Un portale di accesso a riviste elettroniche multidisciplinari per l'Università e la Ricerca scientifica e tecnologica: l'esperienza del consorzio CASPUR con il suo servizio di *Emeroteca Virtuale*

Ugo Contino
ugo.contino@caspur.it
***CASPUR** – **C**onsorzio per le **A**pplicazioni del **S**upercalcolo **P**er **U**niversità e **R**icerca*
via dei Tizii, 6b – 00185 Roma

Sommario. In questo lavoro viene presentata l'esperienza maturata all'interno del CASPUR con il suo servizio di *Emeroteca Virtuale*, che, da circa un quadriennio, offre un servizio di accesso a testate elettroniche di tipo multidisciplinare, per una base di utenti che comprende una trentina tra università ed enti di ricerca. Vengono in particolare illustrate: le caratteristiche salienti dell'infrastruttura utilizzata e recentemente sottoposta ad una sostanziale rivisitazione, con l'obiettivo di migliorarne la qualità verso l'utente finale; l'innovativa soluzione infrastrutturale adottata per i server che fanno da interfaccia all'utenza; i parametri utilizzati per stimare la qualità dell'interfaccia WEB dell'Emeroteca, sulla base dei tempi medi di risposta del portale all'utente finale. Chiude l'articolo una rassegna delle attività *in fieri* che vedrà impegnato il CASPUR, da solo o con partner accademici o scientifici nell'ambito di questo progetto.

Introduzione

La crescente diffusione delle strutture che offrono servizi di *Digital Library* (in seguito identificati dall'acronimo D.L.) in un contesto universitario e di ricerca e l'adozione di soluzioni tecnologiche dirette ad una maggior integrazione delle risorse digitali disponibili nei centri che offrono questo tipo di servizi, è ormai una realtà consolidata, come dimostrato dalla presenza in rete di un ampia raccolta di articoli scientifici o di rassegna legati al contesto della biblioteca digitale[1].

Molteplici sono i centri universitari o di ricerca che offrono servizi di D.L. [2] o servizi connessi al mondo delle D.L. (per es. sistemi di *Archive & Indexing*), soprattutto in un contesto di *collezioni aperte* (archivi di pre-print; *scholarly e-journals* distribuiti gratuitamente; archivi *istituzionali*) in linea con quanto espresso dal manifesto dell'iniziativa di *Open-Archive*[3] (OAI).

In questo contesto l'Emeroteca Virtuale del CASPUR rappresenta un servizio di *Digital Library* per l'accesso ad un insieme di testate elettroniche multi-disciplinari pubblicate da sette editori specializzati nell'editoria dei periodici elettronici. Nel successivo paragrafo verrà tracciata una storia dell'Emeroteca Virtuale (di seguito denominata E.V.) dall'origine del servizio fino ad oggi, per poi passare successivamente alla discussione sulle disegno dell'E.V (in relazione alle particolari condizioni imposte dalla tipologia di utenza finale ed alle caratteristiche dei dati da trattare) e sulle misure di tipo *prestazionale* eseguite sulla nuova architettura.

Il servizio di Emeroteca Virtuale: una breve storia

Questo servizio si inquadra storicamente nel contesto delle attività iniziate alla fine degli anni 90 nell'ambito della collaborazione CIBER (Coordinamento Interuniversitario per le Basi di dati e l'Editoria in Rete) [4] , che ha visto coinvolti 5 atenei (il Politecnico di Bari e le Università di Bari, Lecce, Roma "La Sapienza" e Roma Tre) ed il consorzio CASPUR, con il quale questi atenei erano consorziati.

Scopo del CIBER è quello di[1]:

- cooperare per la costituzione e lo sviluppo di biblioteche digitali (informazioni scientifiche su matrice elettronica) nelle Istituzioni partecipanti e per promuovere la crescita professionale del loro personale tecnico-bibliotecario;
- facilitare l'acquisizione, da terzi, di servizi bibliografici e documentari in rete;
- agevolare il trasferimento di informazioni e di servizi tra le Istituzioni partecipanti;
- cooperare alla scelta e la definizione delle risorse da sviluppare, acquistare o affittare;
- identificare le tecnologie informatiche necessarie, al miglior rapporto costi/benefici;
- curare lo sviluppo di prodotti per la consultazione e la archiviazione dell'informazione scientifica in formato elettronico;
- richiedere alle autorità competenti, a nome delle Istituzioni partecipanti, il sostegno necessario per dare alle Biblioteche digitali le risorse di cui hanno bisogno per perseguire i loro fini.

Sin dall'inizio il CASPUR ha offerto, da un lato, un supporto consulenziale di tipo amministrativo e tecnico, per la promozione e la diffusione delle risorse digitali inerenti a specifiche basi dati[2] sia di tipo umanistico, che medico, scientifico e tecnologico, e dall'altro permettendo l'accesso a periodici elettronici accademico-scientifici, la maggior parte a *testo completo*, tramite il suo servizio di Emeroteca.

Relativamente a quest'ultimo servizio, nell'arco di poco meno di un anno a partire dal 1999, e a fronte di un accordo di acquisto consortile delle testate elettroniche distribuite dall'editore Elsevier, sono state caricate su un server dislocato al CASPUR più di 500 testate a *testo pieno*, disponibili dal 1995 in poi, per un totale di circa 300.000 articoli. La particolare forma di accordo commerciale scelta con l'editore permetteva l'accesso all'insieme delle testate possedute da tutti gli atenei partecipanti (*Cross-Access*), approccio successivamente superato dal contratto *big-deal*, nel quale l'accesso è consentito a tutti i periodici elettronici pubblicati dall'editore, sia per Elsevier che per gli altri editori che si sono successivamente aggiunti al servizio di Emeroteca.

La bontà della scelta consortile del CIBER (sempre aperto ad accettare l'adesione di nuovi atenei) nell'approccio con i vari editori di *scholarly e-journals* ed il concreto supporto dato dal CASPUR a quest'iniziativa, hanno fatto si che il numero di atenei che hanno risposto positivamente all'iniziativa sia andato progressivamente crescendo nel tempo come mostrato nel diagramma seguente, raggiungendo il numero attuale di 26 membri [8] (pari al 40% della popolazione universitaria italiana).

[1] Estratto dal regolamento del CIBER – Art. 1 – Denominazione e Scopi [5]

[2] Attività svolta in collaborazione con il Cordinamento SIBA [6] – Servizi Informatici Bibliotecari di Ateneo – dell'Università di Lecce e l'Università "La Sapienza" di Roma nell'ambito del progetto BIDS [7] – Progetto Biblioteca Digitale della "Sapienza"

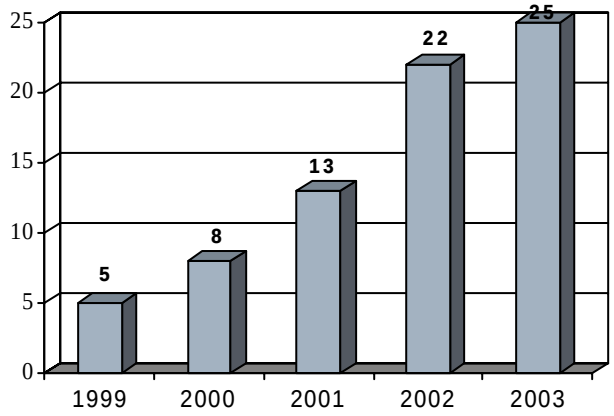


Fig.1 – Numero di atenei partecipanti al coordinamento CIBER (agg. Ottobre 2003)

Parallelamente si è avuto un sostanziale incremento del numero di riviste disponibili in *full-text* all'interno dell'Emeroteca Virtuale, sia ampliando il numero di editori coinvolti nelle trattative, sia incrementando il numero di riviste dei contratti preesistenti che andavano in scadenza.

Attualmente il numero di editori, con i quali sono stati sottoscritti contratti o sono in atto dei *trial*, è pari a sette, per un totale di più di 3500 riviste consultabili tramite l'Emeroteca (di cui più di 3300, pari al 93% del totale, residenti in locale), numero che rappresenta il 34% del totale delle pubblicazioni di tipo *scholarly* (tab. 1).

	n.ISSN	Disponibile dal	Tipo di accesso
Elsevier	1825	1995-	FT
Blackwell	590	1996/97	Metadata
Kluwer	760	1996/97-	FT
ACS	30	1879-	Metadata
IOPP	41	1875-	FT*
Wiley**	300	1996-97	sul sito dell'editore
Nature	16	1996/97-	sul sito dell'editore
Totale	**3566**		

(aggiornamento: 22 Agosto, 2003)
* dal 1991
** in Trial

Tab.1 – Numero di testate accessibili tramite il servizio di Emeroteca Virtuale

Nel grafico di fig. 2 è invece mostrato l'andamento delle nuove acquisizioni nel corso degli anni, andamento che mostra come sia ancora in atto un incremento del numero totale di riviste accessibili tramite l'E.V. sia in formato *full-text* che di soli meta-dati (*abstract*).

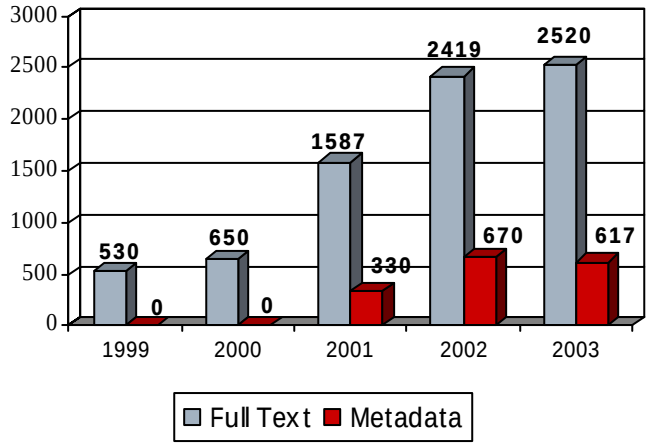
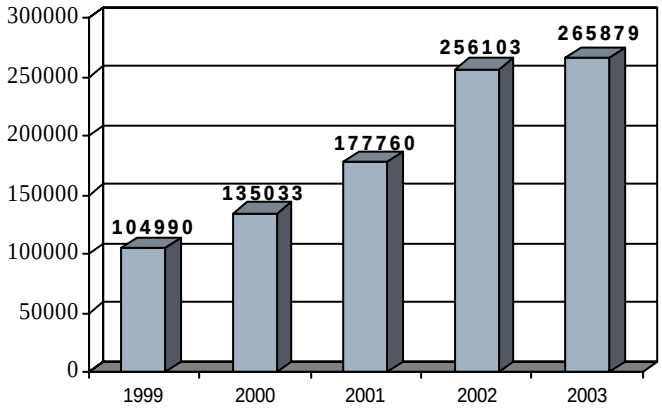


Fig.2 – Andamento del numero di riviste accessibili tramite l'E.V. (agg. Agosto 2003)

Nei diagrammi di fig. 3 è infine mostrato l'incremento degli utenti (potenziali) del servizio di Emeroteca rappresentato dal numero di studenti iscritti e dal numero del personale docente appartenente agli atenei del CIBER (fonte MIUR [9]).



Full Time Equivalent (dati '99)

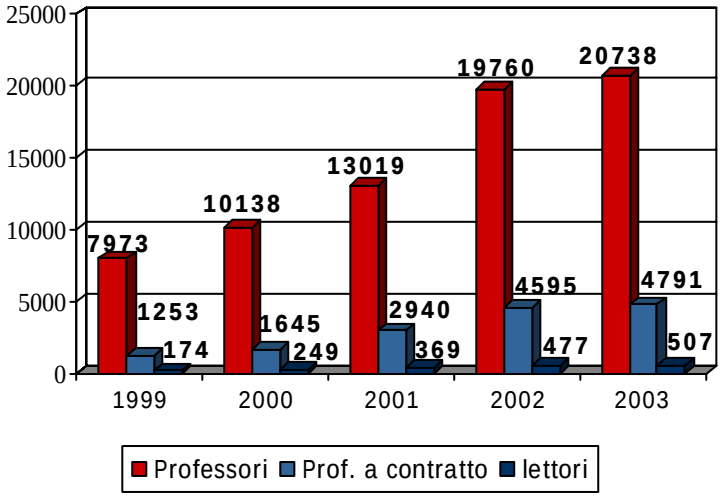


Fig.3 – Incremento del numero di utenti potenziali del servizio di l'E.V. (agg. Agosto 2003)

Risulta evidente come la necessità di fornire livelli nella qualità del servizio di fruizione dell'informazione disponibile nell'E.V. per una base utenti così vasta, mantenendo i costi dell'infrastruttura tecnologica ad un livello accettabile, sia stato sin dall'inizio un parametro fondamentale nel disegno adottato al CASPUR della soluzione di *Digital Library* .

Nel prossimo paragrafo verranno affrontate queste tematiche, chiarendo soprattutto quali siano stati i parametri hanno pesato nel definire le scelte tecniche operate sul servizio negli ultimi quattro anni.

Progettare una Digital Library: motivazioni e scelte

Nel disegnare la soluzione alla base del servizio dell'Emeroteca Virtuale del CASPUR hanno pesato motivazioni di tipo gestionale, scelte di tipo contrattuale e l'obbiettivo di rendere in breve tempo disponibile un servizio di *content management* certamente complesso.

La prima motivazione è stata dettata dalla necessità di fornire al coordinamento CIBER una struttura centralizzata per il *repository* degli articoli delle testate elettroniche in abbonamento e dei loro *metadati.* Il bisogno di garantire un *profilo di accesso* corretto alle varie testate dovuto al fatto non tutte le università hanno acquisito i medesimi diritti alla consultazione (solo all'inizio del servizio l'accesso alle risorse elettroniche è stato comune per tutti gli atenei), ha spinto verso la soluzione centralizzata di un unico server (dislocato al CASPUR) opposta rispetto a quella viceversa adottata per il servizio di accesso alle banche dati in abbonamento (che prevede una struttura di server distribuita su tre sedi). Questo disegno presenta infatti l'indubbio vantaggio di essere più facilmente gestibile per ciò che concerne il controllo del meccanismo di accesso (basato nello specifico caso dell'E.V. sulla classe IP di provenienza) e la corretta attribuzione del *profilo di consultazione*.

Relativamente alle scelte contrattuali derivanti dalla forma di abbonamento scelta dalle università che implicava non solo (e non tanto) l'accesso ai descrittori (*abstract*) degli articoli contenuti nelle varie testate elettroniche in abbonamento, quanto piuttosto la possibilità di avere in locale il testo completo dell'articolo stesso a testo pieno (*full-text*), acquisibile in formato PDF (secondo la terminologia, coniata dall'editore Elsevier, si parla in questi casi di accessi *SDOS, Science Direct On Site*, distinti da quelli *SDOL*, *Science Direct On Line*, per i quali il *full-text* risiede sul sito dell'editore), queste hanno non poco pesato sulla scelta fatta per l'infrastruttura di *storage* dell'Emeroteca. Il dover infatti offrire un servizio di *local repository* che fosse scalabile per ciò che concerne il numero e la mole dei dati memorizzati ha portato, sin quasi dalla genesi del servizio, verso la scelta di soluzioni di memoria di massa *esterne* all'infrastruttura rappresentata dal server (si parla in questi casi di *SAN*, *Storage Area Network*), composte da *array* di dischi a bassa latenza, possibilmente in tecnologia SCSI, dotate da opportuni meccanismi di *fault-tolerance* (RAID) e connesse al server utilizzando uno standard aperto ed interoperabile tra *vendor* diversi. Le stesse scelte contrattuali hanno anche pesato sulla tipologia del servizio offerto: il servizio di E.V. è riservato alle sole università che hanno sottoscritto abbonamenti con l'editore dal momento che l'accesso è verso riviste la cui proprietà è dell'editore e degli enti che ne hanno acquisito i diritti. Da questo punto di vista l'E.V. potrebbe essere classificata come una *closed-digital-library*, a differenza di un'altra tipologia di sistemi (*open-digital-library*) che nell'ambito OAI, danno accesso gratuito a collezioni di articoli in forma di *pre-print* ovvero gestite da strutture editoriali scientifiche e di tipo non commerciale (ad esempio centri di ricerca, strutture *no-profit* o università).

Per ciò che concerne l'ultimo punto c'è da osservare che la necessità di offrire un servizio di *content management* che avesse caratteristiche professionali per l'utenza dell'Emeroteca e questo cercando per quanto possibile di ridurre i tempi di messa a punto del sistema, ha portato alla scelta di un software commerciale, scelta per altro resa più semplice dal numero sempre più crescente di società specializzate nel settore dell'*information management*. Per l'Emeroteca si utilizza il software *Science Server* della Endeavor Information Systems, sussidiaria della società Elsevier

Science[10]. Questa scelta non ha tuttavia interrotto i processi di personalizzazione sia dell'interfaccia utente, che dei meccanismi di gestione delle metodologie di accesso, dal momento che il *look-and-feel* del portale di accesso dell'Emeroteca ha già subito due profondi restyling e che (come accennato più avanti) è allo studio un progetto per affiancare al motore di ricerca proprietario, un nuovo motore basato su approcci innovativi che fanno uso di reti neurali[11].

Nei prossimi paragrafi si cercherà di scendere maggiormente in dettaglio su quelle che sono le caratteristiche tecniche e funzionali dell'attuale servizio di E.V., mostrando come esso abbia acquisito col tempo una struttura più articolata, che, partendo dall'installazione di un singolo server corredato del software di gestione ed accesso alla *digital library*, vede attualmente un insieme, fortemente sinergico, di più sistemi informatici. Possiamo aggiungere a margine delle motivazioni esposte in questo paragrafo che, la gestione di un *oggetto* complesso quale può essere rappresentato nel nostro caso dalla E.V., si è dovuta confrontare con una gestione tecnica e amministrativa, parimenti complessa, di un coordinamento interuniversitario oggigiorno composto da 26 elementi. Sono proprio questi due aspetti che hanno contribuito di pari grado all'aspetto fortemente articolato con il quale il servizio, nel suo complesso, si presenta.

L'Emeroteca Virtuale del CASPUR: di cosa stiamo parlando

In questo paragrafo si vuole descrivere in maggior dettaglio quali elementi funzionali contribuiscono all'erogazione di questo servizio, avendo cura di illustrare i ruoli che ciascun elemento ha relativamente a tre tipologie di utenza:

- l'utente remoto che (tramite il portale dell'E.V.) accede al servizio di consultazione
- gli *electronic resources librarians* che, per ciascuno dei 26 atenei o centri di ricerca partecipanti, curano la gestione biblioteconomica del posseduto verso i propri utenti e verso il CASPUR
- gli amministratori del sistema (personale CASPUR)

Al fine di garantire una migliore comprensione di cosa si intenda realmente quando si parla del servizio di Emeroteca Virtuale del CASPUR, si può far riferimento alla visione schematica mostrata in fig. 4.

La *Digital Object Repository*

E' la struttura che contiene gli articoli in *full-text*, sia in formato PDF, che HTML, insieme ai loro descrittori (metadati), ovvero solo i metadati nei casi di alcuni editori (Blackwell-Synergy, ACS). Fa parte di questa struttura anche la gerarchia dei descrittori XML utilizzati dal *Browsing Engine* e l'*Indexing Data Base* utilizzato dal *Searching Engine*. E' l'elemento funzionale *centrale* del servizio di Emeroteca ed è per questo quello su cui si concentra maggiormente l'attività di *back-up* (descritta più avanti).

Il *Browsing Engine*

A questo blocco funzionale è demandato il compito di permettere all'utente il *browsing* all'interno delle riviste dell'Emeroteca. Tipicamente queste funzioni vengono gestiste da particolari programmi (CGI) eseguiti sul server che eroga il servizio di *digital library*; questi programmi interagendo con i descrittori XML che raccolgono i titoli delle testate in ordine alfabetico, ovvero per editore, ovvero

infine per categoria di pubblicazione, ne traslano il contenuto in formato HTML. Nel caso dell'E.V., tuttavia, al fine di ottimizzare i tempi di accesso a liste contenenti in certi casi anche più di un migliaio di testate, si è deciso di utilizzare liste statiche, composte da file già in formato HTML. Questa scelta non ha pesato sulle attività di aggiornamento delle liste, dal momento che queste vengono svolte periodicamente con cadenza giornaliera da specifiche procedure automatiche.

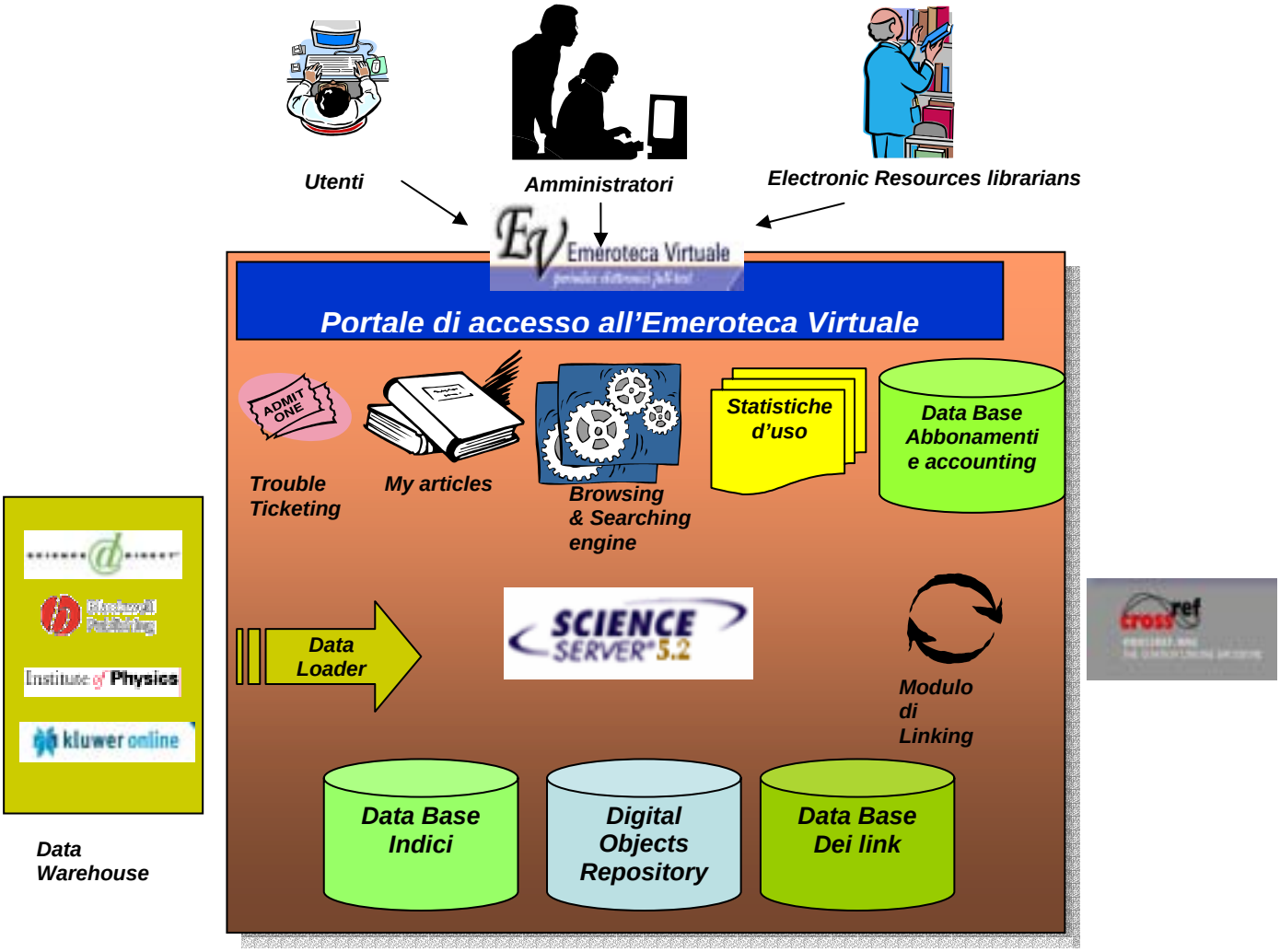


Fig.4 – Quadro di insieme del servizio di Emeroteca Virtuale

Il *Searching Engine*

Come intuibile, questo blocco funzionale ha lo scopo di permettere la ricerca all'interno trai i *digital objects* presenti nella *Repository*. Il meccanismo che è alla base dell'attuale versione del *Searching Engine* è di tipo *full-text* (basato sul *pattern-matching*), ed include l'uso dei classici operatori Booleani (AND, OR, NOT), oltre che di particolari operatori *di prossimità nel contesto*.

L'attuale software impiegato per la gestione di questo servizio prevede diverse tipologie di ricerca:

- ricerca semplice
- ricerca avanzata
- ricerca esperta
- ricerca automatica

In particolare l'ultimo tipo di ricerca permette ad una determinata tipologia di utenti (quelli che si *registrano* nell'Emeroteca) di accedere ad una funzione che automaticamente e periodicamente esegue delle ricerche per conto dell'utente, utilizzando *chiavi di ricerca* da lui impostate, segnalando eventuali nuovi risultati tramite *e-mail* all'utente (ciò spiega la necessità della registrazione). Questo servizio è pressoché uno standard tra i software di *digital library* commerciali e non commerciali (*articles alerting*).

Il blocco *My Articles*

Solo per gli utenti che si registrano al servizio, è disponibile un modulo che permette di accedere non solo alle ricerche salvate, permettendone la loro modifica, esecuzione o rimozione, ma anche ad un area nella quale l'utente ha salvato i puntatori (*handles*) agli articoli di suo maggior interesse. E' così possibile per lui avere una propria area riservata, accessibile sul sito, dove conservare questo tipo di dati, senza doverli scaricare in locale sul proprio computer.

Il sistema di *Trouble Ticketing*

Il dover interagire con una utenza remota che è andata negli anni sempre più crescendo, la necessità di definire una metodologia ottimizzata che permettesse la corretta ripartizione dei compiti all'interno del gruppo degli amministratori che danno supporto su problematiche inerenti al servizio, oltre che la completa diffusione interna delle informazioni fornite come supporto all'utenza e la costituzione di un archivio dei *problemi risolti*, hanno portato il CASPUR ad integrare, all'interno del servizio di E.V., un meccanismo di *trouble ticketing*. Questo modulo, sviluppato mediante l'uso di tecnologia *open-source* e basato su un DB MySQL[12] con interfaccia scritta nel linguaggio PHP[13], permette all'utente, che necessiti di supporto, di accedere ad una specifica pagina WEB, *modulo di segnalazione errori*, dove può indicare:

- la rivista, il volume ed il fascicolo che presentano problemi di consultazione
- l'editore della rivista
- un campo note dove fornire ulteriori informazioni

La sottomissione di una segnalazione (*ticket*) viene notificata al supporto dell'E.V. tramite un'e-mail (*e-mail alerting*) nella quale è contenuto il *link* al servizio di gestione dei *ticket* anch'esso basato su WEB. Dall'introduzione del servizio di *trouble ticketing*, che ha sostituito nel gennaio del 2003 il precedente approccio basato su e-mail, sono state registrate un centinaio circa di chiamate (novembre 2003). Il tempo medio di risposta ad una chiamata è di circa un giorno; la percentuale delle chiamate che vengono risolte nella prima risposta è del 70 %.

L' Holdings & Accounting Data Base

Mentre il modulo precedente è dedicato agli utenti finali, questo modulo è stato appositamente progettato per gli *electronic resources librarians*, per gli amministratori CASPUR dell'E.V. e per il personale amministrativo (sempre del CASPUR) che gestisce i contratti consortili verso gli editori. Lo scopo di questo Data Base è di raccogliere e rendere disponibili informazioni, pertinenti al servizio, per ciascuna università che accede all'Emeroteca, quali:

- i contatti tecnici e scientifici
- le testate possedute (*holdings*) di ogni ateneo suddivise per anno e per editore
- le classi di indirizzi IP di appartenenza
- la situazioni contabile per i contratti collettivi sottoscritti dall'ateneo

Questo modulo, sviluppato internamente e, anch'esso, con tecnologie *open-source*, verrà ufficialmente rilasciato all'utenza nel prossimo mese di dicembre. Attualmente sono raccolte informazioni sulle *holdings* delle sole riviste dell'editore Elsevier. A regime verranno raccolte le collezioni, possedute dai vari atenei, per gli altri editori accessibili tramite l'Emeroteca.

Il modulo di accesso alle statistiche d'uso

Questo modulo permette, solo per gli utenti autorizzati (gli *electronic resources librarians*), di accedere ad un portale dove possono impostare i parametri di analisi dei file di *log* nei quali vengono registrati gli accessi all'E.V. Dopo aver impostato questi parametri (periodo di analisi; università di appartenenza; *holdings* di riferimento), il sistema di *back-end* (basato sul software di analisi statistiche SAS[14]) provvede ad inoltrare all'indirizzo di e-mail specificato dall'utente, un *report* PDF, particolarmente dettagliato, nel quale sono contenute le seguenti informazioni:

- cumulative sugli accessi ai contenuti (riviste, fascicoli, *abstract* e *full-text*)
- cumulative sugli accessi agli editori
- dettagliate sugli accessi alle singole testate
- distribuzioni *top-n* delle riviste consultate

Il modulo di analisi statistica, attivato nel corso del secondo semestre del 2003, è in grado di gestire la generazione di quattro *report* contemporanei. Il tempo medio di generazione di un rapporto si aggira intorno ai 20 minuti.

Modulo di caricamento automatico delle riviste (*Automatic Data Loader*)

I dati inviati dai vari editori sono raccolti in *cluster* di file chiamati *dataset*. Ogni *dataset* contiene gli aggiornamenti di una o più testate pubblicate sul sito. Il metodo di aggiornamento delle riviste offerto dagli editori ha, da un paio di anni circa, sostituito quello tradizionale della distribuzione su CD, per affidarsi ad un meccanismo di *on-line delivery* basato su procedure FTP, coadiuvato da un meccanismo di "*new dataset available" alerting* via e-mail. Questo ci ha permesso di sviluppare all'interno del *framework* dell'E.V., il modulo di *Data(set) Loader* automatico; questo processo su base giornaliera:

- verifica la presenza messaggi di *alerting* ricevuti nelle ultime 24 ore

- si connette via FTP ai siti (*dataware house*) dove sono contenuti i *dataset* da acquisire
- procede all'acquisizione dei dati, archiviando successivamente i *digital objects* nelle directory suddivise per editore
- nomina le directory con i *dataset* utilizzando uno schema basato sul giorno di caricamento (ad eccezione dell'editore Elsevier, che ha un proprio schema di identificazione dei *dataset*)
- effettua il *back-up* automatico dei dati acquisiti
- esegue gli script (*loader*) per la generazione dei descrittori XML e degli *handle* locali per gli articoli dell'Emeroteca
- procede con l'indicizzazione delle riviste e degli articoli per aggiornare il DB del motore di ricerca
- analizza i *log* file prodotti dal processo di caricamento e produce dei *report* sintetici inviati via e-mail agli amministratori del sistema

Questo modulo (sviluppato in linguaggio PERL[15]) è stato introdotto per semplificare procedure gestionali pertinenti all'aggiornamento dell'Emeroteca, in precedenza svolte o manualmente o in maniera semiautomatica. Attualmente il gestore del sistema interviene solo quando viene riportata un'anomalia dal modulo di caricamento automatico.

Il Modulo di *Linking*

Con l'ultima versione di Science Server (5.2) è stata introdotta una funzionalità di *linking* alle citazioni che permette all'utente di avere disponibile (quando previsto) un *link* agli articoli che citano o che sono citati dall'articolo scaricato, presenti o sulla stessa E.V., ovvero sul sito dell'editore. Questo meccanismo si basa sul descrittore DOI[16] per l'identificazione univoca degli oggetti digitali e fa uso dello standard OpenURL[17] per la risoluzione dei *link* da *reference archive sites* (*Cross-Ref*[18]). Il modulo di *linking*, parte della distribuzione standard di Science Server, procede ad aggiornare i riferimenti alle citazioni (salvandoli poi in uno specifico Data Base) distinguendo tra:

- *link interni* nel caso in cui l'articolo referenziato (o che referenzia) sia già presente nell'E.V.
- *link esterni* verso siti di editori o archivi aperti (nel caso di *pre-prints*) se l'articolo da cercare non è presente nell'emeroteca

C'è da osservare come questo modulo, benché introduca a nuove funzionalità di *information retrieval*, ha ancora dei limiti legati al fatto che la risoluzione nei link interni avvenga per il solo editore Elsevier, e che poca flessibilità sia data al meccanismo di ricerca delle risorse *collegate* alla citazione ed alla caratterizzazione dell'utente, flessibilità invece offerta dallo standard OpenURL. Per questo motivo è allo studio una soluzione che implementi pienamente la tecnologia del *context-sensitive-linking*; attualmente si stanno considerando due possibilità: la scelta di un software di tipo commerciale (come fatto per Science Server); la scelta di una soluzione *semi-proprietaria*, ovvero sviluppata in collaborazione con strutture universitarie e partner commerciali.

Il portale di accesso all'Emeroteca Virtuale

Costituendo uno dei punti di accesso *privilegiati* al servizio di *digital library* offerto dal CASPUR, questo modulo rappresenta il perno centrale dell'interfaccia utente. Abbandonata l'idea di utilizzare la pagina di *default* disponibile con il software Science Server, ci si è concentrati su una soluzione che fosse maggiormente rispondente ai criteri di usabilità e personalizzazione per le utenze dell'Emeroteca. La versione corrente dell'home page di accesso all'Emeroteca è mostrata nella figura seguente, così come si presenterebbe ad un utente autorizzato proveniente da uno degli IP del CASPUR.

Facendo riferimento alla fig.5, è possibile suddividere la pagina di accesso all'E.V. in più aree:

- un'*intestazione*, nella quale sono raccolte le informazioni pertinenti al servizio, ed i link per l'accesso alle pagine riservate
- un settore relativo alla *ricerca* nel sito dell'Emeroteca ed alla *registrazione dell'utente*
- una parte relativa alla navigazione (*browsing*) nelle riviste contenute all'interno dell'Emeroteca suddivise per editore
- una relativa agli accessi sui siti degli editori per la consultazione *off-site* (dove previsto dalle norme contrattuali)
- un'area specifica per ogni ente (*area ente*) dove poter inserire informazioni di pertinenza locale

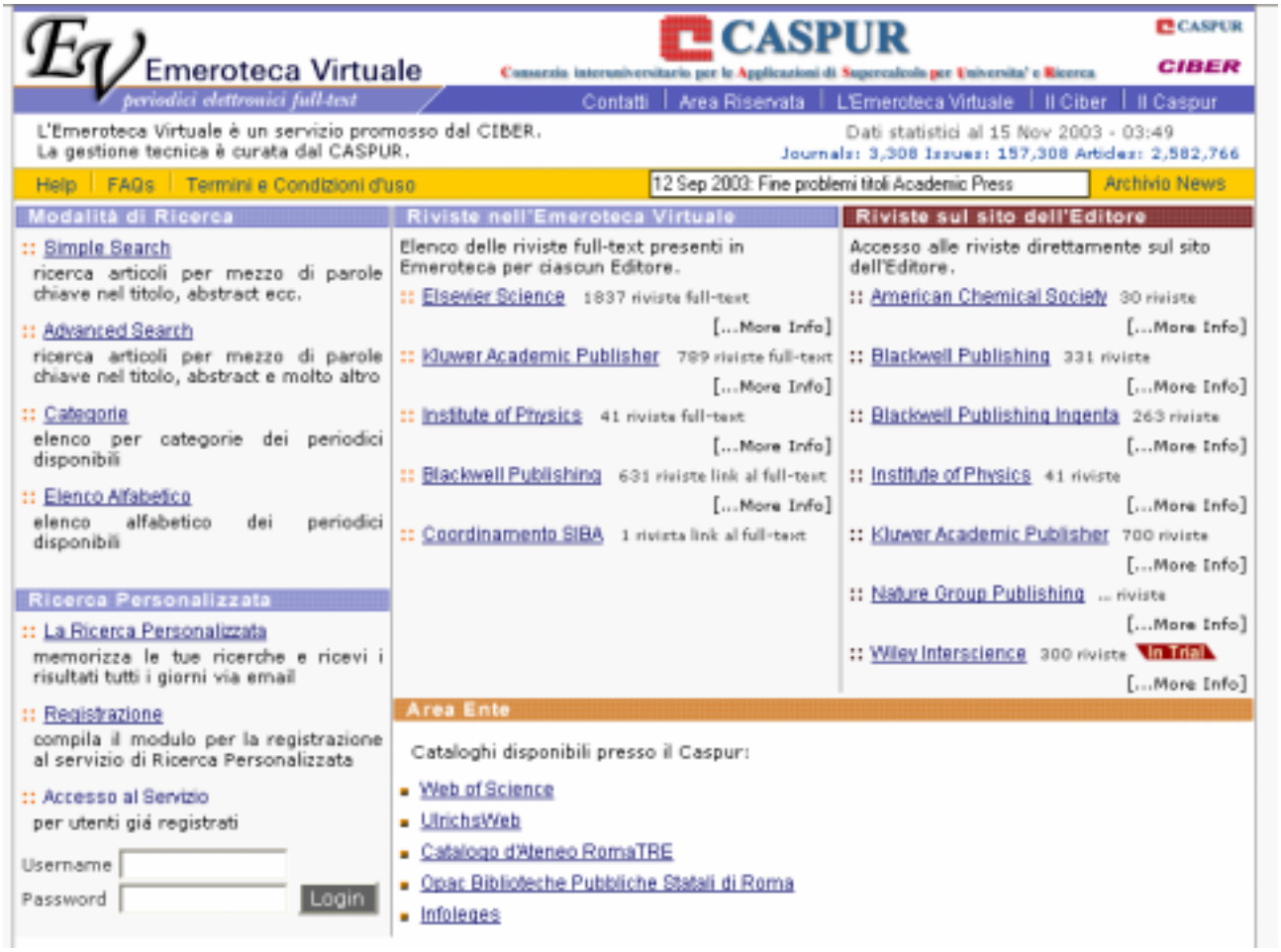


Fig.5 – Pagina di accesso al servizio di Emeroteca Virtuale (agg. Novembre 2003)

E' allo studio il test di questa interfaccia relativamente alle regole di accessibilità WEB così come emanate nell'ambito del programma W3C della *Web Accessibility Initiative*[19].

Quale architettura per l'Emeroteca?

La scelta dell'infrastruttura tecnologica da adottare per offrire un servizio di *Digital Library* pone sovente profondi interrogativi su quale sia l'architettura migliore relativamente:

- alle prestazioni del sistema lato utente (*Users' Quality of Service, U-QoS*) e alla *disponibilità* del servizio (intesa dal punto di vista della *service availability*)
- al numero di utenti potenzialmente coinvolti
- alla scalabilità nello *storage* e nella capacità di I/O (Input/Output) verso le memorie di massa del DB degli *Indici* e dell'area dei *metadati* e dei *full-text*
- alla *flessibilità* della soluzione per ciò che concerne l'integrazione con sistemi *non*-proprietari
- ai costi sostenibili dall'organizzazione che decide di mettere in piedi questo servizio

Nell'approccio seguito al CASPUR si è cercato di tener conto di queste condizioni al contorno, adottando soluzioni architetturali che meglio rispondessero alle esigenze del servizio di E.V. e che nel contempo fossero in linea con il *budget* di spesa previsto.

Tuttavia l'approccio, scelto fino all'agosto del 2003, basato sull'impiego di un unico server (SUN Enterprise-450) non è riuscito a far fronte a due richieste fondamentali:

- di *disponibilità* del servizio di E.V.
- di *flessibilità* della soluzione

Per questi motivi è stata avviata, in collaborazione con il settore sistemi del consorzio CASPUR, una sperimentazione su architetture distribuite ad accesso condiviso. Il risultato di questa sperimentazione è l'architettura attualmente adottata per l'Emeroteca, schematicamente illustrata in fig. 6.

Il *cluster* mostrato è composto da due server Linux indicati come ***periodici1*** e ***periodici2***, tra di loro speculari: ciascuno di essi, basato su un'architettura interna a doppio processore Xeon® della Intel con clock a 2.4 GHz, possiede una RAM di 4 GB ed è equipaggiato con la tecnologia dell'*Hyper-Trading* [20], grazie alla quale, sfruttando i cicli di *idle* del processore è possibile raddoppiarne virtualmente le capacità computazionali.

L'accesso alla *Storage Area Network* avviene su *bus fibre-channel* [21] che permette un *throughput* complessivo di 2 Gbit/sec in modalità *full-duplex*. I dischi utilizzati negli *array* sono in tecnologia Ultra-ATA[22], installati in sistemi RAID5 della Infortrend[23], e offrono una capacità di memoria complessiva pari a 4 TB (4000 GB).

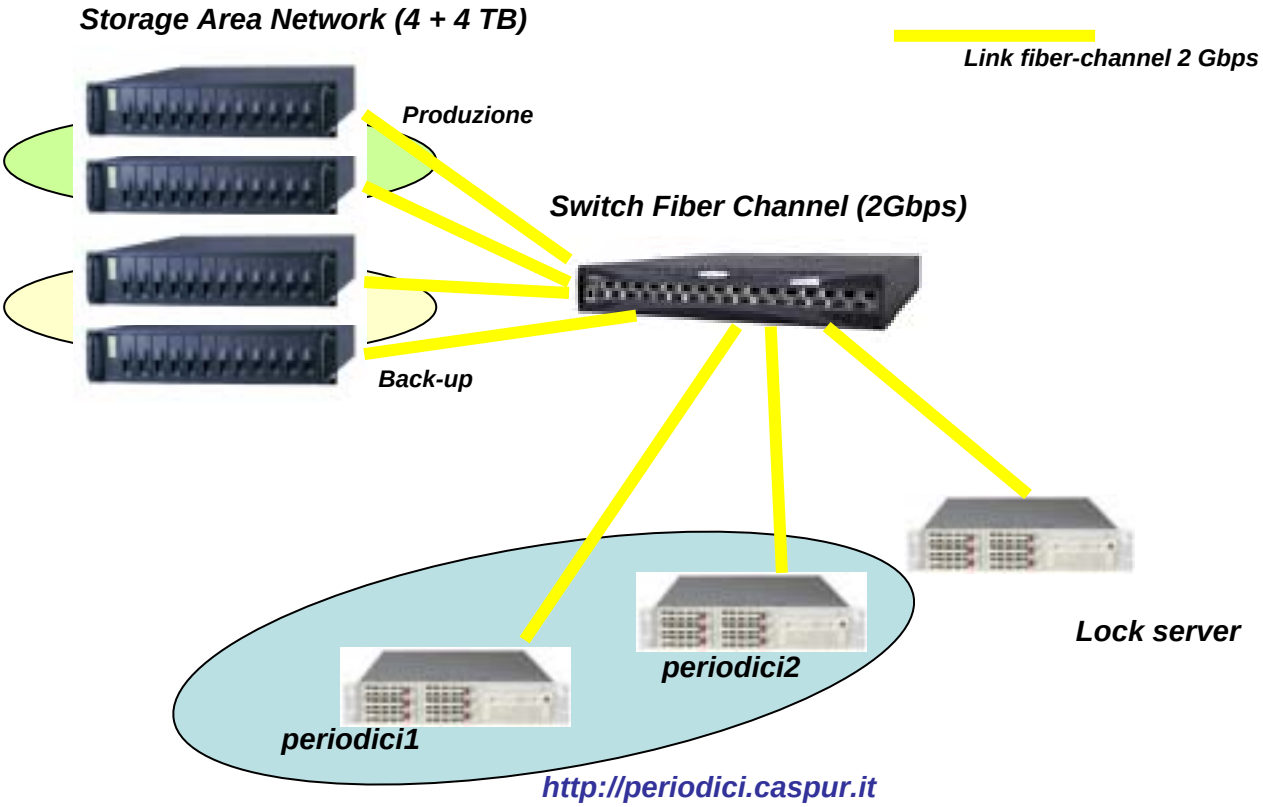


Fig.6 – Struttura dell'Emeroteca Virtuale del CASPUR

La macchina indicata come ***lock-server*** ha lo scopo di gestire l'accesso alla risorsa condivisa rappresentata dall'array dei dischi esterni, sui quali risiedono:

- la *Digital Object Repository*
- il DB degli Indici e dei *link*
- la gerarchia delle directory XML
- le pagine statiche dell'Emeroteca
- il codice eseguibile di Science Server e degli script *custom*

Il meccanismo di accesso condiviso al disco si basa sull'adozione di un particolare tipo di *filesystem* denominato *GFS, Global File System*. Questo sistema proprietario, distribuito dalla *Sistina Software* [24], permette di poter accedere (sia in lettura che scrittura) agli stessi files da qualunque nodo del cluster (la gestione dell'accesso è fatta in maniera *trasparente* rispetto all'utente), dimostrando notevoli doti di scalabilità e prestazioni. E' ovvio notare come questa soluzione garantisca un elevato grado di disponibilità dei dati (e quindi dell'informazione ad essi associata), poiché questi continuano ad essere fruibili anche in seguito alla caduta di uno dei due nodi del *cluster*.

Come detto in precedenza su questa architettura sono stati eseguiti dei *benchmark* (nei quali sono stati impiegati quattro nodi ed un *lock-server*), il cui scopo è stato quello di valutarne i limiti prestazionali e l'affidabilità. I risultati di questi test sono mostrati nella tabella seguente (tab. 2), che mostra come nella configurazione attuale dell'E.V. si possa arrivare ad un throughput di accesso (in lettura) alla *repository* dei dati dell'E.V. di 230 MB/sec.

	Read	Write
1 client	**122**	**156**
2 clients	**230**	**245**
3 clients	**291**	**297**
4 clients	**330**	**300**

Tab.2 – Test di I/O per un cluster composto da 4 nodi GFS ed un lock server (i dati sono espressi in MB/sec)

Uno dei pochi limiti di GFS consiste nella dimensione massima del singolo *filesystem*, limitata a 2 Terabyte (2000 GB). Al momento, per superare questa limitazione bisogna orientarsi su soluzioni diverse, tra le quali e' sicuramente da citare il *filesystem* StorNext della Adic [25]. Come GFS, anche *StorNext* e' un *filesystem* distribuito SAN-based che supporta unità logiche con dimensioni fino a 16 Petabytes (16 milioni di GB) e presenta un'architettura più semplice del *filesystem GFS*, in quanto non necessita del *lock-server*..

La soluzione illustrata risolve, a nostro avviso, efficacemente il problema del collo di bottiglia dell'architettura a *singolo server*, per ciò che attiene il parametro di disponibilità del servizio e si mostra sicuramente scalabile (sono stati eseguiti test al CASPUR fino a 7 nodi GFS contemporaneamente in funzione). E' sufficiente, infatti, definire il server *virtuale* ***periodici.caspur.it*** all'interno del sistema di risoluzione del DNS, in modo che quest'ultimo risolva alternativamente, per ogni richiesta esterna di risoluzione, le coppie di indirizzi (tecnica della *round-robin IP name resolution*):

- IP1 -> periodici1.caspur.it
- IP2 -> periodici2.caspur.it

e:

- IP1 -> periodici2.caspur.it
- IP2 -> periodici1.caspur.it

perché il servizio di Emeroteca risulti sempre disponibile indipendentemente dal numero di sistemi in quel momento attivi (nell'ipotesi, ovvia, che sia attivo almeno un server del *cluster*). L'unico effetto eventualmente visibile lato utente è quello di un possibile degrado prestazionale (traducibile in un ritardo maggiore con il quale il servizio *risponde* alle sue richieste).

Inoltre la scelta operata verso il binomio Intel®-Linux è sicuramente in linea con l'esigenza di contenere i costi di questa infrastruttura distribuita e offre interessanti prospettive di crescita in un contesto di server eterogenei (cosa non permessa con sistemi proprietari), nei quali la scelta della macchina da aggiungere al *cluster* è fatta sulla base del miglior rapporto qualità/prezzo.

Nelle figg.7a-7b-7c sono mostrate le immagini rispettivamente del cluster dei server SuperMicro®[26], dello switch *fiber-channel* della Brocade®[27] e della *Storage Area Network* basata su sistemi Infortrend®.

7° **7b** **7c**

Fig.7°,7b e 7c – Fotografie del sistema di Emeroteca Virtuale dislocato nella sala macchine del CASPUR

Il problema del back-up dei dati on-line

Il classico ed annoso problema del back-up dei dati dei sistemi informatici, assume per i sistemi di *digital library* qual è quello dell'E.V., toni ancora più esasperati non solo per ciò che concerne la criticità del servizio, ma soprattutto per la gran mole di informazione immagazzinata all'interno della *repository*. Una caratteristica comune a tutti i sistemi che offrano accesso *full-text* in locale ad almeno più di un migliaio di testate elettroniche, è che la mole dei dati associati agli articoli disponibili si traduce non solo in una dimensione complessiva della *repository* elevata (> 1000 GB), ma in un numero complessivo di file (di dimensioni comprese tra 1KB ed 1 MB) che può facilmente raggiungere e superare i 10 milioni.

Nello specifico caso dell'E.V. il numero di testate elettroniche disponibili *on-line* supera le 3300 unità, mentre quasi 2 milioni e 600 mila sono gli articoli disponibili a testo completo (al primo novembre 2003). La dimensione complessiva dell'archivio dei *digital objects* supera ampiamente i 2 TB (2000 GB), mentre il numero dei file associati sfiora i 50 milioni di unità (includendo anche i descrittori XML). Giornalmente, inoltre, vengono caricati 6 *dataset* di 4 diversi editori, per un totale di 2 GB distribuiti su circa 25000 file, ed il processo di caricamento delle riviste di Science Server (*loader*) genera per lo meno altri 10000 nuovi file XML.

Anche impiegando tecnologie di accesso al disco estremamente efficienti per ciò che riguarda l'I/O (quale quella offerta dal link *fibre-channel* a 2 Gbit/sec) una copia completa e speculare dell'archivio dell'E.V., verrebbe ultimata non prima di un paio di giorni nella migliore delle ipotesi, ed escludendo un eventuale *post-processing* di validazione per i dati copiati. E' ovvio che durante tale periodo il servizio di E.V. sarebbe inutilizzabile, a meno di non indirizzare tutta l'utenza verso gli archivi dei vari editori (sempre che tutti permettano la consultazione *on-site*, cosa non vera per l'editore Elsevier).

E' per questo motivo che è sempre in linea un sistema SAN tecnologicamente identico a quello utilizzato in esercizio (si faccia riferimento allo schema di fig.6), e collegato ad uno dei due server GFS. Prima di mettere in esercizio la nuova Emeroteca si è provveduto ad allineare i due sistemi SAN, in modo che risultassero speculari; l'allineamento giornaliero e/o settimanale è garantito da procedure automatiche che

- copiano i *dataset* caricati dall'editore sia sull'archivio in esercizio che su quello di back-up
- eseguono una copia periodica (settimanale) delle directory con i descrittori XML, degli Indici e dei *link*
- eseguono una copia giornaliera delle liste statiche di accesso alle riviste e delle directory nelle quali è installato il software Science Server e quello delle procedure *custom*

In questo modo è possibile garantire al servizio di E.V. un *fuori servizio* non superiore ad un'ora nel caso in cui il *filesystem* GFS dovesse andare incontro ad un grave problema tecnico che compromettesse l'integrità dell'archivio.

Valutazione della *Quality of Service* lato utente: un approccio quantitativo

Nel valutare la qualità del servizio di E.V. come risulta essere percepita da un utente remoto, bisogna tener conto di molteplici fattori, pertinenti alcuni alle infrastrutture (sistemi informatici e telematici), altri al software di *information retrieval* , altri ancora alla *qualità* ed alla *organizzazione* del materiale digitale offerto (che investono la sfera del *content management*), altri infine che riguardano il *supporto* che nell'ambito del servizio viene offerto all'utenza (potremmo dire *user care* parafrasando l'espressione, tipica in ambiente commerciale, di *customer care*). Il servizio infatti può essere pensato, nel suo complesso, come una struttura a più livelli (fortemente sinergica), con i sistemi al livello più basso, per passare allo strato di astrazione del software di *information retrieval*, al *content management* e, infine, a quello di *user care*.

La misura di una *QoS* complessiva del servizio di E.V. è quindi una funzione complessa che ben difficilmente può essere ricondotta a misure puramente quantitative, ma che investe anche un ambito soggettivo nella misura in cui l'utente *medio* si *sente soddisfatto* del servizio che sta utilizzando. Certamente poter avere una comprensione più approfondita della tipologia di utenza che interagisce con il servizio può servire a curare gli aspetti più pertinenti ai livelli superiori, ed uno studio in tal senso è stato recentemente eseguito dal CASPUR per il proprio servizio di E.V. [28].

Quello che si vuole comunque illustrare in questo paragrafo è una tipologia di misura di tipo quantitativo il cui scopo è stato quello di valutare l'effettivo grado di miglioramento introdotto con la nuova architettura, confrontando i dati, ottenuti utilizzando lo stesso metodo di indagine, del vecchio e del nuovo sistema di E.V.

In questo metodo di indagine utilizzato si sono misurati (con una frequenza costante di 30 minuti) i tempi di risposta del portale di accesso all'E.V. utilizzando una *postazione esterna* al server (si faccia riferimento alla fig. 8).

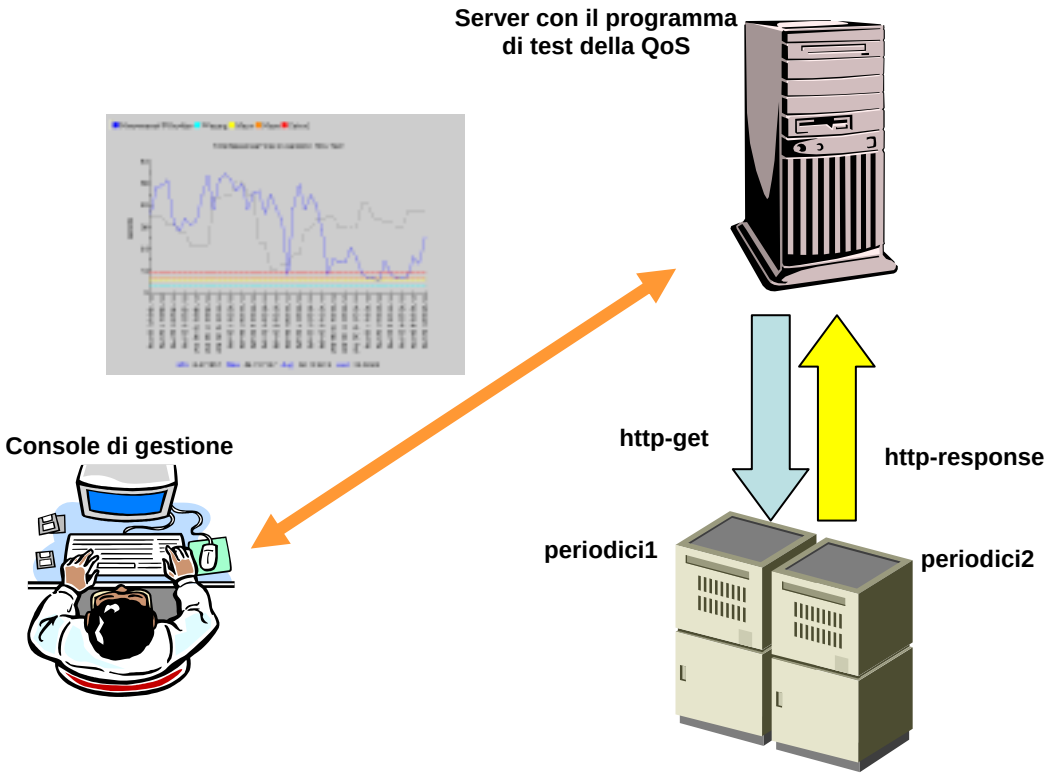


Fig.8 – Architettura utilizzata per i test di QoS dell'Emeroteca Virtuale

Il software di analisi utilizzato [29] è in grado di *simulare* un utente che si connette alla *home page* del server di Emeroteca, fornendo un dettaglio di tutti i tempi delle singole transazioni delle quali si compone questo processo. Nel caso specifico ci si è concentrati unicamente sul *tempo totale di risposta* del server dei periodici dal momento che questo rappresenta per l'utente l'effettivo tempo di attesa della *home page*.

Nei grafici delle fig. 8 e 9 sono mostrati gli andamenti di questo parametro per due periodi di analisi:

- dal 16 maggio al 23 maggio 2003 (misure eseguite sul vecchio sistema)
- dal 24 ottobre al 29 ottobre 2003 (misure eseguite sul nuovo sistema)

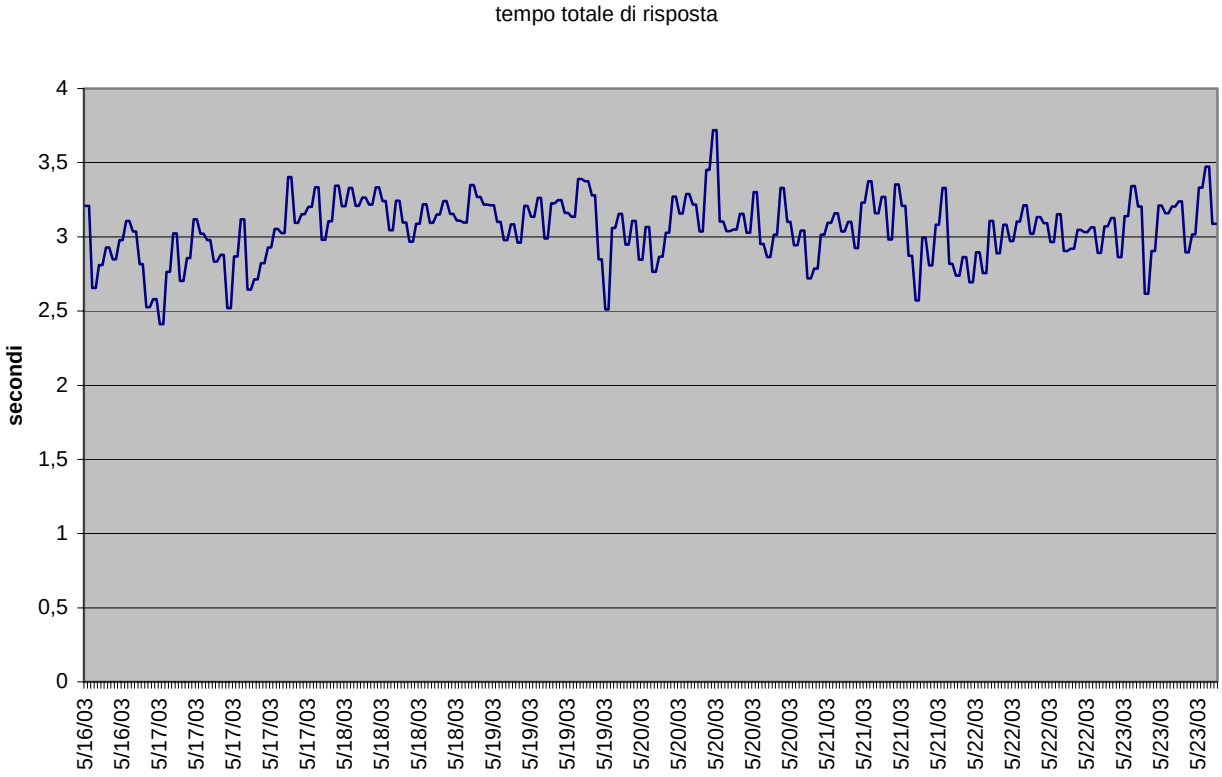


Fig.8 – Misure del tempo totale di risposta ottenuti sull'home-page dell'E.V. sul vecchio sistema SUN

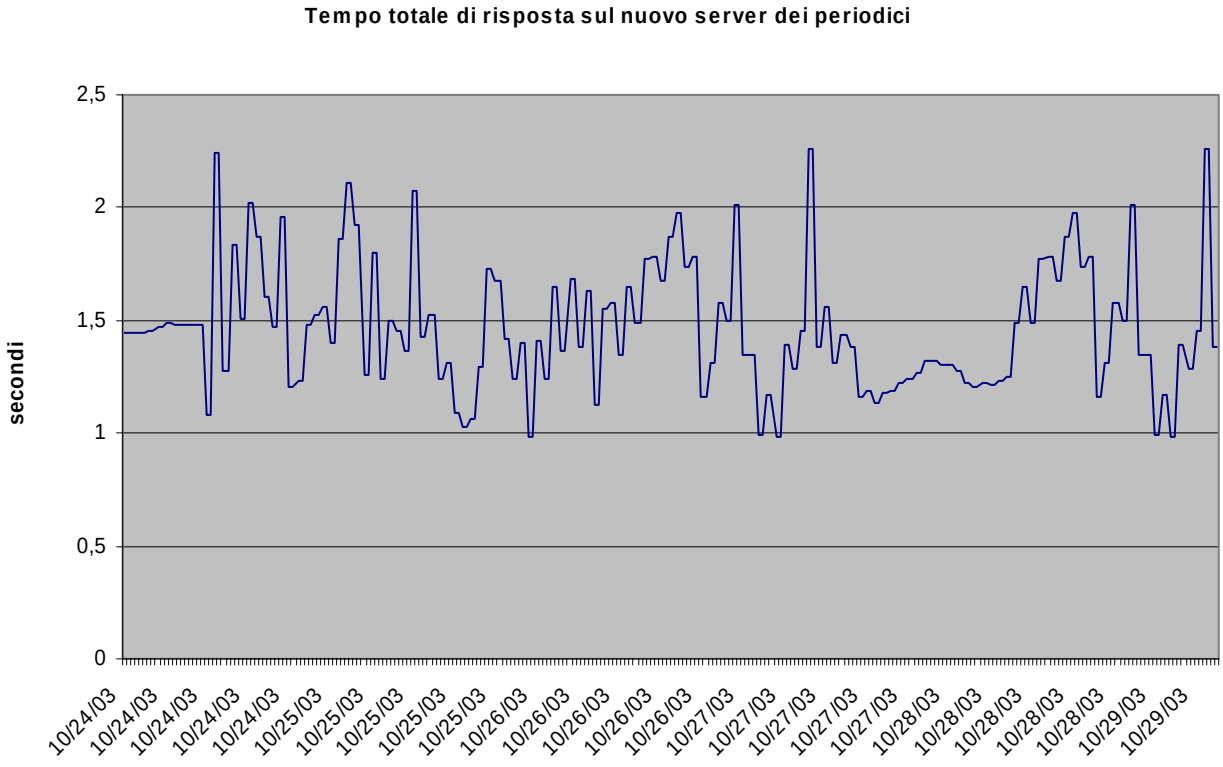


Fig.9 – Misure del tempo totale di risposta ottenuti sull'home-page dell'E.V. sui due nuovi sistemi Linux

I valori medi dei tre andamenti si attestano sui 3 secondi circa per il primo ed 1,5 secondi per i successivi due, e questo a parità di *condizioni al contorno* (numero medio di utenti contemporanei, numero medio di accessi al *full-text* giornalieri, etc.), il che dimostra come la nuova infrastruttura abbia l'ulteriore vantaggio di aver portato *oggettivamente*, come era logico, ad un dimezzamento dei tempi di accesso all'Emeroteca.

Utilizzando il medesimo software di test, avendo come *target*, questa volta, il nuovo sistema dell'E.V., si sono recentemente eseguite analisi dei tempi di riposta del server per ricerche *semplici* ed *avanzate*. Nel primo caso si è utilizzando una sola chiave di ricerca (costituita dalla parola "*cell*", ad altissimo impatto nel DB dei *full-text* indicizzati), mentre nel secondo si è aumentata la selettività proponendo una chiave di ricerca più complessa, includendo comunque al suo interno il termine della prima ricerca (la chiave utilizzata è stata "*cell" AND "gamma irradiation"*).

A titolo illustrativo si riportano i grafici delle figg. 10 e 11, contenenti i tempi totali di risposta ottenuti sul nuovo sistema per le due ricerche; nella prima il numero di occorrenze tocca quasi le 380.000 unità, mentre nella seconda queste ultime sono 168. In questa analisi si è utilizzato come *target* l'indirizzo *virtuale* http://periodici.caspur.it.

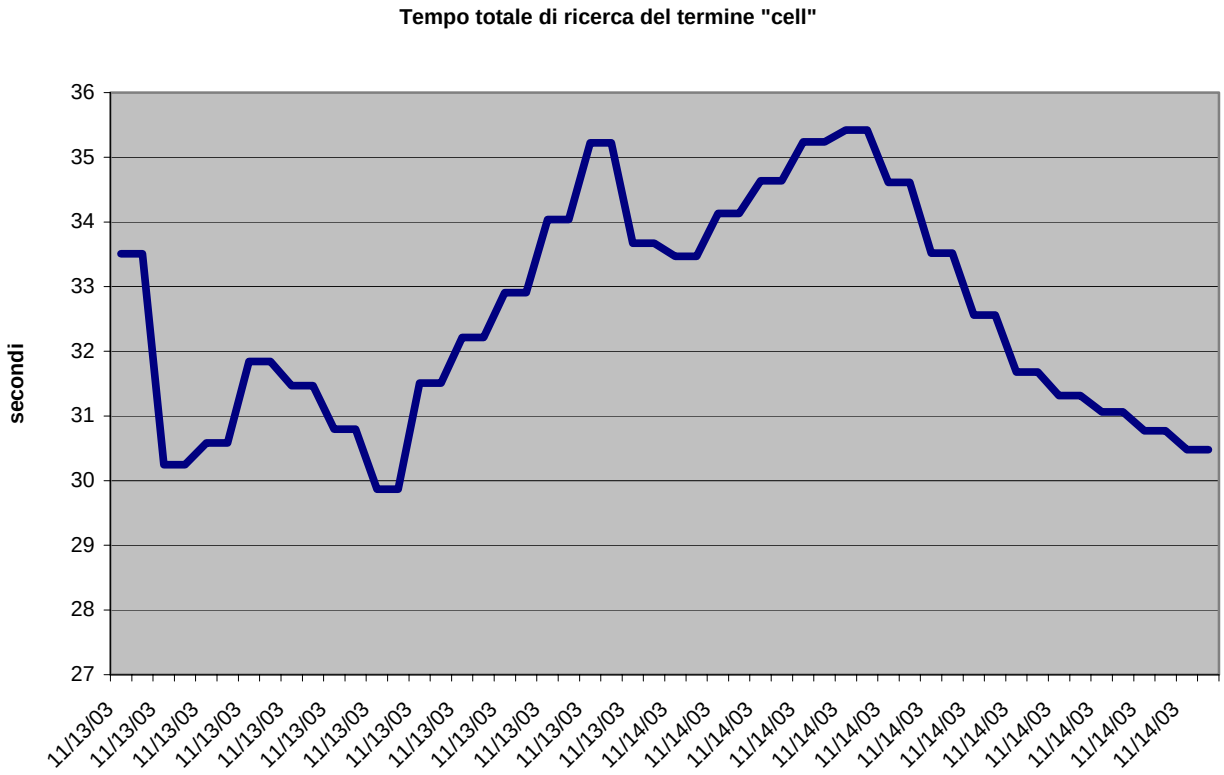


Fig.10 – Tempo totale di ricerca del termine "cell" sui due nuovi sistemi Linux dell'E.V.

A fronte di questi dati possono essere fatte alcune considerazioni:

- l'operazione di ricerca è certamente una delle attività più *computational intensive* per l'E.V., dal momento che si possono avere anche 35 secondi nell'attesa dei risultati
- una ricerca semplice e *ad ampio spettro* impegna effettivamente il sistema più di una ricerca *mirata*

- una ricerca troppo generica che *ritorni* una mole così elevata di risultati, oltre che essere inutile per l'utente è anche dannosa per il sistema, perché implica un elevato consumo di risorse computazionali diversamente impiegabili

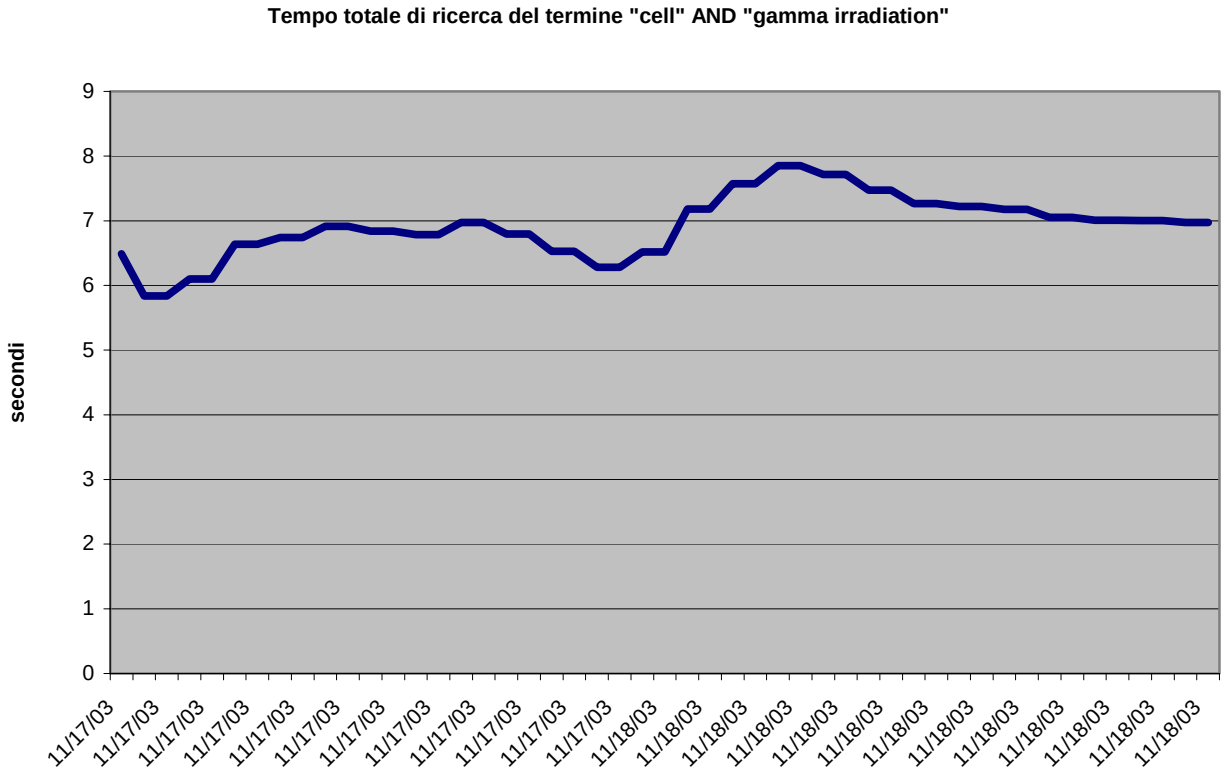


Fig.11 – Tempo totale di ricerca del termine "cell" AND "gamma irradiation"

E' nostra intenzione, comunque, quella di continuare queste analisi includendo altre tipologie di ricerca (c'è un'altra categoria non considerata, la *expert search*), e variando le parole chiave (a basso ed alto impatto) e i campi di ricerca (ricerca nel titolo, nell'*abstract*, limitata nel tempo, ect.). L'obiettivo è quello di fornire un utile ed *oggettivo* strumento di valutazione dell'impatto del meccanismo di ricerca sull'infrastruttura informatica dell'E.V.

Uno sguardo al futuro

L'Emeroteca Virtuale è un servizio in continua evoluzione, sotto molteplici aspetti. Dal punto di vista dei contenuti è prevedibile un incremento del numero di *e-journals* distribuiti legato alla stipula di contratti consortili con nuovi editori; anche l'aspetto di *presentazione* dei contenuti è in evoluzione sia per ciò che riguarda il numero di riviste accessibili nel formato di visualizzazione HTML alternativo al PDF, sia per quanto riguarda il *look-and-feel* del portale di accesso; riferendoci inoltre allo strato *sistemi*, del servizio di E.V., è ragionevole prevedere un incremento di un 20% delle capacità di memorizzazione dell'area di *storage* (attualmente sono utilizzati 4.8 TB degli 8 TB disponibili) per far fronte alla crescita *fisiologica* dell'archivio dei dati, oltre ad un

probabile aumento del *cluster* con l'aggiunta di un nuovo server, se esigenze di *performance* lo dovessero richiedere.

Tuttavia ci interessa menzionare in questo contesto le tre attività che verranno condotte a partire dal primo semestre 2004 e che avranno ricadute sul servizio di Emeroteca, ovvero:

- lo studio per l'applicazione nel contesto dell'E.V. dei risultati ottenuti nell'attività di ricerca promossa da un gruppo CASPUR del *settore applicazioni del calcolo scientifico* [11]su un motore di ricerca *neurale* che possa superare i limiti dell'attuale motore di ricerca *full-text* di Science Server; l'idea è quella di utilizzare tecniche automatiche che possano classificare i documenti dell'Emeroteca in base al loro contenuto semantico (a differenza dell'approccio semiotico dei motori di ricerca standard) mettendolo poi in relazione all'*oggetto* cercato dall'utente e proponendo un *ranking* dei risultati sulla base della maggiore affinità semantica dei documenti trovati con la chiave di ricerca utilizzata
- l'introduzione di una vera ed efficace *linking technology*, sfruttando prodotti già disponibili in commercio [30] che abbiano recepito le raccomandazioni dello standard OpenURL, eventualmente integrandoli con personalizzazioni sviluppate *ad hoc* per l'Emeroteca
- l'avvio di un progetto di *Open-Archive* in collaborazione con le università del CIBER ed in sinergia con analoghe iniziative del CILEA ([31])

Come nota conclusiva si vuole osservare che l'impatto che il servizio di E.V. ha nel CASPUR a livello sistemi e di personale direttamente o indirettamente coinvolto, la *visibilità* che ha all'interno del panorama universitario italiano, relativamente al numero di persone che costituiscono la sua base di utenti potenziali, e (non ultima) le implicazioni tecnologicamente innovative poste dalle *digital library* nel contesto della *computer science* e del *content management* (intimamente connesse in un contesto più ampio di *knowledge management*), fanno di questo servizio una delle *sfide* più intriganti nei prossimi anni non solo per il *settore servizi di automazione per le biblioteche* (direttamente demandato alla sua gestione), ma anche, in un contesto di collaborazione reciproca, sia per gli altri settori nei quali il CASPUR è suddiviso, che, a maggior ragione, per le altre università del CIBER.

Conclusioni

In questo articolo sono state illustrate le caratteristiche del servizio di Emeroteca Virtuale offerto dal CASPUR alle università che fanno parte del coordinamento CIBER, cercando di evidenziare le motivazioni che hanno portato all'attuale architettura. Viene inoltre proposta una metodologia di indagine per un'analisi della *Quality of Service* lato utente, basata su misura quantitative dei tempi di risposta dell'interfaccia WEB di accesso al servizio.

Ringraziamenti

Si desiderano ringraziare la dott.ssa Gargiulo, il dott. Farinelli e la dott.ssa Marquardt, del *Settore Servizi di Automazione per le Biblioteche* del CASPUR, che hanno partecipato alla revisione del testo dell'articolo, e il dott. Giuseppe Palumbo del Settore Sistemi del CASPUR per le informazioni fornite relativamente ai dettagli sull'infrastruttura tecnologica dell'E.V.

Riferimenti

[1] – un sito sicuramente utile da consultare interamente dedicato alle problematiche della *digital library* è quello del D-*Lib*® Magazine (http://www.dlib.org); un interessante riferimento per avere un panorama chiaro del contesto nel quale ci si sta muovendo è rappresentato dal libro *"La biblioteca digitale"* – di Annamaria Tammaro e Alberto Saltarelli – Milano Editrice - 2000

[2] – si faccia riferimento alle pagine della International Federation of Library Association and Institution, IFLA, per avere accesso ad una lista abbastanza estesa di strutture pubbliche che offrono servizi di Digital Library (http://www.ifla.org) – interessante la lista proposta dall'università dell'Indiana all'indirizzo http://www.indiana.edu/~vlib/Digital_Libraries/

[3] – cfr http://www.openarchives.org/

[4] – http://ciber.caspur.it

[5] - http://ciber.caspur.it/informazioni/regolamento.html

[6] - http://siba2.unile.it/

[7] – http://bids.citicord.uniroma1.it/bids/default.htm

[8] - http://ciber.caspur.it/informazioni/membriciber.html

[9] - http://www.miur.it/ustat/Statistiche/BD_univ.htm

[10] – www.scienceserver.com

[11] – Federico Massaioli, Antonino Sgalambro, Angelo Canaletti; Approcci innovativi al document management – Quaderni di AIDA: Le nuove sfide dell'e-content management per utenze differenziate – raccolta degli atti del seminario Bibliocom 2003 dell'AIB – Roma 31 Ottobre 2003 (http://www.aidaweb.it/2003/bibliocom2003.html)

[12] - http://www.mysql.com/

[13] - http://www.php.net/

[14] – www.sas.com

[15] – www.perl.org

[16] - http://www.doi.org/handbook_2000/index.html

[17] – Van de Sompel, Herbert, Oren Beit-Arie, Open Linking in the Scholarly Information Environment Using the OpenURL Framework – D-*Lib* Magazine – Marzo 2001 – Vol.7 – num.3 (http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html)

[18] - http://www.crossref.org

[19] - http://www.w3.org/WAI/

[20] - http://www.intel.com/technology/hyperthread/index.htm?iid=ipp_srvr_proc_xeon+feature_f2htt&

[21] - http://www.fibrechannel.org/

[22] - http://www.seagate.com/support/kb/disc/ultra_ata100.html

[23] – www.infortrend.com

[24] - http://www.sistina.com/products_gfs.htm

[25] - http://www.adic.com/ibeCCtpItmDspRte.jsp?section=10024&item=121889

[26] - http://www.supermicro.com/PRODUCT/SUPERServer/SuperServer7043P-8R.htm

[27] - http://www.brocade.com/products/silkworm/silkworm_3800/sw_3800.jsp

[28] - C.Conti, U.Contino, G.Farinelli, P.Gargiulo, L.Marquardt - Digital Libraries and Users: an Italian Experience. Changes in academic users' attitudes, perceptions and usage of study and research tools in a hybrid context – lavoro presentato al convegno "Toward a User-Centered Approach to Digital Libraries" – 8-9 Settembre 2003; Espoo (Finlandia) (http://www.caspur.it/~contino/diglib_2003_06_20-art.pdf)

[29] – *Firehunter*™ per gentile concessione della *Agilent Technologies* (http://www.firehunter.com)

[30] – ad esempio *SFX* della Ex-Libris, che per primo ha promosso lo standard dello Open-URL (http://www.sfxit.com/)

[31] - http://www.cilea.it/servizi/g/AEPIC/OA/index.html