



## Technical Report TR2.2.12

### MAG REPOSITORY - DEFINIZIONE DI UNO STORAGE SICURO DEI CONTENUTI MULTIMEDIALI GRANULARI, IN GRADO DI GESTIRE CONTENUTI SEMANTICAMENTE ANNOTATI

*ABSTRACT: Il presente technical report contiene una descrizione del MAG repository (MAG Teca) e della sua evoluzione per arrivare a gestire gli oggetti granulari (Contlet), al fine di consentire una gestione delle annotazioni semantiche, la loro archiviazione sicura e il relativo recupero delle stesse (retrieval).*

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



## INDICE GENERALE

1. Una teca digitale per gli Oggetti (Contlet).....	3
2. Evoluzione del sistema Teca.....	6
2.1. Uno storage per Contlet.....	6
2.2. Principali funzionalità.....	8
3. Casi d'uso.....	11
3.1. Attori.....	11
3.2. Azioni.....	14
3.3. Scenari.....	16
4. Architettura di sistema.....	24
5. Specifica dei moduli software.....	33
5.1. Business Logic.....	33
5.2. Presentation Logic.....	34
5.3. XML Parsing Engine.....	35
5.4. Web Services.....	37
5.5. OAI Services.....	38
5.6. Search engine.....	39
5.7. Motore di concurrent per elaborazioni in background.....	39
5.8. Repository Centrale.....	40

## INDICE DELLE ILLUSTRAZIONI

Illustrazione 1: MAG Repository – stratificazione dell'informazione.....	7
Illustrazione 2: MAG Repository – Importazione.....	8
Illustrazione 3: MAG Repository – casi d'uso.....	13
Illustrazione 4: MAG Repository – diagramma di deployment.....	26



## 1. UNA TECA DIGITALE PER GLI OGGETTI (CONTLET)

L'obiettivo che ci siamo posti è quello di far evolvere un sistema di Teca digitale, già in grado di gestire risorse digitali eterogenee e relativi metadati gestionali amministrativi nel formato MAG, al fine di garantire uno *storage* sicuro anche agli oggetti riusabili di contenuto elementare (*Contlet*), frutto di un'attività di annotazione semantica analitica.

In base a quanto già descritto nel TR1.1.4, è possibile rappresentare l'oggetto elementare attraverso un'estensione del formato MAG, introducendo una sezione in grado di contenere le informazioni necessarie ad individuare singole risorse digitali o parti di esse e a collegarle a un'ontologia di dominio.

Il MAG *repository*, in cui vengono archiviati i dati digitali, dovrà rendere disponibile un'interfaccia uniforme per la gestione delle diverse tipologie di dati, sia che si tratti di risorse corredate di opportuni metadati MAG, sia che si tratti di oggetti *Contlet* che, in quanto informazioni di natura digitale, trovano la loro giusta collocazione all'interno del *repository*.

La Teca digitale "evoluta" per la gestione di contenuti multimediali granulari (*Contlets*) mantiene comunque invariate alcune delle sue caratteristiche originarie. Essa nasce come sistema di archiviazione aperto, orientato alla preservazione delle risorse digitali a lungo termine, che gestisce in maniera integrata i contenuti multimediali (immagini, audio, video etc.) ed i relativi metadati digitali, permettendone l'accesso mediante apposita interfaccia.

Di seguito si riportano le sue caratteristiche fondamentali:

- **Conformità allo standard OAIS** (*Open Archival Information System*) -> per la conservazione permanente degli oggetti digitali.
- **Applicazione Web-Based** -> per una gestione remota di tutti i servizi, regolata in base ai privilegi di accesso dell'utente connesso.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



- **Rispetto dello standard MAG** -> utilizzo del formato di metadati amministrativo-gestionali indicato dalle specifiche ICCU.
- **Esposizione dei servizi tramite WEB services** -> per l'integrazione con moduli software esterni.

Il modello logico-funzionale è coerente con gli standard più avanzati per la conservazione permanente degli oggetti digitali (in particolare: **OAIS**, *Open Archival Information System*, [standard ISO 14721:2003](#)), utili a non vanificare nel tempo gli investimenti fatti per la digitalizzazione.

La Teca è stata concepita come applicazione **Web-Based**, che permette una gestione remota di tutti i servizi, esposti in base ai privilegi di accesso dell'utente connesso. In particolare, sono consentite le operazioni di *import/export* oltre che in locale - trasferendo fisicamente i dati dalle unità nastro o ottiche direttamente sul *server* - anche da remoto, ad esempio riversando i dati nelle *directory* di scambio tramite connessione FTP; questo permette una gestione centralizzata delle collezioni digitali.

Altra caratteristica indispensabile, ovviamente, è la compatibilità col formato **MAG 2.0** - metadati amministrativi-gestionali - per le risorse digitali, essendo questo lo standard di metadati richiesto a livello nazionale per la digitalizzazione, ma anche conforme agli standard internazionali per la gestione del digitale.

Infine, l'integrazione del *repository* con altri applicativi esterni è garantita da un'interfaccia **SOAP** (*Simple Object Access Protocol*), per l'interrogazione e il recupero delle informazioni digitali contenute nel *repository*. Tale interfaccia presenta una serie di API già definite, ma ampliabili in base alle nuove esigenze di connessione. A tale protocollo si affianca quello OAI-PMH, in grado di rendere disponibili in modo semplice e standardizzato i contenuti che si vogliono rendere di pubblico dominio.

Le funzionalità già presenti nella MAG-Teca si possono riassumere nelle seguenti categorie:

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



- Gestione multi-utente con livelli differenziati di accesso
- *Import/export* e gestione delle risorse digitali e dei relativi MAG
- Indicizzazione dei dati per la ricerca sugli oggetti MAG a partire dai dati contenuti nella sezione <bib>
- Interfaccia WEB di amministrazione
- Interfaccia per la pubblicazione dei contenuti archiviati

Tali funzionalità sono state potenziate nell’ottica di gestire anche un livello più analitico di informazione, senza stravolgere il sistema lasciando immutata la logica di gestione dei dati sempre basata sui metadati MAG. In particolare, verranno gestiti gli oggetti granulari (*Contlets*) contenuti nel MAG esteso ed esposte, mediante interfaccia SOAP, le funzionalità necessarie ai moduli di annotazione semantica e di ricerca previsti dal progetto LC3, che permettono anche un accesso diretto a ciascun oggetto di contenuto elementare.

<i>PROGETTO LC3</i>	<i>Revisione n*</i>	<i>6</i>	<i>Del</i>	<i>23/01/09</i>
---------------------	---------------------	----------	------------	-----------------



## 2. EVOLUZIONE DEL SISTEMA TECA

L'idea di progettare un sistema di MAG Teca nasce dal fatto che i dati digitali appaiono di per sé ingestibili: si rende pertanto necessario accompagnarli con metadati amministrativi gestionali in formato XML, che descrivono il contenuto del dato digitale, ne garantiscono l'autenticità e integrità, contengono i dati tecnici utili a successive applicazioni con sistemi hardware/software diversificati, mantengono memoria delle apparecchiature con cui sono stati prodotti etc.

### 2.1. Uno storage per Contlet

Nella sezione <bib> i metadati descrittivi sono configurati sottoforma di estensione dello standard Dublin Core: tale descrizione solitamente è utilizzata a livello macroscopico sull'intero documento o sull'unità catalografica più analitica, ad esempio, nel caso di spogli ottenuti mediante l'utilizzo di metadati strutturali <stru>, che permettono di recuperare i legami fra "madri" e "figlie".

In ogni caso, si vuol precisare che la sezione <bib> riesce a recuperare in dettaglio i dati catalografici, senza rivelarsi necessaria per effettuare un'efficace annotazione analitica del contenuto dei documenti. L'idea di fondo della MAG Teca "evoluta" è quindi stata quella di aggiungere un ulteriore strato di informazione di tipo analitico, basata su collegamenti a ontologie di dominio e utile a realizzare uno *storage* in grado di gestire e rendere disponibile questa informazione granulare.

I files MAG raccolgono e contestualizzano i singoli file multimediali: un insieme di pagine digitalizzate, ad esempio, sono riferibili ad un certo libro, nella versione di MAG esteso è possibile poi rappresentare tale libro come composizione di oggetti di contenuto elementare, cioè raggruppamenti di una o più parti di quelle pagine digitali riferibili ad un particolare significato espresso mediante un'ontologia di dominio. Il MAG è legato all'oggetto fisico, mentre il *Contlet* è racchiuso in una parte dell'oggetto fisico che può essere una frazione anche di quella che potrebbe essere considerata la parte indivisibile

(per tornare all'esempio, una pagina del libro, una porzione di illustrazione, uno spezzone di audio etc.)

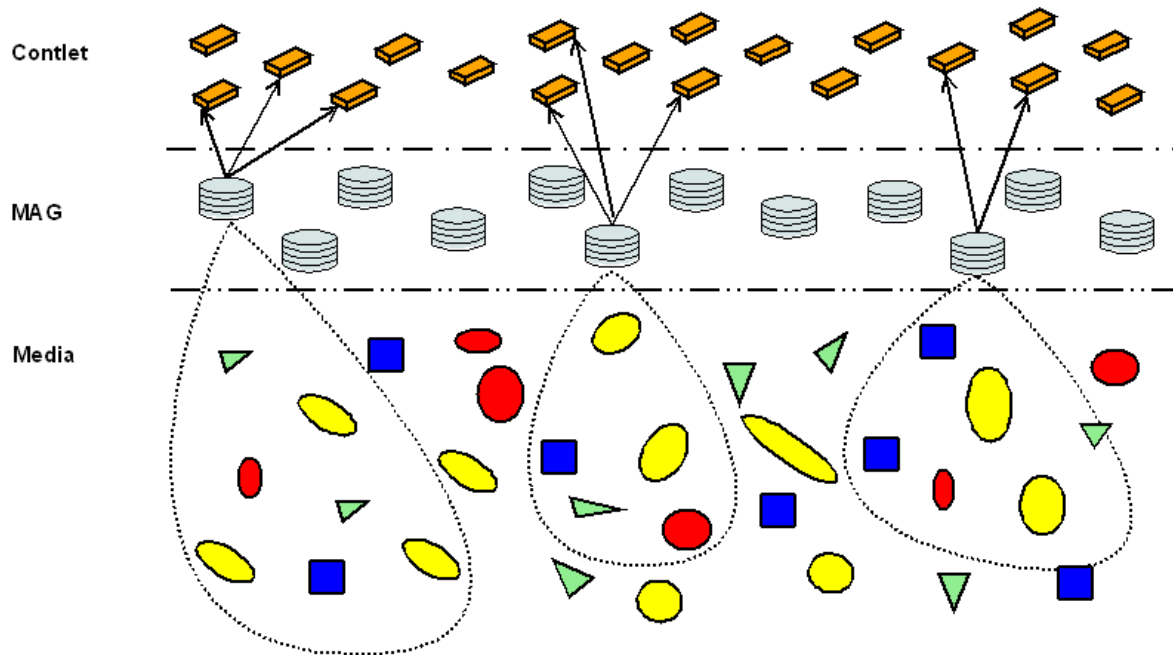


Illustrazione 1: MAG Repository – stratificazione dell'informazione

Di base, anche gli oggetti elementari avranno un proprio set di dati derivati dallo standard Dublin Core e potranno essere ricercati rispetto ad esso. Sarà poi il MAG-Repository a rendere possibile la coesistenza di indicizzazioni esterne, che potranno lavorare ad esempio su base semantica oppure in base ai profili dell'utenza utilizzatrice.

Come per i dati non annotati semanticamente, la procedura gestionale delle informazioni digitali prevede che esse vengano sottoposte ad un'operazione di verifica prima di essere resi disponibili ai *client*. Nel caso in cui questa operazione non abbia esito positivo, l'intero processo di "importazione" viene annullato.

In presenza di annotazioni incapsulate nei MAG, questa operazione sarà potenziata al fine di poter isolare i singoli "contributi" di cui il file XML si compone, cioè i singoli *Contlet*. In questo modo, sarà possibile accedere in modo diretto a ciascun oggetto di contenuto

elementare. Nello specifico, tale operazione verrà fatta isolando ciascun Contlet durante la fase di importazione del file MAG.

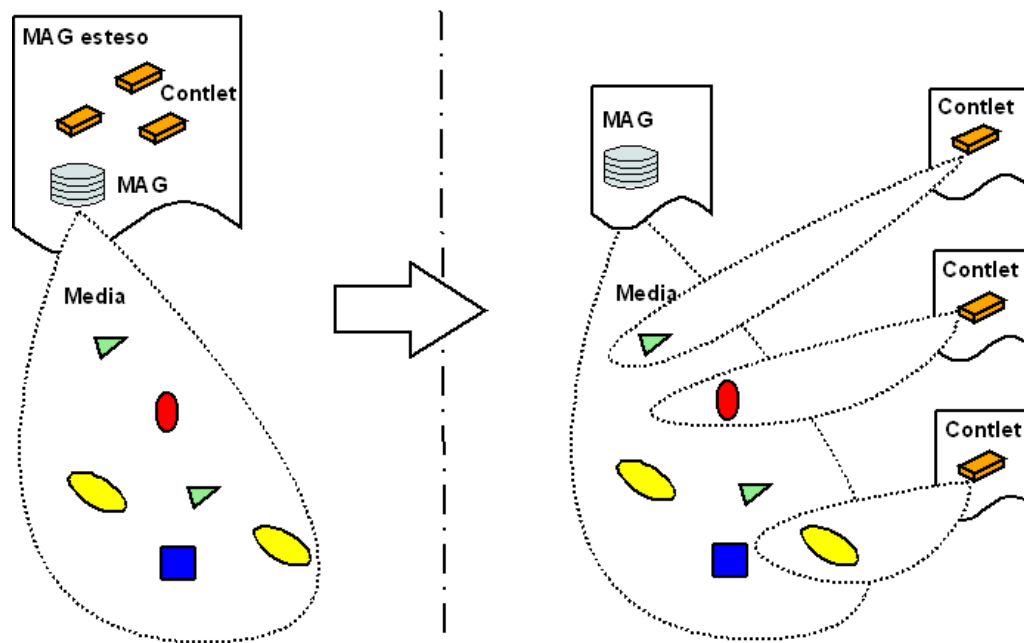


Illustrazione 2: MAG Repository – importazione

L'operazione di importazione ha, inoltre, la funzione di sottoporre a indicizzazione il file contenente i metadati, nonché di informare il server OAI della presenza di un nuovo documento tramite la generazione di un file, dove, oltre ai metadati, saranno contenuti i nuovi link alle risorse digitali.

## 2.2. Principali funzionalità

Il *MAG-Repository* - nel rispetto dello standard OAIS - gestisce risorse digitali intese come oggetti informativi composti da dati e metadati MAG ad essi associati. I dati che possono essere archiviati in *MAG-Teca* sono file digitali di vario genere e formato (immagini, audio, video, testi etc.) derivati dal processo di digitalizzazione delle fonti non digitali. I metadati richiesti devono essere conformi al formato MAG versione 2.0 o successive.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



Tramite l'interfaccia WEB di amministrazione possono essere effettuate operazioni di *import/export*, così come la rimozione di un oggetto informativo (dati e metadati). Queste funzionalità verranno rese disponibili anche tramite protocollo SOAP, oltre a funzionalità specifiche per la ricerca e per l'annotazione semantica previste dal progetto LC3.

Nell'ambito del progetto LC3 verrà ampliato il concetto di oggetto informativo della Teca digitale in modo da comprendere, non solo il pacchetto composto da MAG tradizionale e relative risorse digitali ad esso collegate, ma anche l'oggetto MAG esteso contenente *Contlet*, cioè l'insieme composto da metadati (MAG e *Contlet*) e riferimenti a risorse digitali.

La fase di importazione dati all'interno del *repository* centrale della teca digitale, implica la rielaborazione dei file MAG, trasformando tutti i collegamenti relativi alla localizzazione dei dati digitali collegati (re-indirizzamento), sì da renderli coerenti con la nuova collocazione che gli stessi assumeranno all'interno del *repository* MAG-Teca. Durante questa fase i *Contlet* presenti all'interno del file MAG vengono isolati e viene prodotto per ogni *Contlet* un oggetto distinto in modo da poter rendere fruibile ciascun *Contlet* tramite URI. In parallelo vengono, inoltre, indicizzate tutte le informazioni presenti nella sezione <bib> ed inserite all'interno del *database* relazionale della teca digitale. Tale operazione permette di velocizzare le future operazioni di ricerca da parte degli "attori" che interagiscono con il sistema. I dati soggetti all'attività di importazione devono comunque trovarsi all'interno di un'apposita area di transito temporaneo, denominata *repository* temporaneo.

In fase di esportazione, gli oggetti digitali vengono riversati all'interno del *repository* temporaneo, nella posizione che avevo prima dell'importazione.

Per garantire nel tempo la persistenza dei *Contlet* già immessi nel sistema, in generale i MAG e i *Contlet*, con le relative risorse digitali associate, una volta memorizzati all'interno della MAG-teca non dovrebbero essere modificabili. Eventuali modifiche a tali oggetti comportano sempre la creazione di nuovi MAG e *Contlet*. In particolare, per i *Contlet* è prevista una relazione di *versioning* unilaterale mediante un puntatore alla struttura dati

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



aggiornata: per fare un aggiornamento quindi, deve essere creato un altro *Contlet* come nuova versione di quello annotato precedentemente e deve essere gestito il *versioning* del *Contlet*.

Per casi eccezionali è stata, comunque, prevista la rimozione degli oggetti informativi e viene svolta in due fasi successive:

- **rimozione logica** – permette di impostare un *flag* all'interno del sistema per un determinato oggetto, impedendone ogni ulteriore visualizzazione (tale accorgimento permette di evitare perdite accidentali di dati).
- **cancellazione fisica** – viene effettuata in modo periodico dall'amministratore generale della teca digitale, tramite l'attivazione di un'apposita procedura eseguita in modalità "concorrente" dal sistema, senza dunque comportare alcun rallentamento delle normali operazioni eseguite sulla teca.

Gli oggetti presenti all'interno della MAG-teca di LC3 saranno dotati di un identificatore univoco, che ne garantisce l'individuazione in fase di interrogazione da remoto.

Le risorse digitali presenti all'interno della teca di LC3 possono essere acquisite tramite *web services*. Sempre nell'ambito del progetto LC3, oltre a prevedere una ricerca semplice sui campi del DC sia dei MAG che dei *Contlet*, è stata prevista l'integrazione nel sistema di *MAG-Repository* di una modalità di ricerca di tipo *Full-Text*: il sistema potrà rispondere, dunque, ad un'interrogazione diretta da parte di un'entità esterna (su protocollo SOAP), con risultati multipli collegabili all'oggetto ricercato.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------

### 3. CASI D'USO

Nel presente capitolo, illustriamo alcuni casi d'uso (illustrazione 2) che caratterizzano il software *MAG-Repository* per l'archiviazione dei contenuti multimediali derivati dall'attività di digitalizzazione.

Il diagramma d'uso (o UCD, *Use Case Diagram*) permette di mettere in evidenza le funzioni e i servizi offerti dal sistema in analisi, così come sono percepiti e utilizzati dagli "attori", ovvero da coloro che interagiscono con il sistema di *MAG-Repository*.

Osserviamo come all'interno della presente descrizione venga utilizzato il generico termine di "archivio" per indicare in modo indifferente l'insieme di oggetti digitali depositati nello *storage*. Inoltre, si precisa che il termine "oggetto" viene utilizzato per indicare genericamente sia i dati che i metadati associati al documento digitale.

#### 3.1. Attori

Nell'interazione con l'applicazione sono identificabili cinque tipologie di attori:

- **Amministratore generale di sistema:** è l'amministratore dell'intero sistema *MAG-Teca*, cui spetta il compito di gestire lo *storage*. L'amministratore di sistema può inoltre svolgere tutte le attività di gestione dell'archivio, in particolare accede alle statistiche, gestisce la cancellazione fisica degli oggetti digitali dell'archivio, e tutte le attività compiute dall'operatore d'archivio, rappresentandone pertanto una generalizzazione, ovvero potendo effettuare sull'archivio tutte le operazioni svolte dall'operatore.
- **Operatore:** è l'utente che fisicamente accede all'archivio per svolgere operazioni quali la ricerca, la consultazione dei dati e dei metadati, l'*import/export*, la rimozione logica degli oggetti.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



- **Annotatore semantico:** Tramite operazione di *harvesting* allinea i propri metadati con quelli del *repository* centrale MAG-Teca. A partire da tali metadati può accedere alle singole risorse digitali che, tramite apposite interfacce, potrà visualizzare e annotare. Le annotazioni, che verranno incapsulate all'interno di MAG estesi, potranno essere importate nella MAG-Teca mediante apposita funzione ed, altresì, essere aggiornate e cancellate.
- **Semantic IR Tool:** tramite operazione di *harvesting* allinea i propri dati con quelli della MAG-teca, in particolare accumula gli oggetti *Contlets* su cui rendere possibili le ricerche di natura semantica.
- **Search Manager:** Sistema di ricerca generico esterno al MAG-Repository che esegue ricerche mediate protocollo SOAP.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------

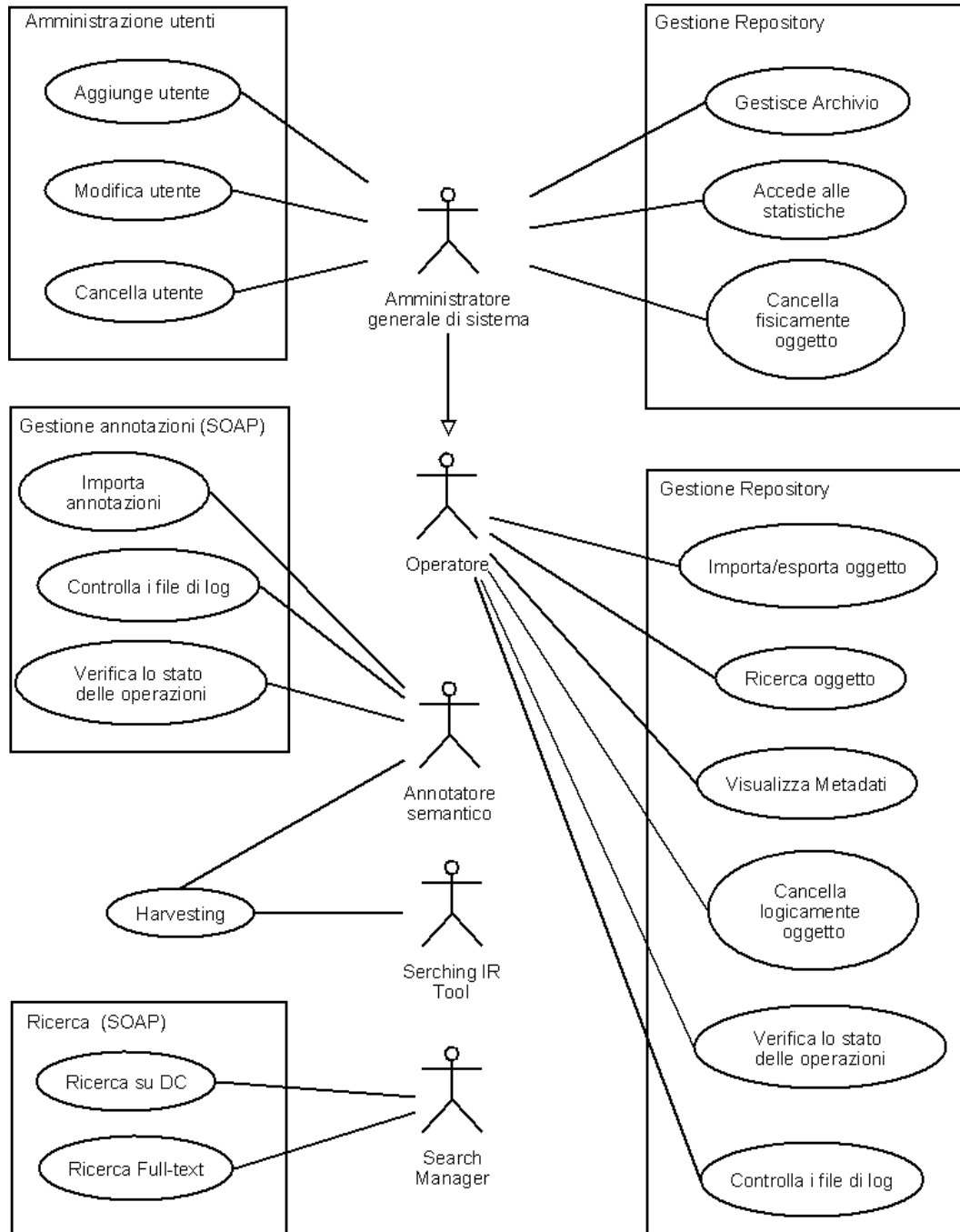


Illustrazione 3: MAG Repository – casi d'uso

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### 3.2. Azioni

Di seguito riportiamo una breve descrizione delle operazioni che ciascun attore sopra descritto è in grado di svolgere sul sistema. Per una descrizione dettagliata del flusso degli eventi associata a ciascuna azione si rimanda al paragrafo seguente.

*Amministratore di sistema:*

- x modifica le informazioni generali che qualificano l'archivio e le impostazioni di accesso;
- x accede alle statistiche relative agli accessi e alle annotazioni
- x cancella fisicamente gli oggetti già cancellati logicamente dall'interno dell'archivio;
- x verifica lo stato delle proprie operazioni effettuate così come di quelle effettuate da altri utenti;
- x consulta i file di log generati dal sistema MAG-Teca relativi a ciascun operatore presente nel sistema;
- x aggiunge un nuovo utente all'elenco di quelli registrati. Le categorie di utenti che l'amministratore può aggiungere sono: operatore d'archivio e amministratore, oppure gli "utenti" Annotatore Semantico e *Sematic IR Tool* (che fornisce l'autorizzazione di accesso al sistema da parte di una particolare installazione del software esterna alla MAG-Teca);
- x modifica i dati relativi agli utenti già presenti nel sistema
- x rimuove un qualunque utente (tranne se stesso). Osserviamo come la cancellazione di un utente non si traduca in una rimozione fisica all'interno del *database* del record associato all'utente, ma comporti la semplice impostazione di un *flag* di rimozione, che ne impedisce la visualizzazione da parte degli amministratori e blocca i permessi di accesso di quell'utente all'interno degli archivi.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



*Operatore:*

- x importa/esporta un oggetto all'interno dell'archivio per il quale ha un *account* di accesso valido;
- x effettua ricerche all'interno del *MAG-Repository* e ne consulta le schede con i dati risultanti dalle attività di ricerca;
- x accede ai metadati nel formato MAG
- x verifica lo stato delle operazioni eseguite;
- x eccede ai file *log* con il dettaglio dell'operazione eseguita;
- x cancella logicamente gli oggetti dall'interno dell'archivio.

*Software Annotatore semantico:*

- x mediante *harvesting* aggiorna i propri metadati con quelli della MAG-Teca
- x importa nel sistema annotazioni prodotte mediante propria interfaccia accedendo ai singoli file digitali tramite uri presenti nel MAG.
- x aggiorna annotazioni già fatte potendole visualizzare nei raggruppamenti originali
- x cancella le annotazioni già inserite nel sistema

*Software Semantic IR Tool:*

- x attraverso *harvesting* aggiorna i propri metadati per allinearsi al *repository* di MAG-Teca.

*Software Researcher System:*

- x esegue ricerche sui campi DC del MAG e dei *Contlet*
- x esegue ricerche *full-text* sui campi DC del MAG o dei *Contlet* o sulle risorse di tipo *Text* ad essi collegate.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### 3.3. Scenari

Nel presente paragrafo vengono approfonditi gli scenari associati ai singoli casi d'uso sopra descritti. Lo scenario fornisce una descrizione “sottoforma di testo scritto”, delle azioni e interazioni che l'attore deve svolgere per portare a termine l'operazione descritta dal caso d'uso stesso. All'interno di ciascuno scenario vengono raggruppate sia la situazione standard, ovvero quella che porta ad un'esecuzione con successo dell'operazione (*scenario base*), sia una o più descrizioni che delineano il comportamento del sistema nel caso in cui l'interazione fra questo, l'utente ed il sistema si discosti da quanto previsto dallo scenario base (*scenario alternativo*). Osserviamo, infine, come all'interno degli scenari alternativi siano incluse sia attività senza successo, che altre con successo (comportamenti alternativi).

#### **Aggiungi un utente**

- L'amministratore di sistema accede alla sezione “gestione degli utenti”
- Seleziona l'operazione per la creazione di un nuovo utente
- Il sistema visualizza la *form* con i campi per la compilazione dei dati associati al profilo utente
- L'amministratore compila i campi e dà l'Ok per attivare la creazione del nuovo utente
- Il sistema effettua la creazione del nuovo utente, mostrando un messaggio di successo al completamento dell'operazione
- Se nel corso dell'operazione di creazione si verificano errori, il sistema mostra all'amministratore un messaggio di avviso con il problema riscontrato

#### **Modifica i dati di un utente**

- L'amministratore del sistema accede alla sezione per la gestione degli utenti dell'archivio

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



- L'amministratore seleziona un utente e, quindi, dà l'Ok per la visualizzazione e la modifica delle proprietà dell'utente
- Il sistema visualizza la *form* contenente i campi riempiti con i dati dell'utente selezionato
- L'amministratore modifica i campi ed attiva la modifica dei dati
- Il sistema aggiorna i dati associati all'utente, mostrando un messaggio di successo al completamento dell'operazione
- Se nel corso dell'operazione di modifica si verificano degli errori, il sistema mostra all'amministratore un messaggio di avviso con il problema riscontrato

### **Rimuovi un utente**

- L'amministratore del sistema accede alla sezione per la gestione degli utenti dell'archivio
- L'amministratore seleziona un utente e quindi dà l'Ok per la rimozione dell'utente
- Il sistema prova a rimuovere l'utente selezionato, mostrando un messaggio di successo al completamento dell'operazione
- Se nel corso dell'operazione di rimozione si verificano degli errori, il sistema mostra all'amministratore un messaggio di avviso indicante il problema riscontrato

### **Gestisci archivio**

- L'amministratore seleziona l'operazione necessaria per accedere ai dati generali dell'archivio
- L'amministratore compila i campi e dà l'Ok per confermare le modifiche

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Accedi alle statistiche**

- L'amministratore seleziona l'operazione per la visualizzazione statistiche
- sceglie la tipologia di oggetto (MAG o CONTLET) e il range di date
- accede ad una lista di risultati in ordine al numero di accessi
- selezionando un elemento dalle lista accede alla scheda dell'oggetto .

### **Cancella fisicamente un oggetto**

La rimozione fisica degli oggetti presenti all'interno dell'archivio può essere effettuata solo su quegli oggetti che sono stati cancellati logicamente in precedenza. La rimozione fisica dell'oggetto comporta la rimozione del record con i metadati dal *database* di sistema, oltre che la rimozione dei dati e del file xml associati, archiviati all'interno del *File System* del sistema

- L'amministratore dell'archivio visualizza l'elenco dei record cancellati logicamente
- L'amministratore seleziona gli oggetti che vuole rimuovere definitivamente dall'archivio e dà l'Ok per l'attivazione dell'operazione
- Il sistema procede con la rimozione fisica degli oggetti selezionati
- Se l'operazione è stata eseguita con successo, il sistema mostra un avviso di successo; se, invece, si sono verificati dei problemi e l'operazione non è stata portata a termine correttamente, il sistema mostra all'amministratore un avviso che indica il problema riscontrato

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Controlla file di log**

- L'amministratore seleziona l'operazione per la visualizzazione dei file di *log*
- Il sistema apre in una finestra il file di *log* con tutte le informazioni relative alle operazioni eseguite, gli errori verificatesi etc., indicando la data di richiesta e la fine del processo, il riferimento al file MAG, lo stato dell'operazione, le attività svolte etc.

### **Importa/esporta un oggetto**

Nel caso in cui l'utente voglia importare un oggetto presente all'interno della propria area di transito, le attività svolte saranno le seguenti:

- L'utente seleziona l'opzione per l'*import* di un nuovo oggetto all'interno dell'archivio
- Il sistema mostra l'interfaccia per la selezione dell'oggetto da importare
- L'utente seleziona, fra gli oggetti presenti all'interno della *directory* temporanea di scambio dati del sistema MAG-Teca, l'oggetto da importare e dà l'Ok al sistema per attivare l'operazione
- Il sistema effettua l'*import* del nuovo oggetto selezionato, pertanto:
  - Viene analizzato il file xml con i metadati
  - Se l'analisi ha successo, inserisce i metadati all'interno del *database* di sistema
  - Se l'analisi e l'inserimento dei metadati si concludono con esito positivo, copia i dati ed il file xml all'interno del *File System*
  - Se tutte le operazioni sopra descritte hanno avuto successo, rimuove tutti i dati dalla *directory* temporanea di scambio dati
- Se l'operazione di *import* dati ha avuto successo, mostra all'utente in risposta un messaggio di successo

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



- Se nel corso dell'operazione di *import* sono occorsi alcuni problemi che hanno impedito di completare l'operazione, mostra un messaggio esplicativo del problema occorso

Nel caso in cui l'utente voglia esportare un oggetto presente all'interno del proprio archivio, le attività svolte saranno le seguenti:

- L'utente seleziona l'opzione per l'*export* di un oggetto dall'archivio
- Il sistema mostra l'interfaccia per la selezione dell'oggetto da esportare
- L'utente seleziona, fra gli oggetti presenti all'interno dell'archivio, l'oggetto da esportare e dà l'Ok al sistema per attivare l'operazione
- Il sistema copia i dati ed il file xml associati all'oggetto, presenti all'interno del *File System*, nella *directory* di scambio dei dati
- Se l'operazione è stata eseguita con successo, il sistema mostra un avviso che attesti la riuscita; se si sono verificati dei problemi e l'operazione non è stata portata a termine (ad esempio, nel caso in cui il sistema non possieda temporaneamente i permessi di scrittura), il sistema mostra all'utente un avviso che indica il problema riscontrato

### **Ricerca un oggetto**

- L'utente seleziona l'operazione per la ricerca
- Il sistema mostra l'interfaccia comprendente la *form* di ricerca
- L'utente compila i campi per i quali effettuare la ricerca all'interno dell'archivio e dà l'Ok per avviare la ricerca, da questa *form* potrà anche selezionare la tipologia di oggetto *Contlet*
- Il sistema esegue la ricerca all'interno dell'archivio e mostra in risposta un elenco con tutti gli oggetti che soddisfano i criteri di ricerca impostati
- L'utente seleziona uno fra gli oggetti forniti come risultato
- Il sistema visualizza la scheda dell'oggetto selezionato

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Visualizza metadati**

- L'utente seleziona l'operazione per la ricerca
- Il sistema mostra l'interfaccia comprendente la *form* di ricerca
- L'utente compila i campi per i quali effettuare la ricerca all'interno dell'archivio e dà l'Ok per avviare la ricerca
- Il sistema esegue la ricerca all'interno dell'archivio e mostra in risposta un elenco con tutti gli oggetti che soddisfano i criteri di ricerca impostati
- L'utente scarica il MAG in formato xml dell'oggetto selezionato.

### **Cancella logicamente un oggetto**

- L'utente seleziona uno o più oggetti presenti all'interno dell'archivio, quindi sceglie l'operazione per la rimozione degli oggetti selezionati
- Il sistema procede con la rimozione degli oggetti selezionati. La rimozione avviene solo in modo logico (settaggio di un *flag* di cancellazione)
- Se l'operazione è stata eseguita con successo, il sistema mostra un avviso che la documenta; se si sono verificati dei problemi e l'operazione non è stata portata a termine, il sistema mostra all'utente un avviso che indica il problema riscontrato
- La rimozione logica degli oggetti fa sì che il sistema non mostri nelle visualizzazioni successive degli oggetti tutti gli elementi cancellati, nonostante questi continuino ad essere presenti all'interno del *database* del sistema. La rimozione fisica degli oggetti viene effettuata dall'amministratore del sistema tramite il caso d'uso "cancella fisicamente un oggetto".

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Verifica Stato Operazioni**

- L'operatore seleziona l'operazione per la verifica e la visualizzazione dello stato delle operazioni da lui attivate (soltanto l'amministratore potrà verificare lo stato di tutte le operazioni attivate sia da parte propria che dagli utenti dell'archivio)
- Il sistema fornisce in risposta una finestra all'interno della quale visualizza le informazioni sullo stato delle operazioni, configurate secondo le tipologie "eseguita", "in fase di esecuzione", "non eseguita".

### **Importa annotazioni (SOAP)**

- Il software di annotazione semantica si connette al sistema di MAG-Teca
- Dopo aver eseguito le annotazioni desiderate, il software di annotazione semantica richiama la funzione di inserimento nuova annotazione e provvederà a passare il dato in formato xml (MAG esteso).
- Il sistema riceve in risposta il codice del *concurrent* associato all'operazione.

### **Verifica stato importazione (SOAP)**

- Il software di annotazione semantica si connette al sistema di MAG-Teca
- Il software richiama la funzione verifica stato operazione passando il codice *concurrent*
- Il software di annotazione riceve il codice di stato dell'operazione, che può assumere i valori "in attesa", "eseguita", "in fase di esecuzione", "non eseguita"

### **Accedi al file di LOG (SOAP)**

- Il software si connette al sistema di MAG-Teca
- Il software richiama la funzione scarica stato operazione passando il codice *concurrent*
- Il software riceve il file di *log* relativo all'operazione.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Ricerca su Dublin Core (SOAP)**

- Il software si connette al sistema di MAG-Teca
- Il software richiama la funzione di ricerca su DC e fornisce l'elenco dei criteri di ricerca
- Ottiene l'elenco degli oggetti che soddisfano i criteri

### **Ricerca full-text (SOAP)**

- Il software si connette al sistema di MAG-Teca
- Il software richiama la funzione di ricerca full-text e fornisce l'elenco dei criteri di ricerca
- Ottiene l'elenco degli oggetti che soddisfano i criteri

### **Esegue l'harvesting su dati indicizzati internamente**

- Il robot di *harvesting* effettua una richiesta all'*OAI services*. Tale richiesta è codificata secondo il protocollo OAI-PMH (*Open Archive Initiative – Protocol for Metadata Harvesting*)
- Il sistema esegue l'operazione richiesta e fornisce in risposta il risultato derivato dall'interrogazione, codificato secondo il protocollo di comunicazione OAI-PMH
- Se nell'effettuare la richiesta si verificano degli errori, l'*OAI services* fornisce in risposta tutte le informazioni necessarie ad identificare correttamente il problema riscontrato

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



## 4. ARCHITETTURA DI SISTEMA

In questo capitolo analizzeremo l'architettura generale del sistema software MAG-Teca e verrà fornita una descrizione del *Deployment Diagram* relativo al software in oggetto, corredato di una prima descrizione dei singoli componenti.

Il *Deployment Diagram* (letteralmente “diagramma di dispiegamento”) è un diagramma di tipo statico previsto dal linguaggio di modellazione *object-oriented* UML per descrivere un sistema in termini di risorse hardware (detti “nodi”) e di componenti, mettendo in evidenza le relazioni che sussistono fra i singoli nodi e fra i relativi componenti software.

Nello sviluppo di MAG-Teca sono state adottate soluzioni software ed architetturali basate su prodotti *open source*. In particolare, sono state operate scelte mirate all'utilizzo di prodotti software, di base e d'ambiente, tali da rispondere ai requisiti attesi in termini di prestazione, sicurezza, affidabilità.

Alla base della progettazione dell'architettura di sistema vi è il concetto di servizio, che rappresenta l'elemento strutturale su cui le varie componenti software sono state sviluppate. In particolare, l'accesso ai dati avviene mediante un unico strato di interfacciamento, denominato *Business Logic*, secondo criteri orientati alla riusabilità delle componenti applicative.

L'intero software applicativo è stato sviluppato in modo da poter essere eseguito su di un *servlet container*. Tale scelta garantisce l'utilizzo di funzionalità avanzate come il *connection pooling* e la scalabilità del sistema, senza dover necessariamente farsi carico dell'investimento di alcun *application server* propriamente detto.

Dal momento che le funzioni di importazione ed esportazione dati possono richiedere tempi lunghi di esecuzione (a causa del tempo reale necessario allo spostamento di grosse quantità di dati), tali funzionalità sono state realizzate tramite applicativi lato server, attivati in modalità *on-demand* dall'amministratore, ed eseguiti dal sistema in modo autonomo, per evitare di caricare o inficiare ulteriormente le attività amministrative del

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



sistema. Per uniformità con il resto del sistema anche le funzionalità di importazione e di esportazione dati sono sviluppate in Java2SE e condividono gli stessi metodi per l'accesso e la manipolazione dei dati. Tale approccio consente di avere una sola *business logic* condivisa tra le operazioni interattive (interfaccia *web*, *SOAP*) e le operazioni non interattive (importazione/esportazione dati), con evidente riuso di codice e robustezza nella manutenzione dello stesso.

Il sistema di Teca ad oggetti, nata per una gestione analitica dell'informazione mediante MAG esteso, presenta un'architettura che abbiamo schematizzato nella figura sottostante mettendo in evidenza le componenti che hanno subito una maggiore evoluzione, ovvero: il sistema di analisi del MAG esteso per la deframmentazione dell'informazione contenuta nella sezione <sem>, il modulo di indicizzazione che agisce anche sugli oggetti *Contlet* e il *web services* con funzionalità aggiuntive per il dialogo con sistemi esterni.

La parte interattiva (*web* e *SOAP*) utilizza per l'accesso ai dati un *connection pool*, per ottimizzare le connessioni al *database* e, al tempo stesso, assicurare un elevato numero di accessi da parte dei *client*.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------

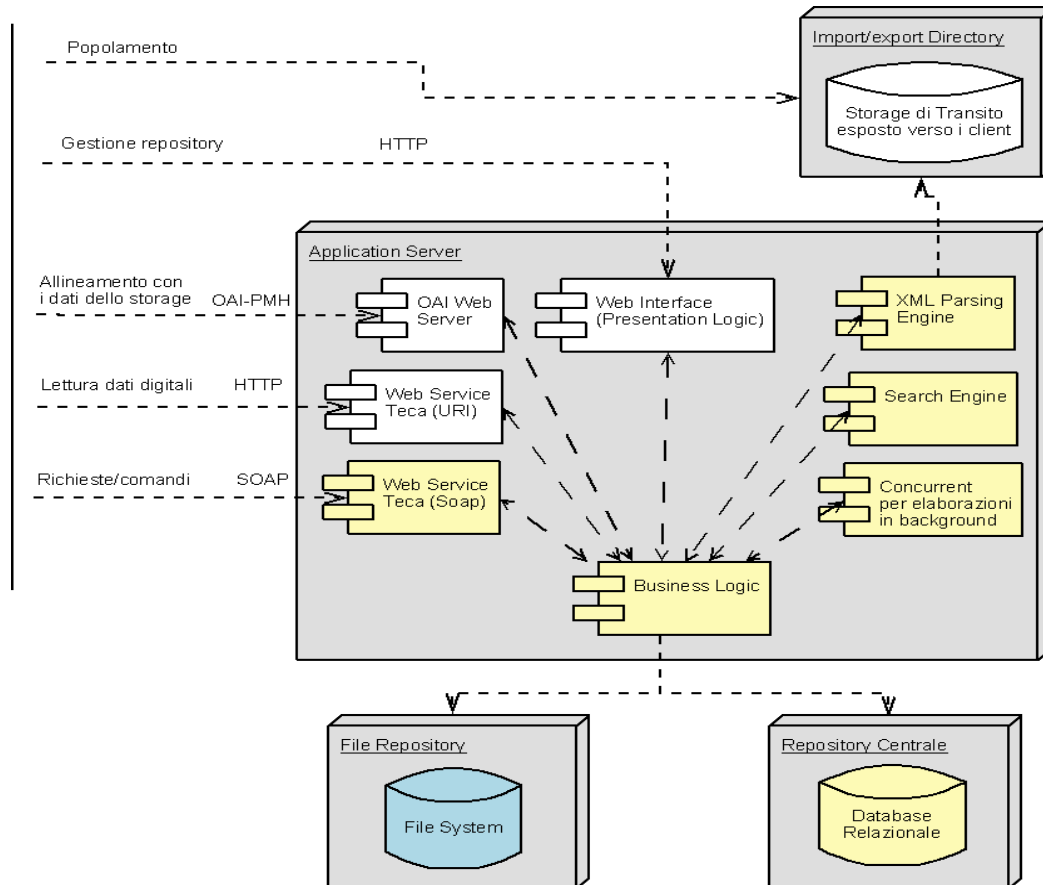


Illustrazione 4: MAG Repository – diagramma di deployment

La MAG-Teca ingloba al suo interno anche un *server* OAI (*Open Archive Initiative*) che consente l'accesso da parte dei servizi di *harvesting*. Tale motore è basato su una versione personalizzata del motore OAI-Cat. I dati XML forniti sono memorizzati su un'apposita *directory* del *FileSystem* del *server* e vengono generati al momento dell'importazione di un volume. La soluzione adottata nell'implementazione del modulo OAI consente di separare il *server* OAI stesso, dal *server* Teca propriamente detto. L'unico "vincolo" imposto è di avere una *directory* comune tra i due *server*, con permessi di scrittura per il primo e di lettura per il secondo.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



Come è possibile vedere dall'illustrazione 3, *MAG-Teca* è suddiviso sostanzialmente in otto componenti software. Di questi, la **business logic** riveste un ruolo centrale, rappresentando il cuore vero e proprio dell'intera applicazione, con gli altri componenti che interagiscono fra loro passando tramite la *business logic* con specifiche chiamate di sistema.

La *business logic* a sua volta “poggia” su un substrato costituito dalle librerie Java e Tomcat (per la gestione specifica dei *web services*).

La **presentation logic** è il componente che gestisce l'interfaccia utente per l'interazione fra amministratore di sistema, amministratore dell'archivio e operatori. L'interfaccia utente prescelta è un'interfaccia web basata sul linguaggio di presentazione HTML, che comunica su web mediante i protocolli TCP/IP e HTTP. Tale scelta è giustificata dall'esigenza di fornire un'ampia flessibilità al sistema, potendo essere visualizzata su qualunque sistema operativo (MS Windows, così come Linux o MAC Os) mediante un qualunque *browser* web grafico. Inoltre, l'adozione di un'interfaccia utente, il cui paradigma di interazione risulta attualmente essere uno dei più diffusi ed impiegati, riduce le difficoltà di impiego da parte degli utenti chiamati ad interagire con il sistema software.

Per l'accesso alle informazioni il modulo effettua opportune chiamate dirette sul componente della *business logic* attraverso il quale tutte le richieste devono passare.

Il modulo **xml parsing engine** si incarica di analizzare il file xml descrittivo di un oggetto e di estrarne tutti i metadati associati. Le informazioni così estratte vengono passate alla *business logic* che provvederà ad inserirle all'interno del *database* PostgreSQL.

L'*xml parsing engine* supporta inoltre le operazioni per l'acquisizione ed il trasferimento dati dalla *directory* di transizione del sistema all'interno del *repository* del sistema e viceversa. Tale scelta permette di specializzare il modulo qui descritto, evitando di esporre metodi specifici per l'interazione verso l'esterno.

I **web services (URI, SOAP)** costituiscono l'*entry point* per applicazioni esterne al sistema *MAG-Teca*. La comunicazione verso il *web services (URI)* avviene attraverso un canale

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



comunicativo basato sui protocolli HTTP e permette di accedere ad un risorsa digitale a partire da un indirizzo. Viceversa, per accedere alle funzionalità esposte dal *repository* i messaggi inviati al sistema ed in risposta da esso sono codificati mediante il protocollo SOAP.

Le funzionalità offerte dal *web service* sono esposte verso l'esterno tramite un *subset* di API. Queste consentono l'accesso ad una qualunque applicazione esterna, in modo semplice ed indipendente dai dettagli implementativi interni (quali il linguaggio implementativo utilizzato, o il *database* adottato). Tale scelta risulta, dunque, essere premiante in un'ottica a medio e lungo termine dello sviluppo dell'applicazione.

Il *web services (SOAP)* è stato progettato principalmente per permettere all'altro sistema software nell'ambito del progetto LC3, di annotare semanticamente le risorse digitali e di ricercare e accedere ai dati contenuti all'interno del *repository*.

Tramite il **Search Engine** vengono gestite le richieste SOAP relative alle ricerche sui metadati MAG e sui *Contlet* presenti nel sistema.

Il modulo **gestione concurrent** si occupa di tutte le operazioni *batch* previste nel sistema e che riguardano principalmente quelle attività che richiedono molto tempo per essere eseguite, perché coinvolgono l'elaborazione di file anche di grosse dimensioni.

L'ultimo componente, propriamente detto, è l'**OAI services**, che permette ai Robot OAI di accedere, per l'acquisizione, ai dati indicizzati contenuti all'interno degli archivi memorizzati dal sistema MAG-Teca. La comunicazione verso l'*OAI services* avviene tramite il canale comunicativo basato sui protocolli TCP/IP. I messaggi inviati al sistema ed in risposta da esso sono codificati mediante il protocollo OAI-PMH.

Gli ultimi tre componenti che completano l'architettura del sistema MAG-Teca sono il *reporitory* di transito, il *repository file system* ed il *database repository*.

Il **repository di transito** è un'area di appoggio temporanea per i metadati e le risorse digitali che devono essere importati nella MAG-Teca.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



Il **repository file system** è quell'area deputata dal sistema software a memorizzare e mantenere inalterati i dati ed il file xml con i metadati associati a ciascun oggetto importato all'interno dell'archivio.

Il **database repository** è, invece, il *database* destinato a raccogliere ed indicizzare i metadati estratti dai file xml associati agli oggetti importati all'interno degli archivi. La scelta di mantenere i metadati all'interno di un *database*, piuttosto che in un file *hash* o un normale file di testo, è dettata da specifiche di efficienza nell'effettuare interrogazioni sui dati; tale esigenza è tanto più evidente, essendo gli archivi librari formati da una notevole quantità di oggetti.

Il *database* prescelto è PostgreSQL. Tale scelta è stata dettata principalmente da due fattori: economico e tecnico. PostgreSQL è un *database open source*, che implementa soluzioni e funzioni tipiche dei moderni *database* relazionali. Confrontato con altri *database* relazionali quali Microsoft SQL Server e Oracle, fornisce prestazioni comparabili, con l'indubbio vantaggio di non richiedere né costi di acquisizione né tanto meno costi di aggiornamento e di supporto nel breve e lungo periodo. A confronto con altre soluzioni *open source*, osserviamo come MySQL non implementi ancora *full join* e *constraint* sui dati, oltre a mostrare limiti di affidabilità nei tempi di risposta in ambienti caratterizzati da un numero elevato di accessi concorrenti (soprattutto in scrittura), ovvero soffra di scarsa scalabilità all'aumentare del carico, e in presenza di molte scritture.

Di seguito, vengono descritti i sottosistemi in cui si articola il sistema MAG-Teca, esaminati sinteticamente in base agli scopi che si prefiggono:

- **Storage di transito:** ha lo scopo di fornire un'area di appoggio in cui depositare i dati digitali prima di procedere alla loro effettiva importazione. Tale spazio deve essere reso accessibile sia all'editore dei MAG che all'annotatore semantico, che avendo a disposizione un'apposita area dedicata, potranno utilizzare direttamente quella funzione. Al termine della fase di importazione i dati possono essere rimossi

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



da tale area, in quanto memorizzati in modo permanente all'interno dello *storage* della Teca.

- **Web Interface:** è l'interfaccia con cui l'editore dei MAG e l'amministratore operano per la gestione del sistema MAG-Teca; compito della *presentation logic* è consentire la visualizzazione di finestre contenenti i comandi eseguibili, gestire il controllo sull'immissione dei dati, trasferire le richieste verso la *business logic* per l'esecuzione dei comandi attivati e gestire i messaggi di ritorno da parte del sistema.
- **Xml parsing engine:** si incarica di analizzare il file xml in formato MAG esteso e di estrarne tutti i metadati associati. Le informazioni così ricavate vengono passate alla *business logic*, che provvederà ad inserirle all'interno del *database* relazionale. L'*Xml parsing engine* supporta, inoltre, le operazioni per l'acquisizione ed il trasferimento dati dalla *directory* di transizione del sistema all'interno del *repository* del sistema e viceversa. Tale scelta permette di specializzare il modulo qui descritto, evitando di esporre metodi specifici per l'interazione verso l'esterno.
- **Business logic:** implementa tutte le funzionalità necessarie al funzionamento del *repository*, supportando la comunicazione con i moduli del livello superiore (*presentation logic*, *XML parsing engine*, *web services* etc.) e quelli del livello inferiore (*repository* dei metadati e dei dati). La *business logic* ha l'obiettivo di raccogliere in un unico componente software tutto ciò che riguarda la manipolazione delle informazioni a basso livello. Tale modulo accede fisicamente sia al *filesystem*, che alla base dati, ed espone verso i livelli superiori metodi in grado di manipolare le informazioni senza dover conoscere il dettaglio della base dati o come i singoli dati multimediali siano archiviati fisicamente.
- **Concurrent per elaborazioni in background:** la fase di importazione dei file MAG coinvolge anche i file digitali e, quindi, può necessitare di molto tempo per essere eseguita. Per questo motivo, si è preferito utilizzare un sistema basato su "batch di

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



elaborazione” per non impegnare troppo l'*application server* e permettere una certa scalabilità, potendo inserire questo motore su apposito server dedicato. Tale componente viene gestita tramite interfaccia Web (usata dagli editori di MAG), ma anche tramite un set di *web-service* appositamente prodotti a tale scopo.

- **OAI Web Server:** tale componente consente il reperimento dei dati digitali a partire da sistemi esterni dotati di un *OAI-client* attraverso l'operazione di *harvesting*, come avviene ad esempio con il Network Turistico Culturale. Dal momento che espone alcune funzionalità di *discovery*, utilizzando un'interfaccia semplice e già collaudata, è stato scelto come metodo utilizzato dagli altri componenti del sistema per conoscere e reperire i nuovi dati digitali inseriti all'interno della Teca digitale o l'aggiornamento dei dati già presenti.
- **Web Service Teca:** tramite questo componente, da un lato vengono restituiti i dati digitali presenti nel sistema Teca, dall'altro vengono eseguite *query* complesse sui dati XML dei MAG. Tali *query* operano essenzialmente sui dati del Dublin Core e, quindi, non sono pensate per interrogazioni semantiche, per le quali è stato progettato apposito componente specifico. Inoltre, tramite tale componente, è possibile controllare l'esecuzione dei *concurrent* (processi concorrenti) per l'importazione dei MAG da parte dell'annotatore semantico.
- **Search engine:** questo componente ha lo scopo di eseguire le ricerche nel MAG esteso che viene anche decomposto nei vari *Contlet* estrapolando le annotazioni semantiche in modo da poter fornire anche una sola “frazione” di annotazione a chi ne fa richiesta, senza dover necessariamente trasmettere l'intero file MAG contenente l'annotazione.
- **Storage su File System:** i dati digitali vengono copiati dallo *storage* di transito in questa area. In questo modo, si ha il vantaggio di non consentire all'utente la rimozione accidentale dei *files* relativi ai dati digitali, in quanto questi vengono memorizzati in un'area a lui inaccessibile. Le operazioni di ripristino (che sono

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



eseguibili tramite il sistema dei *concurrent*) ricostruiranno la struttura originale dei file nello *storage* di transito. L'area di *storage* da utilizzare è un parametro modificabile: in questo modo, è possibile variare quest'area così da ottenere tutti i file prima di una certa data scelta come punto di riferimento raggruppati in una *directory* e sottoposta ad un unico *backup* (eventuali *backup* incrementali coinvolgeranno solo la nuova area che è stata definita). Dal momento che i file vengono rinominati durante l'operazione di importazione, è possibile memorizzare più volte un file avente lo stesso nome, in quanto comunque recuperabile dalla sua tracciabilità all'interno della Teca. Tramite questo meccanismo risulta semplificata la generazione di un sistema di *versioning* di dati digitali.

- **Database relazionale:** questo componente memorizza tutti i riferimenti ai dati digitali, gli utenti abilitati, eventuali autorizzazioni, i dati per l'indicizzazione. Il fatto di tenere i dati digitali separati su apposito *file system* consente al *database* di avere dimensioni ridotte e, quindi, semplifica anche la sua gestione e manutenzione, nonché le operazioni di *backup*.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



## 5. SPECIFICA DEI MODULI SOFTWARE

Nella presente sezione andremo a descrivere ciascuna delle componenti software del *MAG-Repository* indicando le principali funzionalità e le interfacce.

### 5.1. Business Logic

La *business logic* implementa tutte le funzionalità necessarie al funzionamento del *repository*, supportando la comunicazione con i moduli del livello superiore (*presentation logic*, *XML parsing engine*, *web services*, *OAI services*) e quelli del livello inferiore (*repository* dei metadati e dei dati).

La *business logic* ha l'obiettivo di raccogliere in un unico componente software tutto ciò che riguarda la manipolazione delle informazioni a basso livello. Tale modulo accede fisicamente sia al *filesystem*, che alla base-dati, ed espone verso i livelli superiori metodi in grado di manipolare la informazioni senza dover conoscere il dettaglio della base dati o di come i singoli dati multimediali siano archiviati fisicamente.

#### Funzionalità

- Query* di interrogazione del *database* per la ricerca, l'inserimento, la modifica e la cancellazione di record all'interno del *database* dell'applicazione
- Copia/rimozione di file dati (associati ad un oggetto) nel *repository* interno
- Esportazione dati codificati secondo il protocollo SOAP per l'esposizione verso l'esterno tramite l'applicazione *web services*
- Esportazione dati codificati secondo il protocollo OAI-PMH per l'esposizione verso l'esterno tramite l'applicazione *OAI services*

La *business logic* utilizza librerie standard dell'ambiente Java (in particolare J2SE 5), oltre ai driver JDBC per l'accesso alla base-dati. Al fine di ottimizzare i collegamenti alla base

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



dati viene utilizzata una libreria per la gestione del *connection pool* integrata con Tomcat, mediante l'uso di risorse JNDI.

Dal momento che la *business logic* può essere richiamata sia dall'ambiente dell'*application server* che dai programmi lanciati direttamente sul *server*, la libreria può usare sia risorse JNDI (all'interno di Tomcat) sia tradizionali connessioni dirette JDBC.

### **Interfaccia**

Non vengono adoperati specifici protocolli di interscambio dati. Per le modalità di interrogazione dei metodi e delle funzioni Java si rimanda al codice sorgente dell'applicazione. Viene utilizzato SQL per il collegamento con il *Database Server*.

## **5.2. Presentation Logic**

La *presentation logic* è il modulo funzionale deputato alla gestione dell'interfaccia utente attraverso la quale gli operatori, gli amministratori degli archivi e quello di sistema interagiscono con il sistema stesso. Compito della *presentation logic* è visualizzare le finestre con i comandi eseguibili, gestire il controllo sull'immissione dei dati, trasferire le richieste verso la *business logic* per l'esecuzione dei comandi attivati e gestire i messaggi di ritorno da parte del sistema.

### **Funzionalità**

Il componente viene usato per generare le maschere in HTML da visualizzare in un *browser*. Attraverso tale componente viene amministrato l'intero sistema Teca, dalla creazione degli utenti, alla gestione delle biblioteche.

Il componente utilizza le librerie di classi rese disponibili dalla *business logic*.

### **Interfaccia**

L'interfaccia utente è costituita da pagine HTML all'interno delle quali vengono visualizzati i dati richiesti dall'utente, e mostrati i comandi per l'interazione con il sistema.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



Le operazioni vengono attivate mediante la selezione di opportuni pulsanti inseriti all'interno di *form* HTML. Le operazioni di modifica o alterazione del repository interno vengono inviate mediante metodi POST, secondo quanto previsto dallo standard HTTP. Le richieste di acquisizione dati (visualizzazione) vengono inviate mediante metodi GET, secondo quanto previsto dallo standard HTTP.

L'adozione dello standard HTTP/HTML per la gestione dell'interfaccia utente assicura la comunicazione con il sistema anche da remoto e un'ampia possibilità di personalizzazione del layout dell'interfaccia grafica, potendosi adattare – ove richiesto – alle specifiche esigenze manifestate da un particolare ente.

I protocolli utilizzati sono i seguenti:

- TCP-IP per la comunicazione da e verso i *browser* degli utenti all'interno dei quali viene visualizzata l'interfaccia utente con le operazioni disponibili.
- Chiamate dirette alle classi Java implementate all'interno della *business logic*
- Viene utilizzato HTTP, per la comunicazione delle interfacce utente verso i *browser* web degli operatori del sistema.

### 5.3. XML Parsing Engine

L'*XML parsing engine* è il modulo funzionale dedicato alla gestione dell'*import/export* dei dati e dei metadati da e per il *repository* interno del sistema informatico MAG-Teca. Il modulo *XML parsing engine* supporta, inoltre, le funzionalità per l'analisi dei documenti XML all'interno dei quali sono codificati i metadati associati agli oggetti importati.

L'attivazione dell'*XML parsing engine* viene effettuata dal modulo funzionale *business logic*, a sua volta attivato da un utente attraverso il modulo *presentation logic*. Notiamo, infatti, come i moduli funzionali posti nello strato di più alto livello non comunichino mai fra

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



loro in modo diretto, ma lo facciamo sempre ed inderogabilmente attraverso il substrato inferiore costituito dalla *business logic*.

### **Funzionalità**

- Parsing* del file xml contenente i metadati descrittivi dell'oggetto da importare e la loro estrazione
- Verifica dell'integrità dei file da importare mediante algoritmo MD5
- Trasferimento dei file che caratterizzano l'oggetto da importare nel *repository* interno
- Trasferimento dei file che caratterizzano l'oggetto memorizzato nel *repository* interno nella directory temporanea di interscambio dei dati
- Ricerca di Contlet nel MAG e salvataggio come oggetto distinto

Il componente utilizza le librerie di classi rese disponibili dalla *business logic*. Algoritmo per la generazione della signature, MD5 (*Message Digest algorithm 5*).

### **Interfaccia**

Chiamate dirette alle classi Java implementate all'interno della *business logic*.

Vengono scambiati i seguenti dati:

- XML – MAG per la codifica dei metadati descrittivi degli oggetti, da importare all'interno del *repository* interno del sistema *MAG-Teca*.
- Formati multimediali generici (ad esempio AVI, mpeg etc. per i documenti video; mp3, wav etc. per i formati audio; txt, doc, ODG, PDF etc. per i formati testuali, e così via) per la codifica dei dati multimediali associati all'oggetto da importare.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



#### 5.4. **Web Services**

Il *web services* è il modulo funzionale destinato a gestire la comunicazione con applicazioni esterne facenti parte della famiglia di soluzioni LC3 (come Annotatori semantici), così come di applicazioni realizzate da terze parti.

Il *web services* impiega il protocollo SOAP per l'interscambio di dati. Le funzionalità implementate permettono di effettuare ricerche all'interno del *repository* della teca digitale, e di acquisire oggetti da impiegare in altri contesti, astraendosi dai dettagli implementativi interni al sistema MAG-Teca.

L'impiego di un *web services* per rendere accessibili gli archivi digitali contenuti all'interno del *repository* interno, fa del sistema software MAG-Teca un sistema veramente flessibile e aperto ad integrazioni fra più software realizzati da soggetti differenti.

#### **Funzionalità**

- Autenticazione delle credenziali dell'utente per l'accesso alle funzionalità rese disponibili dal *web services*
- Ricerca di oggetti mantenuti all'interno della teca digitale (la ricerca avviene specificando i criteri di ricerca codificati mediante una chiamata SOAP)
- Esecuzione di un concurrent
- Conoscere lo stato di esecuzione di un concurrent
- Leggere il log di un concurrent
- Importazione di un MAG contenente solo annotazioni passando direttamente il file XML.

Viene utilizzato *Tomcat Application Server* e Libreria Axis1.3 per la gestione del protocollo SOAP

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



### **Interfaccia**

Chiamate dirette alle classi Java implementate all'interno della *business logic*

Impiego di protocollo SOAP

### **5.5. OAI Services**

L'*OAI services* è il modulo funzionale destinato alla gestione della raccolta dati da parte degli *harvester* OAI.

L'*OAI services* impiega, naturalmente, il protocollo OAI-PMH per lo scambio dei *record*. Le funzionalità implementate permettono di acquisire le indicizzazioni dei metadati contenuti all'interno del *repository*, secondo quanto previsto dallo standard OAI.

### **Funzionalità**

Il modulo consente di esporre le informazioni sui volumi contenuti, utilizzando lo standard OAI per la ricerca e l'*harvesting* dei contenuti digitali.

Basato sul progetto OAICat con personalizzazioni specifiche per il sistema LC3.

### **Interfaccia**

Usa il protocollo HTTP, per la comunicazione verso gli *OAI Client*

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



## 5.6. Search engine

Questo componente ha lo scopo di eseguire le ricerche nel MAG e nei *Contlet* archiviati nel MAG Repository.

### Funzionalità

Trasforma i parametri di ricerca passati dalle funzioni SOAP o WEB in una richiesta SQL compatibile con il database server.

### Interfaccia

Utilizza il driver JDBC che comunica con il motore di database usando *socket* JCP/IP.

## 5.7. Motore di concurrent per elaborazioni in background

Il motore di *concurrent* opera su oggetti derivati dalla classe *concurrent* consentendo l'esecuzione di *routine* specifiche in *background*. La presenza di tale motore consente la coordinazione delle attività in *background* in modo da non sovraccaricare la macchina. Inoltre può essere messo in esecuzione su una macchina distinta da quella contenente il software MAG Teca.

### Funzionalità

Coordinamento ed esecuzione dei *concurrent*.

### Interfaccia

La comunicazione avviene tramite database e l'output viene riversato su *file* system.

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------



## 5.8. Repository Centrale

Il *repository* centrale si compone di due componenti distinti: il primo, composto dal *file system* è dedicato all'archiviazione dei dati e dei file xml contenenti i metadati associati ad un oggetto, acquisiti attraverso l'operazione di *import*. I dati sono organizzati internamente in sotto cartelle, ciascuna dedicata alla raccolta dei file associati ad un singolo oggetto.

Il secondo componente è costituito dal *database* relazionale PostgreSQL, che raccoglie al suo interno, in apposite tabelle, sia i dati destinati all'amministrazione del sistema (tabella utenti con i dati per l'accesso dei vari *account*), sia i metadati associati agli oggetti digitali contenuti all'interno della teca digitale. L'indicizzazione e la memorizzazione dei metadati all'interno di un *database* relazionale assicura un elevato livello di performance sia in fase di interrogazione dell'archivio che in fase di inserimento e modifica dei dati stessi.

### **Funzionalità**

- Query* di interrogazione del *database* relazionale per la ricerca, l'inserimento, la modifica e la cancellazione dei record.
- Creazione/rimozione, all'interno del *file system*, di *cartelle* destinate alla raccolta dei dati multimediali e del file xml contenente i metadati associati ad un oggetto digitale importato.
- Copia/rimozione di file all'interno di una cartella sul file system.*

Basato su *Database server PostgreSQL* utilizza la Libreria JDBC per l'accesso ai dati

### **Interfaccia**

- Libreria nativa per l'accesso ai dati (la libreria di accesso ai dati viene installata come Driver JDBC).
- Query SQL*

PROGETTO LC3	Revisione n*	6	Del	23/01/09
--------------	--------------	---	-----	----------