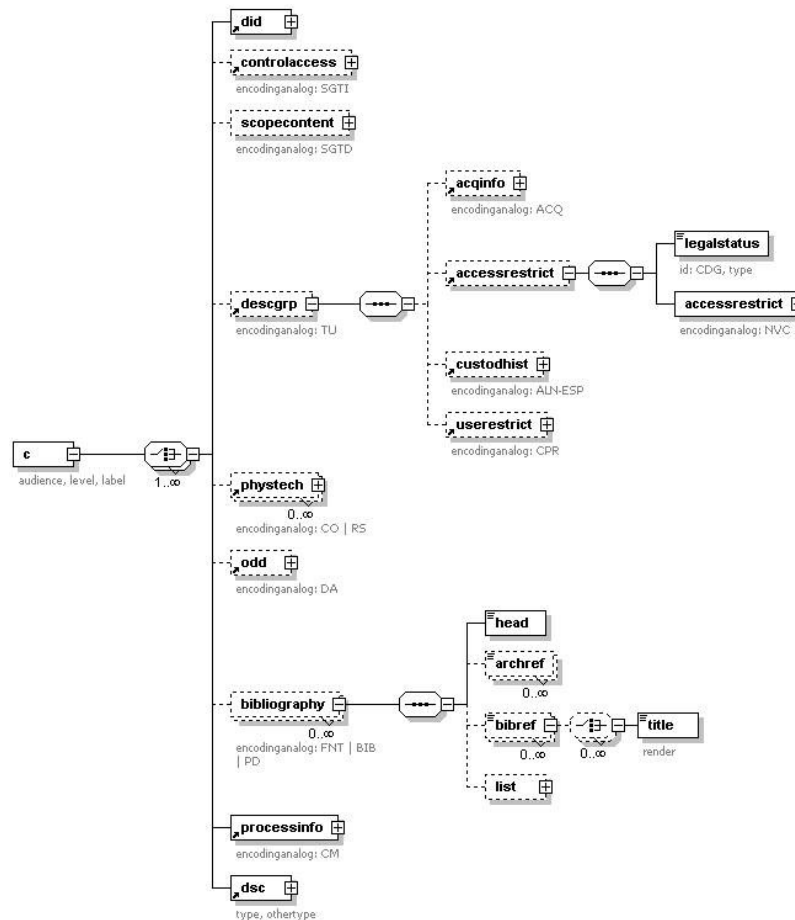
La soluzione XDAMS per gli archivi multimediali

La descrizione dei dati avverrà mediante la definizione di uno specifico schema descrittivo, che tenga conto degli standard descrittivi specifici implementati sulla piattaforma XDAMS; in

particolare, si fa qui riferimento al diverso grado di analiticità della descrizione del materiale fotografico che viene seguita per il trattamento dei diversi fondi che compongono l'archivio fotografico dell'Istituto.

Particolare attenzione sarà prestata in vista dell'accessibilità via Internet di queste banche dati, in modo da offrire agli utenti risposte non equivocate e immediatamente comparabili, indipendente dalla grana descrittiva utilizzata.

Per quanto attiene agli standard descrittivi adottati, la piattaforma XDAMS utilizza EAD (Encoded Archival Description) come infrastruttura descrittiva di base e strumento per la descrizione multilivello (principio di provenienza), integrando per la parte informativa specifica dell'immagine fotografica parte degli standard emanati dall'ICCD e per la parte informativa specifica agli audiovisivi un sottoinsieme delle normative FIAF.



Il materiale fotografico: la scheda F

La fotografia può essere considerata:

0. un documento storico
1. un'opera d'arte
2. come documentazione di corredo

La catalogazione delle fotografia riveste notevole importanza per chiunque si trovi ad utilizzarla: storici, architetti, etnografi, archeologi, ecc. Per tali ragioni, l'ICCU ha approntato uno standard descrittivo assai ricco (contiene infatti 347 campi), partendo dal presupposto che la fotografia ha le seguenti caratteristiche:

- fornisce informazioni direttamente collegate al reale "hic at nunc"
- presenta un contenuto, ossia le immagini che riproduce, che va preservato e descritto in maniera univoca
- documenta aspetti storici
- presenta il modo di vedere del fotografo

Tali aspetti fanno riflettere, in definitiva, sull'importanza che assume la capacità di descrivere il contenuto. Molti sono i campi del tracciato descrittivo della scheda F nei quali è possibile indicare notizie descrittive fino ad un livello di specificità molto elevato.

Il tracciato della scheda F privilegia la foto in quanto oggetto. In ogni caso il catalogatore potrà servirsi di un ampio spettro di possibilità descrittive, così da facilitare la ricerca ed il recupero delle informazioni storiche, artistiche, sociologiche, ecc., e da rendere customizzabile a seconda delle esigenze specifiche dell'istituzione responsabile dell'accesso ai documenti audiovisivi.

Il modello dati della scheda F permette di descrivere il contenitore in cui sono conservate e raggruppate le fotografie, nonché i singoli oggetti.

La scheda "scatola" contiene i seguenti elementi descrittivi:

1. Numerazione scatole possibilmente cronologica)
2. Supporto
3. Formato
4. Emulsione
5. Colore / bianco e nero
6. Marca
7. Note (consistenza, collocazione, ...)

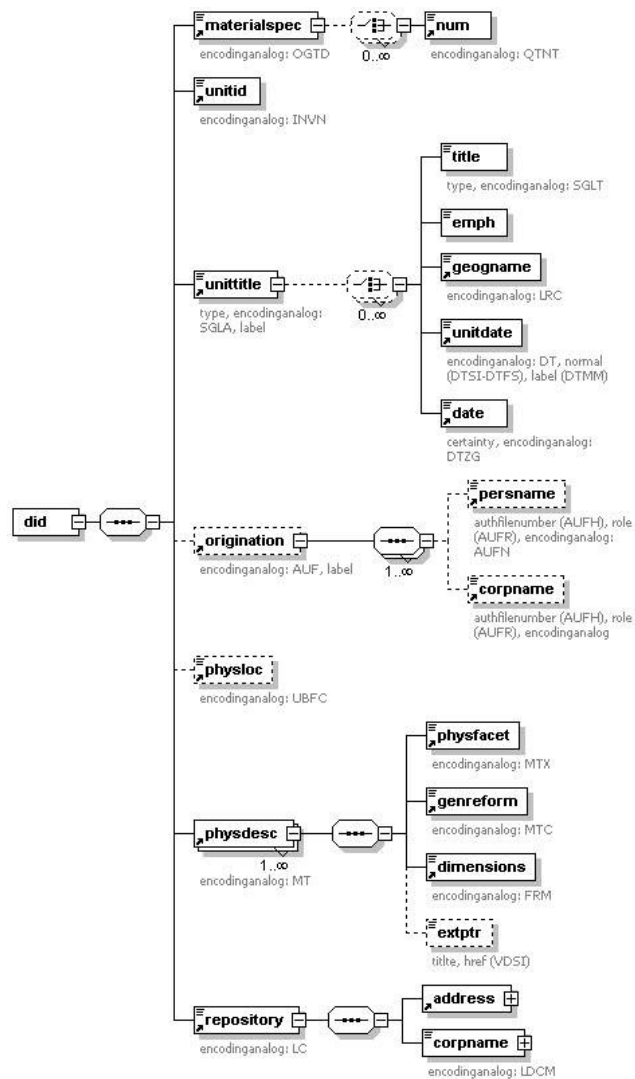
La scheda della singola fotografia presenta invece i seguenti elementi descrittivi:

1. Originale o immagine digitalizzata
2. Contenuto (indicato sulla scatola)
3. Luogo della ripresa
4. Periodizzazione
5. Occasione
6. Identificazione: descrizione del soggetto
7. Indicazioni sul soggetto
8. Icon. Class.
9. Autori fotografia
10. Altri autori (nel caso di fotografia di opera d'arte complessa)
11. Generi fotografici: codificazione sperimentale
12. Da attribuire
13. Fiction: quando c'è un set (fotoromanzo)
14. Foto artistica
15. Soggetti.

16. Scheda supporto sull'oggetto fisico: tipologie dei materiali, dimensioni, sezione di provenienza
17. Numero busta di provenienza
18. Titolo attribuito (non dell'autore)
19. Dati analitici (indicazioni sul retro)

Per ogni oggetto multimediale dovranno essere almeno censite le seguenti informazioni:

- le informazioni anagrafiche: titolo e descrizione, autore, data e luogo di esecuzione (attività di schedatura);
- le informazioni tecniche: supporti, formato, cromatismo, time code etc..
- le informazioni di contenuto: descrizione particolareggiata dell'oggetto e dei componenti significativi;
- le informazioni del contesto di appartenenza
- le informazioni relative alle voci, persone luoghi, eventi, ecc., da inserire nell'apposita banca dati di Authority, che sarà implementata dai catalogatori sulla base della documentazione informativa disponibile.



Le norme FIAF per la descrizione dell'archivio audiovisivo

Le norme FIAF (Fédération internationale des Archives du Film)⁵ sono un insieme di regole per la catalogazione degli archivi composti da materiale audiovisivo, allo scopo di facilitare lo scambio di informazione tra archivi della medesima tipologia. I lavori per la definizione del modello dati descrittivo cominciarono alla fine degli anni '60, per arrivare alla pubblicazione de The FIAF Cataloguing Rules For Film Archives, nel 1991. Poco prima, nel 1987, erano state pubblicate le norme ISBD(NBM) (International Standard for Bibliographic Description of non-Book Material). Da tale standard le norme FIAF hanno mutuato molti degli elementi descrittivi generali, per poi approfondire la descrizione in altre direzioni.

Per tali ragioni, le norme FIAF contemplano il principio della separazione tra informazioni univoche, relative agli elementi anagrafici del singolo elemento audiovisivo, e quelle relative alla descrizione fisica e alle informazioni tecniche di ogni copia del film presente in archivio, su differenti supporti.

Il modello descrittivo è infatti composto dalle seguenti aree informative:

- 0. NOTE PRELIMINARI
 - 0.1. AMBITO
 - 0.2. PROPOSITO
 - 0.3. UTILIZZO
- 1. TITOLO E DICHIARAZIONE DI RESPONSABILITÀ
- 2. EDIZIONE/VERSIONE/VARIANTI
- 3. PRODUZIONE, DISTRIBUZIONE, ECC.
- 4. DICHIARAZIONE DI COPYRIGHT
- 5. DESCRIZIONE FISICA
- 6. AREA DELLE SERIE
- 7. AREA DELLE NOTE

Mancando uno standard descrittivo normativo, sull'esempio della scheda F-ICCD per questa tipologia di materiale, si è inteso elaborare una specifica DTD di riferimento, mutuata sulla base dello Schema EAD e coerente con quella adottata per il fotografico.

La gestione e l'accesso: MAG

Il MAG si presenta come un modello per l'individuazione e la creazione di un set di metadati descrittivi, strutturali e amministrativo – gestionali sulle risorse o oggetti digitali siano essi quelli derivati da documenti preesistenti su supporti non digitali oppure gli oggetti digitali nativi (i prodotti dell'editoria elettronica).

Il gruppo di lavoro è partito dalle esperienze sviluppatesi in ambito nazionale (progetto ARSBNI, *Arricchimento dei servizi della Bibliografia Nazionale Italiana*) e internazionale (progetto OAIS - **Sistema Informativo Aperto per l'Archiviazione**, Dublin Core metadata iniziative, ecc.) per cercare di cogliere le caratteristiche che maggiormente potessero essere recepite nell'ambito di progetti di creazione di Biblioteche digitali.

La scelta è caduta nell'adozione di uno standard internazionale che prescindendo da qualsiasi soluzione sw proprietaria, XML. Il modello adottato è lo Schema, evoluzione della DTD XML, utile per definire non solo il modello strutturale dell'informazione veicolata, ma anche il contenuto.

⁵ Le norme FIAF per la catalogazione degli audiovisivi sono disponibili per il download al sito web <http://www.fiafnet.org/fr/publications/catrules.cfm>.

Come accennato nella presentazione dei risultati del gruppo di lavoro, lo Schema XML all'interno di un progetto di digitalizzazione ogni postazione di lavoro è in grado di produrre per ogni oggetto digitalizzato un file guida "standard" ovvero conforme allo Schema XML che:

- raccoglie tutte le informazioni sull'oggetto digitalizzato (metadati);
- contiene la mappa di tutti i file generati contestualmente dalla digitalizzazione e relativi all'oggetto digitalizzato.

Scopo dello Schema è quello di dare delle specifiche formali standard per la fase di raccolta e di archiviazione dei metadati e dei dati digitali.

Lo Schema è composto da un elemento wrapper, "metadigit", composto dalle seguenti 4 sezioni⁶:

- **gen**: informazioni generali sul progetto e sul tipo di digitalizzazione
- **stru**: metadati sulla struttura del documento, composto, ad esempio, dal frontespizio, e dal contenuto delle pagine testuali
- **bib**: metadati descrittivi sull'oggetto digitalizzato
- **img**: metadati specifici relativi alle immagini fisse. Per queste si fa riferimento alla norma NISO (National Information Standards Organization) per i metadati sulle immagini statiche. In particolare
- **ocr**: metadati specifici relativi al riconoscimento ottico del testo
- **doc**: metadati specifici relativi ai documenti testuali

Ogni formato di metadati utilizzato è associato ad un [Namespace](#) che fissa in modo non ambiguo la terminologia e ad un [XMLSchema](#) che ne fissa la struttura sintattica.

⁶ Lo schema principale di riferimento si trova su <http://www.bncf.firenze.sbn.it/progetti/index.html>, cliccando su "Schema di dettaglio" nella sezione "Documentazione".

La soluzione tecnica

Descrizione delle funzionalità

La soluzione che abbiamo realizzato si basa sull'utilizzo di specifici moduli applicativi del sistema XDAMS.

1. È un sistema integrato di gestione e accesso agli archivi.
2. Si tratta di un applicativo interamente web based, che utilizza il protocollo TCP/IP per l'accesso ai dati e l'erogazione dei servizi.
3. Utilizza XML come standard di comunicazione e di conservazione fisica dei dati e la piattaforma documentale Extraway™ per la creazione degli indici e l'accesso ai dati: Extraway™ rappresenta l'evoluzione del software Highway™, di cui mantiene e sviluppa tutte le avanzate funzionalità di ricerca e di gestione gerarchica delle schede.
4. Per la descrizione archivistica e la codifica delle risorse archivistiche adotta lo standard EAD2002 (Encoded Archival Description), pienamente compatibile con ISAD(G)

Le principali caratteristiche della piattaforma sono:

1. È un sistema multiutente, in grado di offrire l'accesso differenziato all'utenza, sia essa deputata alla gestione o solamente alla consultazione dei dati.
2. È un sistema multiarchivio, che consente di raccogliere e interrogare dati archivistici di differente provenienza e formato.
3. È un sistema multimediale, che oltre a offrire la descrizione di documenti d'archivio, consente di associare a essi, o di ordinare autonomamente, altre tipologie di documenti, come immagini fotografiche, documenti audiovisivi, file musicali, ecc.

L'architettura del sistema informativo è strutturata per ogni singolo account in due sottoinsiemi descrittivi:

1. una o più banche dati per la descrizione archivistica: ogni fondo archivistico inserito nel sistema risiede su un'autonoma banca dati XML, immagazzinata in un'apposita directory del file server;
2. un Authority file comune a tutte le banche dati esistenti, che svolge una tripla funzione nel sistema:
 - ✓ controlla e normalizza l'inserimento nelle banche dati archivistiche dei valori in una serie di campi qualificati (*access point*);
 - ✓ fornisce le informazioni di carattere generale e di contesto (informazioni identificative, anagrafiche, biografiche, bibliografiche, ecc.) per specifici elementi descrittivi: persone, famiglie, *corporate bodies*;
 - ✓ consente un accesso qualificato a tutti i records descrittivi che contengono quei valori su tutto il sistema informativo.

L'ancoraggio ad un unico authority centrale e la condivisione della struttura informativa di descrizione del contenuto consentono una immediata integrazione di archivi eterogenei con la disponibilità di indici comuni e di strumenti di ricerca condivisi.

Le modalità di utilizzazione del sistema:

1. modalità di gestione (management mode)
Permette di definire e controllare la lista e le proprietà degli archivi, la struttura, gli utenti e i loro privilegi, i livelli di accesso alle informazioni, le relazioni e i workflow operativi;
2. modalità operativa (staff mode)

Consente di accedere agli archivi con privilegi di inserimento e gestione dei dati, oltreché di aggiungere descrizioni e allegati digitali;

3. modalità di navigazione (user mode)

Le tipologie di utenti:

1. *Account Manager*: si tratta dell'amministratore del singolo account ed ha dunque facoltà di:
 - a. creare e aggiungere utenti e archivi
 - b. gestire i criteri di accesso per gli utenti registrati indicando a quali archivi e con quali privilegi di accesso per ogni archivio ogni utente possiede. I privilegi di accesso potranno essere modificati dall'Account Manager in qualsiasi momento.
2. *Archive Manager*: tale utente ha facoltà di amministrare l'organizzazione di uno o più archivi presenti sul proprio sistema attraverso una serie di tools specifici per:
 - a. riordinare e riorganizzare l'archivio, rinumerare le schede descrittive, visualizzare i permessi, ecc.;
 - b. stampare il singolo record o l'inventario di una sezione specifica o dell'intero archivio per l'eventuale pubblicazione;
 - c. esportare in XML l'intero archivio o parte di esso.
3. *User*: sotto tale tipologia di utente rientrano tutti gli utenti che hanno il compito di aggiungere, modificare o eliminare le schede delle unità di descrizione..
4. *Guest*: qualsiasi utente esterno all'istituzione che voglia accedere alla documentazione d'archivio potrà utilizzare tutte le funzionalità di ricerca a sua disposizione ma non potrà modificare il contenuto delle schede descrittive; l'utente guest può agire solo sull'insieme di archivi resi disponibili alla consultazione esterna dall'Account manager; inoltre, all'interno delle banche dati pubbliche la disponibilità delle singole descrizioni archivistiche e di ogni allegato digitale presente sul sistema è subordinata a specifici criteri di visibilità (per la gestione di documentazione riservata o ad uso limitato).

L'accesso alle banche dati avviene attraverso quattro modalità:

1. la ricerca libera su tutte le schede descrittive;
2. la consultazione di liste controllate di parole chiave (persone, enti, luoghi) e dei dizionari di campo;
3. la consultazione degli authority files e la successiva proiezione sulle schede descrittive esistenti su uno qualsiasi delle banche dati archivistiche esistenti
4. la navigazione della struttura gerarchica di un singolo fondo archivistico; selezionando sulla struttura il singolo titolo dell'unità di descrizione sarà possibile visualizzare:
 - la scheda descrittiva breve, con gli elementi identificativi essenziali;
 - la scheda descrittiva completa di tutte le informazioni inserite dallo schedatore;
 - gli allegati digitali, se presenti.

È inoltre possibile, a seconda del tipo di accesso assegnato dal gestore, procedere nella stampa e nel download della documentazione digitalizzata.

Ad ogni scheda descrittiva può essere allegato uno o più files digitali di diverso formato: files immagine, audiovisivi, musicali, documenti, o qualsiasi altro file digitale; tutti gli allegati vengono salvati entro una struttura di directory del file server, creata automaticamente dal sistema. Nel set di procedure di XDAMS è integrato un applicativo per l'upload di allegati digitali singoli (un solo file) o multipli. Per ogni singolo file allegato è possibile:

- specificare un titolo specifico o di gruppo
- indicare uno specifico criterio di visibilità (per default si assume quello della scheda).
-
- Inserimento di schede negli authority files:
- può essere effettuato sia dalla banca dati dell'archivio che da quella dell'authority file specifico;

- da qualsiasi punto venga effettuato, l'inserimento interviene (in inserimento, modifica ed eliminazione) su di un'unica banca dati, dando così modo di verificare in tempo reale l'univocità di ogni singola authority entry.

Funzionalità dell'archive manager:

- riordinamento degli archivi (spostamento di singoli rami o di selezioni di schede, rinumerazione di tutto o parte della banca dati, eliminazione e inserimenti di livelli descrittivi)
- produzione di strumenti inventariali a stampa:
 - è possibile stampare una selezione di schede o uno specifico ramo d'inventario,
 - definire dinamicamente la lista dei campi da stampare e il loro layout grafico,
 - indicare se accodare alla stampa delle schede quella degli indici dei controlaccess della selezione,
 - salvare il documento in formato pdf.

Architettura software

L'architettura si basa su una struttura multi-tier che implementa i SERVIZI (funzionalità rese disponibili agli utenti) utilizzando una modellizzazione a componenti (*"Component Software Modelling"*).

Più in particolare la tecnologia di riferimento sarà JAVA– XML ed il modello di riferimento per lo sviluppo dell'applicazione è quello previsto dal Model View Controller.

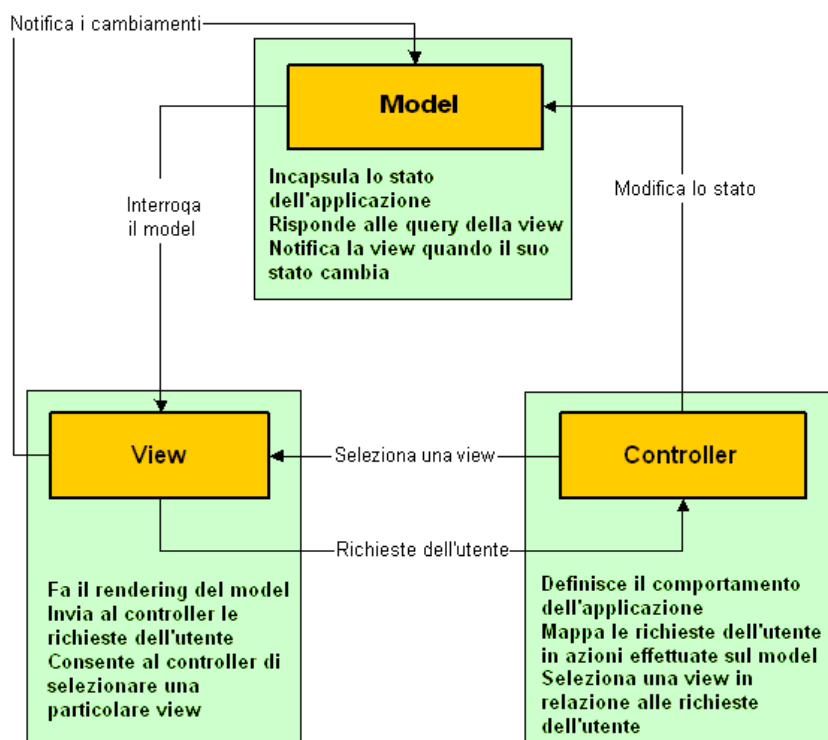
Il modello MVC si basa nella suddivisione funzionale degli oggetti dell'applicazione, in modo da disaccoppiarli fra loro il più possibile: Nell'architettura prevista quindi ci sono oggetti che hanno a che fare con gli aspetti legati alla sua presentazione (view), altri che riguardano le regole legate alla business logic ed ai dati (model), altri ancora che accettano ed interpretano le richieste degli utenti e sono responsabili di controllare che tali richieste siano più o meno legittime per poi esaudirle se è possibile (controller). Il model rappresenta quindi i dati dell'applicazione e le regole che governano le operazioni con cui tali dati vengono acceduti e modificati. Spesso esso rappresenta un'approssimazione software degli oggetti realmente presenti nel dominio dell'applicazione.

Il model notifica la view dei suoi cambiamenti e mette a disposizione della view un modo per interrogare il model circa il proprio stato.

La view ha invece il compito di effettuare il rendering del model. In altri termini, questa accede ai dati contenuti nel model e decide le modalità con cui questi dati debbano essere presentati. Quando il model cambia è responsabilità della view di mantenere la sua presentazione coerente con tali cambiamenti.

La view, infine, ha il compito di inviare le richieste dell'utente al controller.

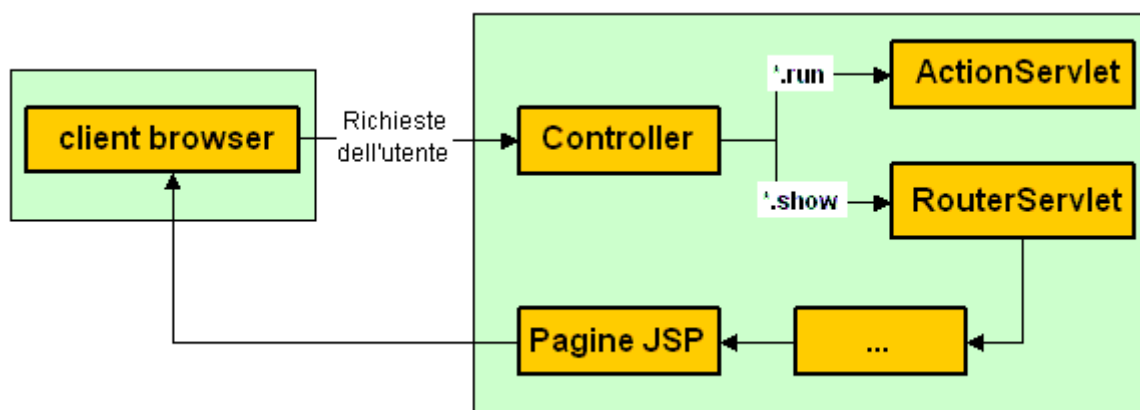
Quest'ultimo definisce il comportamento dell'applicazione: esso interpreta le richieste dell'utente e le mappa in azioni che verranno eseguite sul model. In una web application di questo tipo, tali richieste vengono effettuate al web tier sotto forma di request get e post HTTP. Inoltre, in base alle richieste dell'utente ed agli effetti che tali richieste implicano sul model, il controller seleziona la view che verrà utilizzata per effettuare il rendering del model così modificato.



Come nella maggior parte delle applicazioni EJB-centriche, anche in questo caso, gli EJB di tipo entity, che riflettono i dati immagazzinati nella base di dati, rappresentano il model dell'architettura. Per quanto riguarda la view, i dati memorizzati nel model vengono replicati lato web da componenti JavaBeans. Questi componenti si registrano presso il controller per fare il listening degli eventi di aggiornamento del model: quando ricevono uno di tali eventi interrogano gli entity bean per allineare il proprio stato. Il rendering dei dati presenti nei componenti JavaBeans viene poi ovviamente effettuato mediante pagine JSP. Infine, per ciò che concerne il controller, esso è probabilmente la parte più complessa dell'applicazione. Mentre infatti la view è concentrata sul web-tier e il model risiede unicamente sull'ejb-tier, il controller necessita di estendersi su entrambi i livelli; proprio a causa di tale complessità può essere funzionalmente suddiviso nelle seguenti sottoparti: request processor: converte le request HTTP in eventi comprensibili dal resto dell'applicazione, consentendo di concentrare in un unico punto le processazioni specifiche del protocollo HTTP e quindi rendendo il resto dell'applicazione indipendente dal tipo di client utilizzato; web controller: inoltra gli eventi generati dal request processor all'EJB controller, assicurando che il risultato dell'aggiornamento del model, effettuato da quest'ultimo, venga propagato ai componenti JavaBeans ovvero alla view; ejb controller: accetta gli eventi inviati dal web controller e modifica il model coerentemente con tali eventi; è anche responsabile di mantenere lo stato della sessione dell'utente all'interno dell'applicazione.

Le tre parti di MVC sono così organizzate:

1. View: Una serie di pagine JSP e di fogli di trasformazione XSLT
2. Controller: due servlet che si occupano dello smistamento del traffico
3. Model: dei componenti JavaBeans o EJB dove è inserita tutta la business logic della applicazione.



I 3 livelli previsti sono dunque:

Presentation Layer (Front End)

L'interfaccia di fruizione degli applicativi verrà realizzata attraverso l'uso di pagine dinamiche costruite con tecnologia J2EE-JSP (Java Server Pages) e/o attraverso la renderizzazione HTML svolta dall'utilizzo della tecnologia XSLT.

La pagina JSP sarà costituita da codice HTML e da codice JavaScript.

Il client è costituito dal solo Browser standard che supporti HTML ver 4 o superiore.

Il livello di *Presentation* è rappresentato dai server Web che gestiscono la parte d'interfaccia verso l'utente generico Internet.

In quest'area si collocano tutte le componenti che hanno il compito di riconoscere l'utente e la modalità di accesso di volta in volta adottata (quindi un browser piuttosto che un terminale Wap o quant'altro), validano e profilano l'utente mediante il riconoscimento e la verifica di parole chiave e/o certificati digitali e creano, in base a queste informazioni, una presentazione del portale d'accesso personalizzata.

La verifica delle credenziali dell'utente potrà avvenire in diverse modalità e per diverse finalità; in particolare, vi sarà un primo riconoscimento per il solo scopo di fornire all'utente viste personalizzate del portale di accesso e per la gestione interna della profilazione (utente registrato), in più potrà essere richiesta un'identificazione tramite parole chiave per l'accesso ai servizi (utente riconosciuto)

Application layer (Business Logic)

L'implementazione della logica di business e l'astrazione dal database è costituita classi Java (servlet, JavaBeans o EJB, Web Services e utilizzo del protocollo SOAP) per l'accesso ai dati e la logica dell'applicazione.

All'interno degli Application Server saranno gestiti i meccanismi di autenticazione.

Una volta che l'utente sia stato "abilitato" alla navigazione e alla fruizione delle informazioni e dei servizi disponibili, le componenti presenti in quest'area avranno il compito d'instradare le richieste effettuate verso i fornitori del servizio e di garantire la continuità e la fruibilità dei servizi, mediante tecniche di ripartizione automatica dei carichi in commistione con le componenti di Business Logic.

Il livello di Business Logic rappresenta il vero e proprio cuore del sistema e racchiude la logica applicativa. A essa sono demandate tutte le funzionalità di accesso ai dati e quelle d'integrazione, trasformazione e cooperazione verso sistemi omogenei e/o eterogenei, così come la gestione della consistenza della base concettuale.

Il compito principale di questa componente è quello di ricevere richieste provenienti dalla componente di Presentation, dopo opportune verifiche; le richieste arrivano corredate delle

informazioni di autenticazione e autorizzazione dell'utente, che possono essere nuovamente utilizzate per nuovi controlli e/o filtri sulle informazioni da rendere disponibili.

Sulla base delle richieste pervenute avverrà l'accesso diretto ai dati o ai servizi necessari alla composizione della risposta per l'utente finale. L'accesso ai dati potrà avvenire in modalità transazionale tra sorgenti omogenee, per garantire quanto più possibile la consistenza e la coerenza dei dati, dovunque essi siano memorizzati

Database Layer

Il livello Data Layer permette al Business Logic di reperire le informazioni sulle quali effettuare le elaborazioni; i componenti del Data Layer contengono la logica che permette sia l'interfacciamento verso sistemi di memorizzazione dati (data base) che verso file system gerarchici.

Software di base

Il sistema realizzato attualmente utilizza il seguente insieme di componenti software:

1. Sistema operativo: Microsoft 2003 Server / Linux / Sun Solaris / AIX / SCO Unix
2. Web Server: Microsoft Internet Information Server / Apache
3. Java Application Server : JBOSS-Tomcat
4. Extraway Information Server come piattaforma unica avente funzioni di XML Native Database e di Information Retrieval

Software applicativo

I software applicativi che verranno realizzati sono:

2. Applicativi per la catalogazione (XDAMS) delle varie tipologie di archivi
3. Applicativi per le funzioni di front-office: portale Intranet ed Internet
4. Applicativi di Content Management e Content Publishing
5. Applicativi di monitoraggio e generazione statistiche
6. Applicativi per la gestione degli utenti e delle credenziali sulle varie aree

Infrastruttura

L'architettura del servizio dovrà essere progettata in modo da garantire le seguenti condizioni:

1. Alto grado di affidabilità
2. Fruibilità (interfacce semplici utilizzabili anche da utenti non esperti)
3. Elevati livelli di performance (tempi di risposta adeguati)
4. Scalabilità lineare senza modifiche alla struttura software

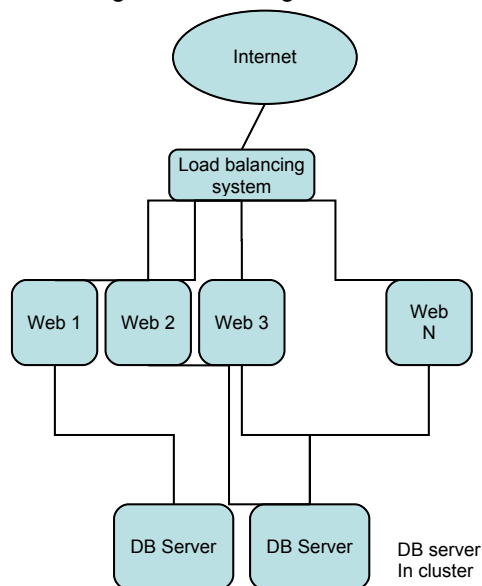
La natura di questo tipo di applicazioni richiede l'utilizzo di una tecnologia che raggiunga l'eccellenza in termini di scalabilità, alta disponibilità ed integrazione degli ambienti preesistenti, per garantire non solo le performance, ma anche la massima affidabilità nella protezione dei dati (*reliability*), la continuità del servizio (*availability*) e la rapida individuazione e correzione degli errori (*serviceability*).

Tali considerazioni trovano la propria applicazione in un'architettura tecnologica che soddisfi i requisiti di espansibilità orizzontale e verticale.

Espansibilità verticale, ovvero un sistema che sia in grado di essere messo in produzione su un unico livello hardware e sia di espandersi sui i n livelli definiti.

L'espandibilità orizzontale deve essere garantita dalla clusterizzazione dei servizi e/o dalla gestione degli stessi su macchine dedicate. Quindi l'applicazione dovrà poter essere gestita, in relazione alle esigenze di servizio, da un'unica macchina, da una macchina per ciascun livello infrastrutturale, da una coppia di macchine per livello (clustering per garantire l'alta affidabilità del servizio), da più coppie di macchine in cluster per ciascun servizio di ogni livello.

Il modello finale al quale giungere è quindi quello illustrato nella figura, e la progettazione iniziale del sistema sarà tale da non preconstituire vincoli al requisito posto di espandibilità verticale ed orizzontale.



Architettura software di dettaglio

Per la realizzazione dell'infrastruttura relativa alla piattaforma xDAMS verrà predisposto un framework applicativo che si basa sulle seguenti componenti software:

HTTP Server distribuito in configurazione di "Load Balancer" su n Front-End WEB: Si può utilizzare indifferentemente Microsoft Information Server 6.0 su piattaforma Microsoft oppure Apache http Server su Piattaforma Linux quale Web Server di riferimento. La nostra soluzione prevede la piena compatibilità con entrambe le soluzioni. FTP Server attivato sui server di Database, attraverso cui gestire l'upload degli allegati digitali sullo Storage server

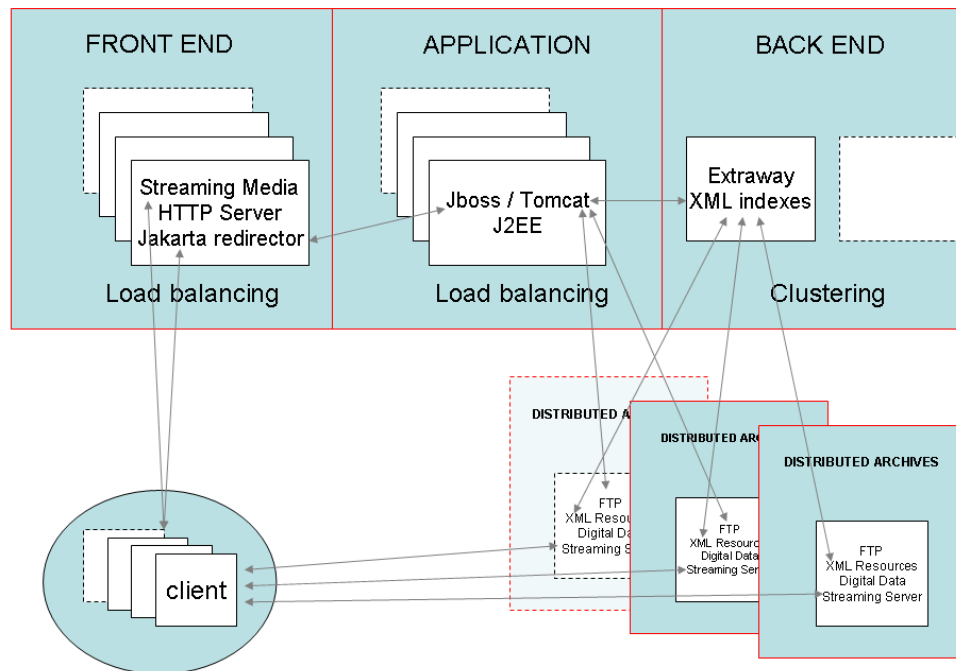
Installazione JK redirector plugin che permette il tunnelling http tra i front-end Web e l'Application server.

Installazione della Java 2 Enterprise Edition ver . 1.4.x

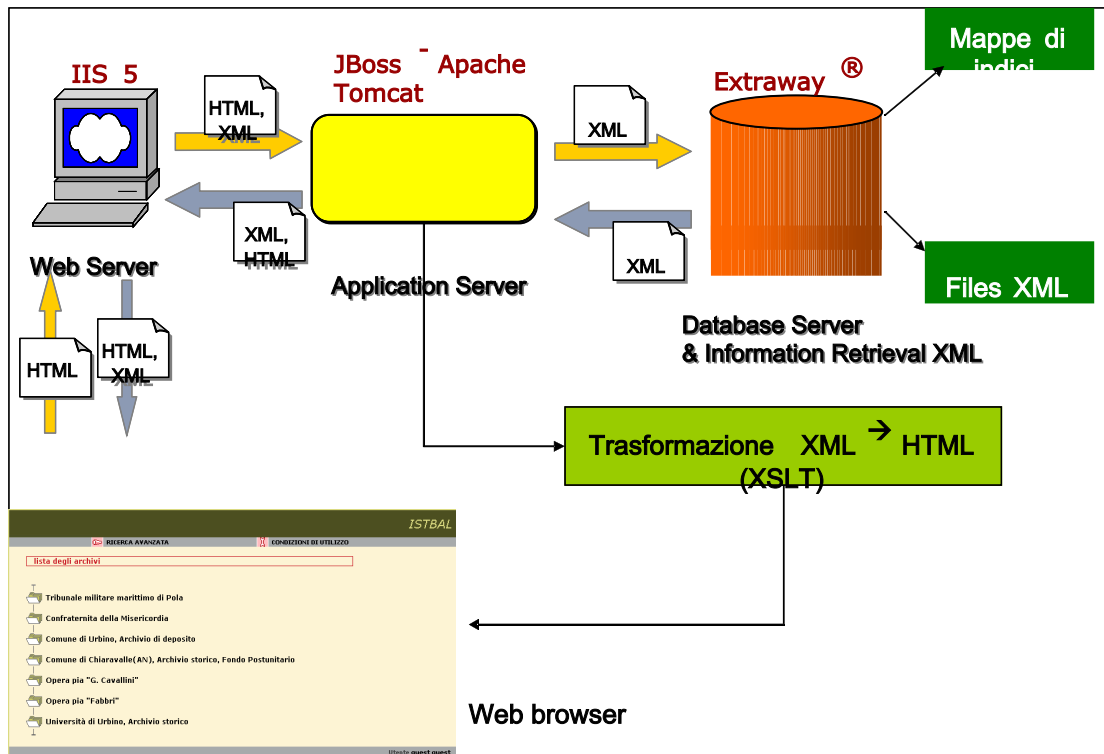
Application Server Java quale Jboss 3.2.1-Tomcat 4.1.24 quale componente per la gestione della logica di business di tutta la piattaforma.

Extraway Information Retrieval quale motore di ricerca di tutta la piattaforma, in grado di gestire le banche dati in XML nativo. Il motore è assolutamente "multipiattaforma" in

quanto può essere installato su Solaris/SPARC, Linux, AIX, SCO Open Server, SCO Unixware, Windows 2K.
Possiamo considerare il seguente schema di riferimento.



Inoltre un esempio di flussi/richieste delle varie componenti interessate è descritto nella figura seguente:



La piattaforma XML di riferimento: Extraway® Information Retrieval

La tecnologia XML, formalizzata come standard permette da un lato di definire nuovi formati di scambio informazioni e comunicazione (ad esempio il WML per i cellulari), dall'altro di rappresentare oggetti in un formato aperto, non proprietario, e quindi fondamentalmente indipendente da un apposito strumento di software di interpretazione proprietario, soggetto alle evoluzioni dell'azienda produttrice e del mercato

Extraway® è un database XML nativo, in grado di memorizzare, indicizzare e rendere disponibili i documenti per la ricerca. I dati non sono conservati in un formato proprietario, ma sono mantenuti nella forma nativa (XML) sul file system.

E' un database XML nativo per applicazioni di archiviazione con le seguenti principali caratteristiche:

1. Indicizzazione full text (boolean, similarity searches, relevance ranking, ecc)
2. Funzioni avanzate d'Information Retrieval quali ad esempio:
 - Tipo di dato (stringa, insieme di parole, numero data ecc.)
 - Prossimità (Adj, Near, Context)
 - Stoplist
 - Wildcard
 - Ordinamento per rilevanza
 - Documenti simili
 - Doppia indicizzazione
 - Concatenazione indici
 - Estensione di motore per indici ad alte performance (ad es. per link automatici)
 - Vocabolari
 - Analisi spettrali

- Gestione thesauri sia in inserimento controllato che in ricerca
- 3. Ricca definizione degli indici
- 4. Accesso esteso ai vocabolari, analisi spettrali e thesauri multilivello
- 5. Vista Gerarchica su intere collezioni di documenti XML
- 6. Estrazione dei frammenti XML con le evidenze di ricerca
- 7. Modulo add-on di linguaggio naturale
- 8. Motore (XML+mappe e indici)
- 9. Livello applicativo (JSP, Generic+Specs.JAR)
- 10. Livello trasformazione file (XSLT ISAPI)
- 11. Livello interfaccia (HTML4, CSS, JPG, applets)
- 12. Piattaforme: Solaris, Linux, Windows

Extraway® XML Native Database è una tecnologia innovativa che muove dai seguenti obiettivi principali:

- 13. tutti gli oggetti gestiti dal motore documentale, indipendentemente dall'interfaccia con cui sono stati prodotti sono fisicamente file o parti di file XML;
- 14. il sistema è in grado di catalogare e trattare nel tempo file XML prodotti e aggiornati da altri applicativi;
- 15. il sistema è dotato di una libreria di filtri, atti a trasformare file da vari formati nativi al formato XML.

Extraway® consente di gestire direttamente e produrre prioritariamente una '**Information Unit**' XML garantendo i seguenti ed importanti benefici:

- 16. **Intelligibilità nel tempo** - comunque possa evolversi il formato XML gli oggetti descritti dalla prima versione in poi saranno sempre decifrabili nella loro struttura, contesto, gerarchia e presentazione.
- 17. **Efficienza** - non è necessario effettuare conversioni in ingresso ed uscita per scambiare informazioni strutturate in XML con altre applicazioni; inoltre la tecnologia ad oggetti supporta l'utilizzo di puntatori e collezioni che consentono di ottenere prestazioni elevate sia in ricerca sia nel controllo dell'accesso concorrente.
- 18. **Flessibilità** - mediante la catalogazione dei modelli di definizione (DTD e Schema) e delle relative versioni, il modello della base dati può evolvere nel tempo senza dover ristrutturare il formato fisico dei dati già inseriti, è possibile cioè generare nuove DTD e Schema senza ritoccare le informazioni già archiviate.
- 19. **Integrazione con la logica applicativa** - La tecnologia a oggetti permette di dichiarare nell'oggetto stesso i metodi con cui essi vengono utilizzati, quindi la logica applicativa risulta definita e associata in modo diretto ai dati che gestisce.

Extraway® apre rilevanti possibilità per lo sviluppo di sistemi documentali informatici, soprattutto perché consente, oltre alla gestione dei riferimenti esterni al documento e alle sue partizioni, anche il trattamento della struttura logica e semantica dei contenuti e dei metadati.

Questi sviluppi si possono tradurre nella decisione di:

- 20. **promuovere**, all'interno di un'organizzazione, interventi di razionalizzazione e semplificazione delle tipologie documentarie mediante la definizione di rappresentazioni specifiche con lo scopo di ottimizzare l'elaborazione automatica dei documenti, garantire coerenza, qualità e uniformità dei materiali;
- 21. **sviluppare** strumenti di recupero e riutilizzo di documenti (o di componenti interne) ai fini di una distribuzione/condivisione di contenuti destinati a durare nel tempo,
- 22. **gestire** formati multipli;

23. **utilizzare** i sistemi di validazione XML anche a fini di sicurezza e di integrità;
24. **controllare** e ottimizzare i cicli di gestione delle informazioni.

Extraway® XML Native Database e' l'unico DATABASE XML in cui ogni oggetto (Information Unit) è trattato come un documento archivistico e quindi corredato, in automatico, dai metadati indispensabili per la costruzione del contesto e la preservazione a lungo termine. La durata nel tempo e' ottima e l'intellegibilità sull'intero archivio e' alta.