

EDITORIALE	1
<i>Marco Lanzarini</i>	
Il nuovo sito del CINECA	3
<i>Maria Gervasio, Patrizia Coluccia, Gabriele Neri</i>	
Anagrafe Nazionale Studenti	6
<i>Luca Donati, Gianna Fabiani, Francesca Pruneti,</i>	
Un chip per studiare il genoma: i Microarray	10
<i>Silvia Giuliani, Elda Rossi</i>	
VisIVO Virtual Observatory e visualizzazione di dati Astrofisici	13
<i>Ugo Becciani, Marco Comparato, Claudio Gheller</i>	
Un TTS per l'assistenza ai clienti dei sistemi gestionali per le Università	17
<i>Lucino Canino, Stefania Caponetti</i>	
OSIRIS e l'evoluzione dell'e-learning	20
<i>Eugenia Rinaldi</i>	
Riabilitazione a distanza: un progetto sperimentale per la TV interattiva	22
<i>Anna Conti, Andrea Venturi</i>	
CoEdiT: Collaborative Editing Tool	25
<i>Stefania Gardelli, Daniela Lotti, Elena Mantovani, Salvatore Rago</i>	
La Banca Dati Biologica	28
<i>Luca Demattè, Marisa De Rosa, Maria Teresa Marano, Elisa Rinieri</i>	
La "sicurezza ambientale" al CINECA	31
<i>Sergio Biagi Foresti, Alessio Mauri</i>	
<b>RUBRICHE</b>	<b>33</b>

## GESTIONE E ANALISI DELL'INFORMAZIONE

EINS: EUROPEAN INFORMATION NETWORK SERVICES

COMUNICAZIONI E SISTEMI DISTRIBUITI

GESTIONE SISTEMI

## NEWS E APPUNTAMENTI

a cura dell'Ufficio stampa

*In copertina:*

*Immagine tratta dall'articolo di pagina 13: Visualizzazione di dati su griglia (simulazione di un jet rotante) attraverso la tecnica "volume rendering"*

# EDITORIALE

*Marco Lanzarini*

## Avviate le attività della Giunta Amministrativa del CINECA

Nell'inverno del 2004, il Consiglio di Amministrazione del CINECA ha deliberato la costituzione della Giunta Amministrativa, così come previsto dallo Statuto. Si tratta di un nuovo organo di governo che consente di snellire alcuni compiti del CDA che, ormai composto da oltre 28 membri (i rappresentanti delle 25 Università consorziate, il rappresentante del CNR, il Presidente e il Direttore), può così concentrarsi sulle linee di governo del Consorzio. Il CDA ha eletto tra i propri componenti i membri della Giunta (presieduta dal Presidente del CINECA, e composta dal Vice Presidente, quattro componenti del CDA e il Direttore) e in giugno le attività del nuovo organo hanno preso il via. I suoi compiti riguardano: l'approvazione di contratti e convenzioni, l'assunzione di obbligazioni, la delibera di alcune tipologie di spese, l'assunzione di personale. Attività da svolgere comunque nei limiti fissati dal CDA, per il quale cura altresì l'istruttoria degli atti da sottoporre alla sua approvazione. Nell'ambito delle attività del nuovo organo di governo del Consorzio rientra anche la redazione di una relazione da presentare periodicamente al Consiglio di Amministrazione sull'attività svolta.

### *ESSE3 avviato presso l'Università di Ferrara*

Lo scorso maggio, l'Ateneo di Ferrara ha adottato ESSE3, il Sistema informativo per la gestione della Didattica e degli Studenti, realizzato dal CINECA. L'avviamento di ESSE3 presso l'Ateneo estense porta a diciassette il numero delle università che utilizzano il Sistema (altri sette atenei seguiranno nei prossimi mesi) e porta a circa 300.000 il numero totale degli Studenti amministrati sul territorio. Oltre alla gestione della didattica dell'ateneo, ESSE3 porterà una forte evoluzione nell'erogazione dei servizi agli Studenti via Internet, non solo con l'obiettivo di ridurre il carico di lavoro delle segreterie e snellire l'espletamento delle pratiche burocratiche anche da parte degli studenti, ma anche con l'intento di supportare lo sviluppo di "comunità" fra studenti e docenti, per una facilitazione ed arricchimento della didattica.

Per meglio comprendere la portata dell'adozione di ESSE3 nell'Ateneo di Ferrara, è bene ricordare che il completamento dell'avviamento di tutto il sistema richiederà l'intero ciclo di un anno accademico, durante il quale ver-

ranno avviati i diversi moduli in corrispondenza dei diversi appuntamenti che scandiscono l'anno accademico: (prove di ammissione, immatricolazione, iscrizione, pagamento tasse, percorso di carriera, conseguimento titolo, master, dottorati ...). Da parte del CINECA, oltre a porgere le proprie congratulazioni all'Università di Ferrara, c'è la massima disponibilità ed attenzione alla realizzazione degli obiettivi della propria consorziata, attraverso la migliore "fruizione" di ESSE3 e la disponibilità di KION, azienda dedicata allo sviluppo e diffusione del Sistema.

### *Il voto telematico per l'elezione del Rettore del Politecnico di Torino*

Il nome del nuovo rettore del Politecnico di Torino, il Professor Francesco Profumo, è uscito dalle urne elettroniche del sistema di voto telematico del CINECA alla terza tornata elettorale, che si è svolta il 21 giugno. Dopo l'elezione di Giovanni Del Tin nel 2001, l'Ateneo ha scelto nuovamente di utilizzare il sistema del Consorzio. Sicurezza nelle votazioni e velocità nelle operazioni di scrutinio sono solo alcune delle caratteristiche di e-vote apprezzate dall'ateneo piemontese, senza dimenticare che le postazioni di voto sono già attive, essendo le stesse utilizzate presso tutti gli atenei italiani per l'elezione delle commissioni giudicatrici delle votazioni comparative fin dal 1999. Il primo turno di votazioni si è tenuto il 15 giugno e poi a seguire il 17 e il 21 giugno. Oltre 2300 sono stati gli elettori chiamati alle urne per eleggere, tra quattro candidati, il Rettore dell'ateneo per il quadriennio 2005-2009. Al termine delle operazioni di voto, i risultati dello scrutinio (che ha richiesto solo mezz'ora di tempo) e alcuni dati statistici sono stati pubblicati su Internet sul sito del Politecnico.

<http://www.swas.polito.it/services/Elezioni/rettore.asp>

### *La Divulgazione Scientifica in ambito Sanitario*

Lo scorso giugno a Napoli, presso la Colonia Geremicca, il CINECA ha presentato "i dati

sull'appropriatezza prescrittiva nelle cure primarie nella Regione Campania e in particolare nelle 3 ASL del Campione ARNO" (ASL Napoli 1, ASL Napoli 4 e ASL Caserta 2). Tra i relatori, dirigenti e farmacisti delle 3 ASL menzionate oltre ai rappresentanti dell'ACP (Associazione Culturale Pediatri). Questo evento si inserisce nell'ambito di una serie di iniziative legate alla divulgazione in ambito biomedico e sanitario avviate nello scorso inverno con i convegni "Donne e Farmaci" e "Bambini e Farmaci", organizzati dal CINECA in collaborazione con l'Istituto il Mario Negri Sud e l'Istituto di Ricerche Farmacologiche Mario Negri a Roma nella Sala Cenacolo presso la Camera dei Deputati.



Nel corso degli incontri, molto apprezzati non solo dagli addetti ai lavori, ma anche dalla stampa specializzata, sono state presentate due nuove edizioni del *Rapporto Arno* dedicate alla prescrizione farmaceutica nella popolazione femminile e nei bambini in oltre 28 ASL italiane. Le immagini dei convegni di Roma sono state riprese e sono disponibili per la consultazione.

(<http://www.cineca.it/streaming/donnefarmaci/>  
<http://www.cineca.it/streaming/bambini/>)

doi:10.1388/notizie-52-editoriale

# Il nuovo sito del CINECA

di Patrizia Coluccia, Maria Gervasio\*, Gabriele Neri

\*Italian Team

Il nuovo portale CINECA, progettato per essere soprattutto uno strumento utile e funzionale, offre sia informazioni riguardanti il Consorzio e le sue attività sia servizi rivolti al pubblico, ai suoi clienti e partner, valorizzando la specificità e la ricchezza delle sue competenze.

Realizzato con la tecnologia *BackStage Director* - uno strumento di *web content management* (CMS) sviluppato dal

Dipartimento Gestione e Analisi dell'Informazione (vedi riquadro) - il sito si appoggia a un'*Area di lavoro cooperativo* che permette di aggiornare i contenuti in maniera rapida e immediatamente visibile sul web, così come di sviluppare nuove pagine e nuove sezioni.

Il portale è strutturato in un'area istituzionale CINECA, a cui si accede entrando nel sito

## Il sistema di web content management BS-D

di Andrea Buda, Luca Cicale

Nella realizzazione del portale del CINECA, la scelta di utilizzare il sistema di *web content management BackStage Director* (BS-D, descritto nel numero 40 del *Notiziario*) ha semplificato notevolmente gli aspetti tecnici tipici di un portale e ha consentito agli esperti di comunicazione di poter organizzare con una metodologia semplificata la gestione dei contenuti.

L'integrazione di BS-D con la tecnologia *lanus* ha poi permesso l'implementazione di un sistema per la gestione delle redazioni distribuite: una volta definito chi e come poteva interagire con i contenuti, l'implementazione tecnica è stata immediata.

La flessibilità che caratterizza BS-D ha consentito di integrare e riorganizzare l'insieme dei servizi ed informazioni presenti sul portale, con un risultato armonico e ben strutturato. La semplicità di utilizzo della redazione e della regia fanno sì che ogni informazione possa essere pubblicata in tempo reale, senza richiedere interventi di "esperti" di web. In questo modo, news, informazioni e comunicazioni trovano in tempo reale la loro collocazione ideale, e vengono rese fruibili al pubblico nel momento desiderato e senza ritardi.

Il sito del CINECA è stato il primo realizzato nella nuova versione di BS-D. L'applicativo, sviluppato al CINECA nel 1999, è infatti ormai giunto alla *release* 4.0, che ha visto la riscrittura totale in *Java* della CDA (*Content Delivery Application*) ovvero del motore che genera e distribuisce le pagine del portale. I risultati ottenuti si possono riassumere in un aumento delle performance, maggiore affidabilità, portabilità e ottimizzazione del codice oltre a una maggiore libertà nella distribuzione. Un'attenzione particolare è stata posta nel mantenere la compatibilità totale con le vecchie versioni del software. Anche la CMA (*Content Management Application*) ovvero l'insieme tutti gli strumenti utilizzati dalla redazione per la creazione, gestione e pubblicazione dei contenuti è stata migliorata nell'usabilità ed è stata introdotta la possibilità di generare pagine in standard XHTML. A tale scopo sono ora disponibili gli strumenti per la ben formazione dei contenuti, la validazione delle pagine secondo gli standard dettati dal W3C e la gestione dei caratteri non occidentali. Inoltre, questa ultima *release* ha consentito un notevole passo in avanti nella strutturazione dei template, rendendoli più organizzati e meglio strutturati, sfruttando al massimo il concetto di riutilizzo. La possibilità di aggiungere "tag" personalizzati, la scrittura di *action* in *JSP*, e le altre *features* della scenografia completano il profilo tecnico di BS-D.

attraverso l'HP, e in una serie di siti "satellite", che corrispondono ai Dipartimenti in cui è organizzato il Consorzio: graficamente caratterizzati nella testata, gestiti direttamente e in maniera autonoma. Poiché le attività e le competenze del CINECA sono molte e differenti, i contenuti del portale sono stati indicizzati e classificati per garantire una ricerca veloce e mirata. Tre i criteri principali: target d'utenza, argomenti, Dipartimenti di riferimento. In questo modo sono stati classificati anche i SERVIZI offerti e i PROGETTI (ulteriormente suddivisi in nazionali e internazionali) che vedono protagonista il CINECA.

La grafica, semplice ed essenziale, pone l'attenzione sullo spazio contenuti della pagina e sugli strumenti di navigazione. Studiata in maniera coordinata per le pagine istituzionali e per quelle dei singoli Dipartimenti, ripropone ovunque lo stesso schema di navigazione attraverso i menu e gli strumenti di ricerca disponibili.

Il portale propone in apertura le ultime notizie (news) e quelle di Primo Piano, presentando diverse opportunità di consultazione, innanzi-

tutto attraverso il menù posto sotto la testata. Cliccando sulla voce prescelta (macroarea), si apre a sinistra dello schermo l'elenco delle aree e sottoaree disponibili.

Oltre al menu del portale, che riguarda i contenuti istituzionali, ogni Dipartimento offre il proprio menu di presentazione dei contenuti, mentre una MAPPA dettagliata del sito consente la panoramica completa della sua struttura. Ma la navigazione si attua anche attraverso strumenti più rapidi e mirati.

Innanzitutto tramite il CERCA, che consente la ricerca (e la ricerca avanzata) di specifici argomenti all'interno del portale. Poi con la funzione A-Z, che propone un elenco dei contenuti ordinato alfabeticamente e cliccabile.

In alternativa è possibile ricercare i contenuti per target, ovvero seguendo specifici PERCORSI mirati alle diverse tipologie di utenza, che possono così accedere alla consultazione diretta di materiali d'interesse preminente. Al momento sono: Università e Ricerca, Pubblica Amministrazione, Aziende sanitarie e farmaceutiche, Imprese, Professioni, Scuole,

## L'accessibilità e il portale del CINECA

di Elena Mantovani

La legge n. 4 del 9 gennaio 2004 "Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici", definisce gli obblighi per l'accessibilità nella realizzazione e modifica dei siti Internet per le pubbliche amministrazioni.

Nella realizzazione del nuovo portale, il CINECA ha riposto grande attenzione in materia di accessibilità, grazie all'esperienza maturata a partire dall'anno 2001, a seguito della normativa italiana che indicava le linee guida e i criteri per l'accessibilità dei siti web. Attualmente è in corso un'attività di verifica e di valutazione della rispondenza del portale ai criteri di accessibilità delle linee guida del W3C e della recente normativa italiana. La verifica si basa sull'Allegato A del Decreto Ministeriale 8 luglio 2005, pubblicato sulla Gazzetta Ufficiale n. 183 dell'8 agosto 2005.

Il Decreto, a firma del Ministro Stanca, stabilisce le linee guida recanti i requisiti tecnici e i diversi livelli per l'accessibilità e le metodologie tecniche per la verifica dell'accessibilità dei siti Internet, nonché i programmi di valutazione assistita utilizzabili a tale fine. Il testo del decreto si basa sullo studio elaborato dalla "Commissione Interministeriale permanente per l'impiego delle ICT a favore delle categorie deboli o svantaggiate" del CNIPA, con il contributo della Segreteria tecnico-scientifica, di cui il CINECA fa parte con un suo rappresentante tecnico.

Il processo di valutazione della rispondenza del portale ai criteri di accessibilità è articolato in varie fasi: la prima riguarda l'analisi di un campione di pagine significative del portale allo scopo di identificare le caratteristiche generali di accessibilità; la seconda riguarda la verifica tecnica vera e propria articolata in 11 attività distinte.

Dai risultati ottenuti si redigono due rapporti:

- "Verifica della lista dei punti di controllo del WCAG" contiene la verifica di conformità ad almeno il livello A delle linee guida WCAG del W3C così come richiesto dalla direttiva 30 maggio 2002 del Ministro per l'Innovazione e le Tecnologie,
- "Verifica dei 22 requisiti tecnici previsti dal Decreto Ministeriale 8 luglio 2005"

Dai risultati ottenuti si farà un percorso di revisione del portale, valutando cosa fare in sede di prima applicazione e cosa pianificare per l'adeguamento del portale al rispetto della nuova normativa italiana.

Maggiori informazioni: <http://www.w3.org/TR/WCAG10/>; <http://www.pubbliaccesso.it/normative/>.

Editoria. La classificazione dei contenuti permette inoltre di trovare direttamente l'elenco dei SERVIZI offerti, suddivisi per ambito di riferimento.

È possibile, inoltre, accedere direttamente ai siti dei DIPARTIMENTI sfogliando l'elenco proposto dal menu in testata. La loro navigazione ripropone quella del portale istituzionale e ad esso si ricollega. In sostanza, infatti, i contenuti inseriti nel database - sia a livello istituzionale sia a livello di singolo Dipartimento - possono essere richiamati e riproposti agli utenti ogni volta che occorre.

Trasversali ai Dipartimenti sono le aree relative alle opportunità commerciali e alla comunicazione. L'area BUSINESS DEVELOPMENT è un nuovo spazio attivato per consentire di sviluppare, anche tramite web, nuove opportunità di business per quelle che sono le attività più commerciali svolte dal CINECA. In questa area, infatti, vengono presentati in modo omogeneo i servizi commerciali del Consorzio che, essendo erogati da diversi dipartimenti e in diversi ambiti, sono più così più facilmente raggiungibili in modo unitario rispetto alle altre forme di navigazione. Inoltre viene offerta la possibilità al visitatori di interagire con il personale del business development per approfondire rapporti di partnership o per ricevere informazioni commerciali via mail.

Una sezione del sito è dedicata interamente alla COMUNICAZIONE istituzionale (per la stampa ma anche per i visitatori che abbiano la necessità di avere maggiori informazioni sul Consorzio): oltre alle notizie e alle iniziative in programma, in questa area sono disponibili la rassegna stampa e i comunicati stampa emessi, immagini e documenti di approfondimento e il form per iscriversi alla Mailing list, che consente di ricevere nella propria casella di posta le notizie pubblicate sul sito.

Si è scelto di utilizzare la forma del modulo online, dettagliato e articolato, anche per la raccolta delle candidature spontanee per lavorare al CINECA e per l'iscrizione ai corsi. Alla sezione LAVORO, infatti, oltre alla pubblicazione delle ricerche di personale in corso, compilando il form è possibile inserire il proprio curriculum nel Data Base dell'ufficio personale del CINECA e partecipare così alle selezioni periodiche per collaborazioni o stage. Nella sezione CORSI, la pubblicazione



dei prossimi corsi in programma è automatizzata, viene infatti gestita direttamente dai singoli responsabili. In questa area, sempre tramite form, è possibile iscriversi sia ai corsi per i quali sia già stata fissata la data di svolgimento, sia, in forma non vincolante, a quelli in fase di preparazione.

Diversi sono, poi, i contenuti audio/video in streaming pubblicati sul sito: dai convegni trasmessi in diretta su Internet e poi resi disponibili in modalità on demand, ai filmati delle applicazioni realizzate nell'ambito della visualizzazione scientifica, dai video che descrivono il CINECA e le sue attività per immagini fino ai filmati, sempre molto visitati, relativi alla ricostruzione dell'incidente di Ayrton Senna.

Il nuovo portale, pubblicato nella sua prima versione alla fine di marzo, è in continuo sviluppo. Entro la fine dell'anno saranno disponibili le prime funzioni del Portale vocale, e a breve sarà pubblicata la versione in lingua inglese. Stiamo lavorando, inoltre, per raggiungere i massimi standard previsti dalla normativa in materia di usabilità e accessibilità (vedi riquadro nella pagina affianco).

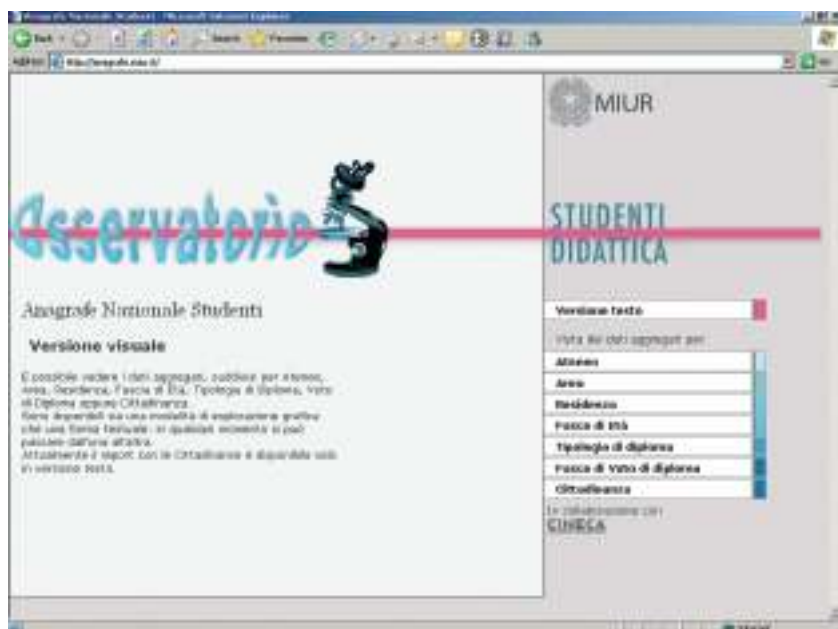
*Per ulteriori informazioni:*  
notizie@ceneca.it

doi:10.1388/notizie-52-01



# Anagrafe Nazionale Studenti

di Luca Donati, Gianna Fabiani, Francesca Pruneti



Nell'aprile del 2004, in seguito al Decreto Ministeriale del 30 aprile 2004 (prot. n. 9/2004), che rendeva operativa l'Anagrafe Nazionale degli Studenti e dei Laureati, il Ministero dell'Istruzione dell'Università e della Ricerca (MIUR), ha ufficialmente dato il via al progetto di raccolta di tutti i dati degli studenti del sistema accademico nazionale che hanno usufruito dell'applicazione della riforma universitaria, attuata con il D.M.509/99.

Il complesso sistema informatico, la cui banca dati viene alimentata direttamente dalle singole università e la cui realizzazione è affidata al CINECA, raccoglierà tutti gli eventi di carriera di uno studente, dal momento del suo ingresso nel sistema universitario fino alla chiusura di quella carriera, registrando nel mentre tutte le attività, che caratterizzano e rendono unica l'evoluzione del percorso di

studio: gli esami sostenuti, le iscrizioni rinnovate,....

L'obiettivo che il MIUR si è posto nella definizione del progetto è quello di rendere disponibile, al Ministero e alle università, uno strumento rapido e semplice da consultare, ma soprattutto sempre aggiornato, tramite il quale sia possibile verificare qual è l'orientamento degli studenti rispetto al gran numero di Corsi di Laurea e di Laurea Specialistica offerti. Inoltre, l'analisi dei dati consentirà di valutare la reale efficacia dei processi formativi attraverso il monitoraggio tempestivo delle carriere degli iscritti ai vari corsi di studio e di monitorare se l'offerta formativa nazionale risponde alle reali richieste, che gli studenti avanzano di anno in anno.

<https://osd.cineca.it>  
(Osservatorio Studenti Didattica)

Per rispondere alle esigenze del Ministero, dunque, il CINECA ha progettato un Data Warehouse, che consente la gestione sia delle problematiche di standardizzazione e omogeneizzazione dei dati (gli atenei utilizzano diversi sistemi informativi per la gestione dei dati, e diversi codici per le classificazioni degli stessi) che delle analisi dei dati a diversi livelli di aggregazione e per molteplici viste. Nella realizzazione di questo sistema, inizialmente è stato dunque necessario predisporre delle codifiche omogenee, e si è scelto di utilizzare le codifiche standard tratte dai sistemi gestionali MIUR, dall'ISTAT o conformi con la normativa ISO.

Successivamente si è passati alla definizione delle strutture dei file che avrebbero alimentato la Staging Area, ovvero l'"area di raccolta"

del sistema, ed alla realizzazione del meccanismo di invio di questi file (in formato txt o xml secondo uno schema XSD opportunamente sviluppato, anche compressi). Infine, sono state create le procedure di validazione dei dati spediti, da applicare prima del caricamento nella Staging Area, e le procedure per il caricamento dei dati nel Data Mart.

Per la parte di “colloquio” tra il nuovo sistema centrale dell’Anagrafe Nazionale Studenti e gli atenei è stato realizzato il sito sicuro <https://osd.cineca.it/> (SSL) con accesso riservato mediante username e password, che consente agli atenei di scaricare i codici da utilizzare per inviare le informazioni richieste e di effettuare l’invio dei file dati.

### **Il meccanismo di raccolta dati**

Per la raccolta dati, il primo passaggio consiste nel caricare i dati nella Staging Area che raccoglie i dati contenuti nei file inviati dalle università, supporta il processo di validazione e, ove sia necessario, la standardizzazione dei dati nel rispetto delle decodifiche di riferimento.

La fase di validazione dei dati avviene immediatamente dopo l’invio (upload) utilizzando delle procedure batch sviluppate ad hoc in linguaggio PHP che consente di avere un unico ambiente per le fasi di acquisizione, validazione e archiviazione dei file dati nella Staging Area.

I dati validati nella Staging Area vengono caricati in una ulteriore base dati per supportare i processi di indagini statistiche e di analisi dati, il Data Mart, che consente due livelli di dettaglio dell’informazione: uno di sintesi con una aggregazione per corso, facoltà e università e uno più analitico con dati di maggior

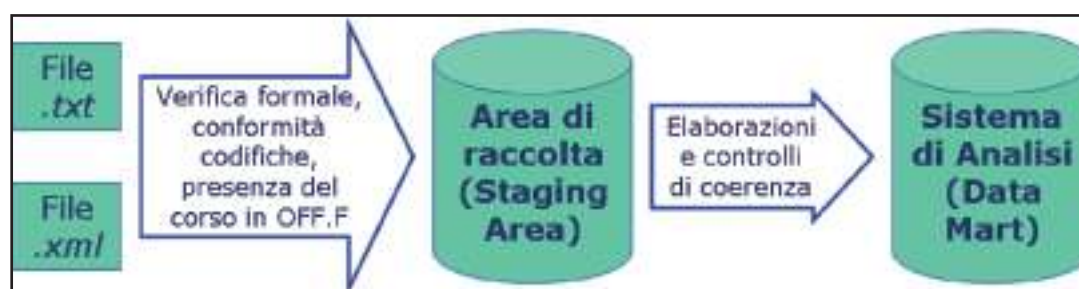
dettaglio, che consente l’analisi fino al singolo studente.

Questa seconda fase di caricamento è stata interamente sviluppata in PL/SQL in modo da minimizzare i tempi di elaborazione e sfruttare appieno le potenzialità del DBMS Oracle. L’insieme della Staging Area e del Data Mart costituisce l’architettura del Data Warehouse dell’Anagrafe Nazionale degli Studenti Universitari.

### **Obiettivi 2004**

Per il 2004 (posta la data ufficiale di chiusura il 31/01/05), visto l’imponente carico di lavoro previsto per gli atenei dall’entrata in vigore del decreto, il Ministero ha deciso di fissare un primo obiettivo: la raccolta degli ingressi nel sistema universitario dei neo-diplomati del 2004 e di tutte quelle persone che non avevano avuto precedenti contatti con il sistema universitario post D.M. 509/99 e che accedevano a corsi di laurea di durata triennale o corsi a ciclo unico. In questo modo il Ministero ha avuto la possibilità di valutare la risposta degli atenei, e questi ultimi hanno avuto modo di riadattare in tempi brevi i propri sistemi informativi per estrarre i dati richiesti, utilizzando le codifiche ministeriali. Il sito <https://osd.cineca.it/> è stato da subito un punto di riferimento per gli atenei, che hanno avuto la possibilità di scaricare codici adottati dal Ministero per comuni, nazioni, atenei, scuole superiori, classi di corso di studio, ambiti disciplinari, etc. Inoltre, per ogni ateneo, CINECA ha messo a disposizione i codici dei corsi di studio, inseriti nella banca dati dell’Offerta Formativa, Off.F, e quindi conformi alle offerte formative annuali, formulate dagli atenei a partire dal 2001/2002.

**La Staging Area raccoglie i dati contenuti nei file inviati dalle università, supporta il processo di validazione e, ove sia necessario, la standardizzazione dei dati nel rispetto delle decodifiche di riferimento**



*Fin dai primi di Settembre 2004, gli atenei hanno iniziato ad utilizzare il sito sia per scaricare i codici, sia per effettuare l'upload dei file di test: l'impegno profuso nell'attività di raccolta delle informazioni è stato testimoniato dalle chiamate e dalle mail che giungevano al Servizio di Consulenza CINECA*

Fin dai primi di Settembre 2004, gli atenei hanno iniziato ad utilizzare il sito sia per scaricare i codici, sia per effettuare l'upload dei file di test: l'impegno profuso nell'attività di raccolta delle informazioni è stato testimoniato dalle chiamate e dalle mail che giungevano al Servizio di Consulenza CINECA. Nel giro di breve tempo il Consorzio è quindi stato in grado di mettere a disposizione sul sito una sezione FAQ e una serie di report che hanno consentito alle università di verificare sia la quantità che la qualità dei dati da loro spediti.

In particolare, il problema della standardizzazione della codifica dei dati è stato molto oneroso per gli atenei che hanno dovuto "convertire" i codici utilizzati nei loro sistemi informativi nelle codifiche ministeriali. Inoltre fino al gennaio 2005, il sistema Off.F è stato riaperto diverse volte, per consentire agli atenei di effettuare correzioni ed adeguamenti dei corsi già inseriti al fine di consentire l'invio di tutti gli studenti effettivamente iscritti. Infine alcune università hanno dovuto aggiornare i propri sistemi informativi per raccogliere le informazioni non gestite.

Per riuscire a comprendere se il sistema poteva essere "accettato" dagli atenei, il Ministero ha stabilito delle date di verifica intermedia, prima della scadenza del 31/01/2005.

Dal primo di settembre 2004 al 31 gennaio 2005:

- sono arrivati all'Osservatorio del CINECA circa 1.350 file (tra quelli di test e quelli definitivi);
- sono stati analizzati e caricati complessivamente 3.300.000 record;
- sono giunte circa 550 mail con richieste di chiarimenti o di supporto;
- entro il 31/01/2005, 73 atenei hanno spedito i dati richiesti.

Nel complesso gli immatricolati a corsi di laurea di primo livello o a ciclo unico, entrati nel sistema Anagrafe Nazionale Studenti, sono stati 339.907.

### Obiettivi 2005

Il successo della prima fase ha indotto il Ministero a pubblicare nel mese di marzo 2005 i nuovi obiettivi 2005, che prevedono la raccolta dei dati completa, ovvero gli atenei devono spedire tutti gli eventi di carriera e devono essere recuperate le carriere avviate prima del 2004/2005, inoltre occorre monitorare anche le immatricolazioni alle Lauree Specialistiche.

La prima scadenza ufficiale di questi nuovi obiettivi è stata il 30/06/2005 ed anche in questo caso la risposta degli atenei non si è fatta attendere. Da febbraio 2005 a fine giugno 2005:

- sono arrivati all'Osservatorio del CINECA circa 3.910 file (tra quelli di test e quelli definitivi);
- sono stati analizzati e caricati complessivamente 13.200.000 record;
- sono giunte circa 500 mail con richieste di chiarimenti o di supporto;
- infine, entro il 30/06/2005, tutti gli atenei hanno spedito i dati richiesti.

### Il Sistema di Analisi

La pronta risposta degli atenei ha consentito al Ministero ed al CNVSU di monitorare in "tempo reale" l'andamento delle immatricolazioni: ad ogni verifica, CINECA aggiornava i dati a loro disposizione con i nuovi invii definitivi ricevuti e con report creati ad hoc con cui potevano controllare l'andamento delle spedizioni. Oltre a questi sistemi di controllo







interni, il Ministero ha richiesto la realizzazione di un sito pubblico di facile navigazione, <http://anagrafe.miur.it>, per mettere a disposizione degli atenei i dati da loro stessi spediti. Attualmente, il sito consente di prendere visione dei dati aggregati relativi agli immatricolati del 2003/2004 e del 2004/2005, suddivisi per “Ateneo”, “Area e Classi”, “Residenza”, “Fasce di Età”, “Tipo di Diploma”, “Voto di Diploma” e “Cittadinanza”.

Il sito è stato realizzato in conformità con la *Legge 9 gennaio 2004, n. 4*, in modo che le informazioni potessero essere disponibili anche alle persone affette da forme più o meno gravi di ipovedenza. Per raggiungere questo scopo, CINECA ha impostato il sito utilizzando l’XHTML ed i criteri di progettazione e scrittura indicati dal W3C.

La parte visuale ha avuto come obiettivo la realizzazione di una facile interfaccia, che mostrasse in maniera chiara i risultati di una qualsiasi navigazione tra i dati, e che desse immediata evidenza della distribuzione degli immatricolati sul territorio nazionale. Questo è stato possibile grazie all’utilizzo di file in formato Flash, che hanno consentito di standardizzare in modo efficace i comportamenti dei vari tipi di browser ricorrendo a plug-in

specifici per la visualizzazione dei risultati, e di utilizzare le funzionalità rese disponibili dal plug-in, ad esempio lo zoom sull’immagine.

### *Prospettive Future*

Gli obiettivi principali dei prossimi mesi di lavoro saranno: l’integrazione del sistema anagrafe studenti con Off.F in modo da semplificare l’aggiornamento dei dati relativi ai corsi da un sistema all’altro; il completamento del sistema di procedure di validazione ed analisi dei dati in modo da ridurre i tempi di caricamento del Data Mart e conseguentemente del sistema di analisi; infine il completamento del sito pubblico in modo da consentire l’analisi dei dati per anni accademici diversi e l’introduzione della possibilità di salvare i dati visualizzati in fogli Excel per analisi off-line.

Per ulteriori informazioni:  
[sito@cineca.it](mailto:sito@cineca.it)

*doi:10.1388/notizie-52-02*

# Un chip per studiare il genoma: i Microarray

di Silvia Giuliani, Elda Rossi

I microarray rappresentano uno strumento di analisi sempre più diffuso nel settore della diagnostica e della ricerca in ambito medico biologico. Si tratta di una tecnologia molto avanzata, capace di produrre dati su migliaia di geni contemporaneamente, e anche molto articolata. Perché possano fornire informazioni biologiche, infatti, i dati prodotti sono soggetti a numerose procedure di analisi e confronto, comportando la gestione di una massiccia mole di dati.

A questo proposito, le nuove generazioni di microarray possono arrivare a contenere alcune centinaia di migliaia di geni, rendendo le normali dotazioni computazionali inadatte a gestire questi dati. Per questo stiamo collaborando con alcuni ricercatori e medici in Italia

per mettere a punto un'infrastruttura che, per caratteristiche, capacità di memorizzazione dei dati e potenza di calcolo permetta di conservare e analizzare dati di microarray.

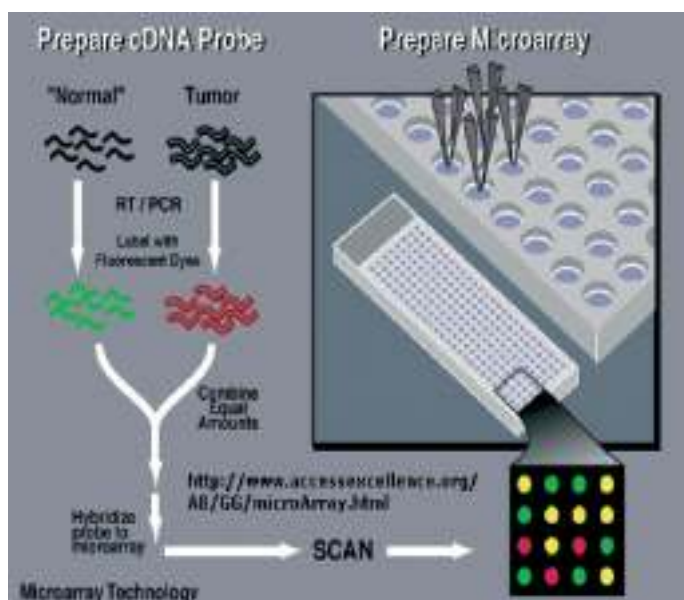
## **Memorizzazione dei dati: scenario di riferimento**

Conservare gli esperimenti di microarray comporta non soltanto l'archiviazione dei dati "raw image" e dei risultati ottenuti nella fase di analisi, ma anche delle informazioni sperimentali necessarie per garantire la riproducibilità ed il confronto tra esperimenti.

A questo proposito, la società MGED (*Microarray Gene Expression Data*) ha realizzato un set di linee guida denominato MIAME (*minimum information about microarray experiment*) che individua le informazioni indispensabili per una corretta e completa annotazione di un esperimento ed un'ontologia apposita per l'impiego di una terminologia comune. È stato inoltre definito un formato standard per lo scambio delle informazioni indicate da MIAME. Si tratta di un linguaggio XML denominato MAGE-ML (*microarray gene expression-markup language*).

Tutte le più autorevoli riviste scientifiche riconoscono il protocollo MIAME e richiedono che i dati citati nelle pubblicazioni siano preventivamente depositati in alcuni repository pubblici di riferimento: Array Express, in Europa, GEO negli USA, e CIBEX in Giappone. Tutti questi repository pubblici sono MIAME compliant.

Figura 1: i geni espressi nel tessuto normale e tumorale vengono spottati sul chip



### ***Rete per la condivisione dei dati per patologie legate al tumore al seno “Hormone Responsive”***

L'idea di una simile infrastruttura nasce dall'esigenza di accedere facilmente a dati di microarray su patologie specifiche (in questo caso tumore della mammella), che sono normalmente depositati nei vari database pubblici oltre che in repository specifici (SMD - database di Stanford). Inoltre, si vuole rispondere anche all'esigenza di archiviare, consultare e scambiare dati prodotti nel corso del progetto dai singoli laboratori.

L'infrastruttura ipotizzata prevede che ogni utente della rete disponga di un archivio personale locale, che verrà utilizzato per conservare i propri dati, oltre che per caricare dati esterni che debbano essere confrontati.

Gli archivi locali saranno collegati tra loro e con un database centralizzato (ospitato su unserver del CINECA), consultabile online, che conterrà i dati scaricati dai repository pubblici oltre che la descrizione degli esperimenti condotti da tutti i partner della rete. Questo permette di rispettare la riservatezza e la pro-

prietà dei dati non ancora pubblicati, pur permettendo una condivisione di conoscenza indispensabile all'interno di un gruppo di lavoro ampio e distribuito.

Inoltre, il database centrale potrà, su richiesta, ospitare aree personali di storage per coloro che non dispongono localmente della strumentazione informatica adatta. A tal fine saranno garantite tutte le precauzioni di sicurezza dei dati e di riservatezza delle informazioni.

Naturalmente deve essere assicurato un facile interscambio dei dati tra tutti i nodi della rete: dai nodi locali si devono trasferire le informazioni sugli esperimenti verso il DB centrale, dal DB centrale si devono scaricare i dati pubblici di interesse sui nodi locali; infine il DB centrale deve collezionare dati dai repository pubblici (vedi Figura 2).

### ***La realizzazione della rete***

Al fine di garantire una tale flessibilità di utilizzo, abbiamo scelto tra i database specialistici esistenti una tecnologia che assicurasse duttilità, portabilità, facilità d'installazione e di manutenzione, sicurezza del dato. Particolare

***L'idea di una simile infrastruttura nasce dall'esigenza di accedere facilmente a dati di microarray su patologie specifiche che sono normalmente depositati nei vari database pubblici oltre che in repository specifici***

### **Microarray**

La tecnica dei microarray nasce negli anni 90 al fine di indagare i livelli di espressione dei geni in una cellula. Essa è in grado di stabilire se un certo gene è attivo in quella cellula, cioè se la proteina corrispondente viene effettivamente prodotta e in quale misura. Oggi è possibile stabilire l'espressione per migliaia di geni contemporaneamente, confrontando allo stesso tempo due diversi tessuti, ad esempio campioni provenienti da pazienti diversi (uno sano e uno ammalato) o dallo stesso paziente (prima e dopo la cura).

Un microarray, o DNA-chip, è un sistema miniaturizzato in cui su un apposito supporto di pochi cm<sup>2</sup> vengono immobilizzati acidi nucleici a sequenza nota, detti sonde o probes, corrispondenti ai geni che si vogliono studiare.

Da un campione biologico si preleva un'opportuna quantità di RNA, che è il prodotto intermedio del processo che dai geni conduce le proteine e che ha la caratteristica di riconoscere le sonde complementari alle quali si lega chimicamente. L'RNA viene sottoposto ad un processo di etichettatura con marker radioattivi o fluorescenti in modo da poterlo rilevare in fase successiva. Infine si procede alla ibridazione, dove la miscela di RNA viene messa in contatto con il microarray in modo che le sue molecole si leghino alle rispettive sonde.

Il microarray viene poi letto da uno scanner, rendendo un'immagine in cui la luminosità o il colore di ogni punto è proporzionale alla quantità di RNA legatosi alla sonda in quel punto. L'immagine è acquisita ed elaborata al computer per produrre dati “raw image”, una misura dell'espressione genica per ogni gene presente sul chip. (figura1)

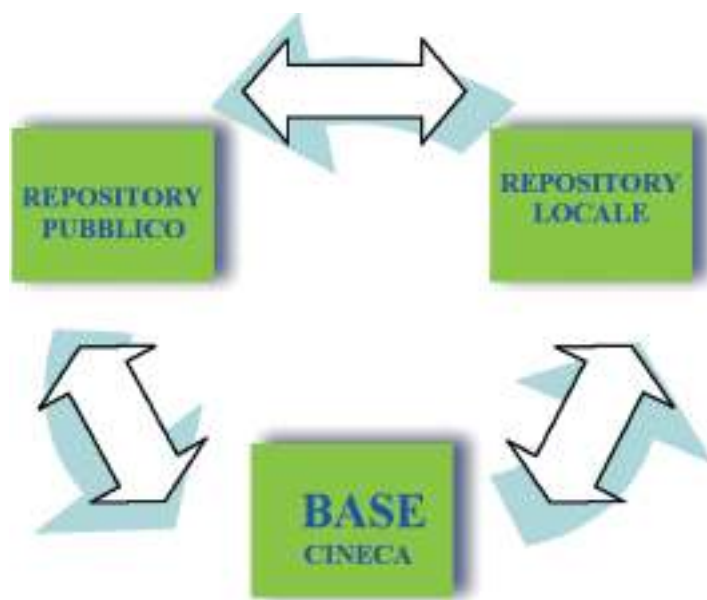


Figura 2: rete per la condivisione di microarray

attenzione, inoltre, è stata posta all'aspetto *open source* e al rispetto degli standard (MIAME e Mage-ML in particolare): la scelta è caduta su BASE.

BASE (*Bio Array Software Environment*) è un sistema per lo storage e l'analisi di dati microarray sviluppato dalla università di Lund in Svezia.

È predisposto per archiviare qualunque tipo di esperimento di microarray (qualsiasi piattaforma) e per una condivisione personalizzata dei dati tra gli utenti.

BASE è stato scelto come repository centrale della rete ed è accessibile via web. Si può accedere liberamente all'area pubblica per consultare dati raccolti dai repository pubblici. È presente poi un'area di progetto ad accesso limitato contenente la descrizione degli esperimenti effettuati dai partner e non ancora pubblicati; infine c'è un'area privata, per la gestione dei propri dati o di dati per i quali si è stati autorizzati.

Per quanto riguarda i nodi locali, BASE è stato proposto, ma non imposto come unica alternativa. Spesso infatti la particolare tipologia dei dati utilizzati presso certi laboratori consiglia database più specializzati. In questo caso vengono approntate procedure di conversione tra il database e BASE nel nodo centrale. Uno degli aspetti più interessanti di BASE è la sua duplice funzionalità, di storage e di ricerca delle informazioni. L'interfaccia uten-

te, MIAME *"compliant"*, propone appositi filtri di ricerca che, sulla base di parametri selezionati, filtrano le informazioni d'interesse.

Un'altra caratteristica risiede nella capacità di BASE di esportare (e, in qualche misura, anche di importare) le informazioni in formato MAGE-ML, così da semplificare l'operazione di pubblicazione sui repository pubblici (tutti MIAME compliant) e l'interscambio con i database locali.

Attualmente, il repository centrale contiene tutti i dati pubblici relativi al tumore alla mammella raccolti dai repository GEO e SMD. L'inserimento, automatizzato tramite una specifica procedura, assicura un aggiornamento continuo dell'informazione. Stiamo lavorando per automatizzare anche l'inserimento dagli altri repository di riferimento.

Alcuni nodi locali sono già stati predisposti e connessi nella rete: due di questi hanno installato BASE, un altro ha preferito un database più specializzato per la piattaforma Affymetrix.

### Prospettive

Questa attività nasce all'interno di un progetto FIRB, "*Hormone Responsive Breast Cancer*". Per questo motivo i dati pubblici collezionati si riferiscono al tumore alla mammella. Nulla vieta però di esportare la competenza acquisita e gli strumenti predisposti per organizzare repository specializzati su altre patologie. Già da ora, l'area privata sul database centrale presso il CINECA è disponibile per chiunque intenda sperimentare questa tecnologia, anche al di fuori del progetto FIRB.

Per ulteriori informazioni:

[info.bio@cinca.it](mailto:info.bio@cinca.it)

BASE:

<http://sirio.cinca.it/HPSystems/Bioinfo/BASE-intro.html>

Progetto FIRB:

<http://www.hrbc-genomics.net/>

BioGate to Microarrays:

<http://www.biogate.it/microarrays/>

doi:10.1388/notizie-52-03



# VisIVO Virtual Observatory e visualizzazione di dati Astrofisici

di Ugo Becciani\*, Marco Comparato\*, Claudio Gheller

\*Osservatorio Astrofisico di Catania

I dati rappresentano oggi un problema centrale per la ricerca scientifica ed in particolare per l'astronomia. Gli strumenti di osservazione (telescopi, satelliti...) producono enormi quantità di immagini e informazioni. I calcolatori e le simulazioni numeriche generano, a loro volta, una mole di risultati altrettanto imponente. Questi dati devono essere immagazzinati, gestiti e analizzati. Ciò richiede un notevole impiego di risorse umane e la disponibilità di strumenti tecnologici capaci, efficienti e potenti. Tutto questo necessita di finanziamenti elevati che, aggiunti al costo degli strumenti di osservazione e dei supercalcolatori, porta ad un enorme e crescente costo del processo di "produzione" della conoscenza. Diventa quindi essenziale che i dati generati da un esperimento, da una serie di osservazioni o da una simulazione, siano sfruttati nel modo più completo, analizzati sotto ogni punto di vista, "spremuti" di ogni informazione utile.

Oggi spesso questo non accade. La mancanza di standard, la difficoltà (tecnologica) di accesso e trasferimento dei dati, la mancanza di strumenti comuni di elaborazione dell'informazione, la proprietà intellettuale del sapere, spesso interpretata in modo troppo restrittivo, costituiscono un limite sostanziale alla possibilità di condividere ed ottimizzare il processo di sviluppo della conoscenza. Di contro, le istituzioni e gli enti finanziatori, vincolano ormai in modo assolutamente imprescindibile al massimo sfruttamento dell'informazione prodotta e del più elevato ritorno dagli investimenti realizzati.

L'insieme di queste motivazioni ha prodotto in

molti campi della ricerca una forte spinta verso la creazione di infrastrutture capaci di garantire la massima diffusione e fruibilità delle informazioni. Da alcuni anni, la comunità internazionale astrofisica lavora alla creazione del Virtual Observatory (VO), definito come:

*«an enabling and coordinating entity to foster the development of tools, protocols, and collaborations necessary to realize the full scientific potential of astronomical databases in the coming decade.»*  
(NVO White Paper, giugno 2000).

La più importante caratteristica del VO è quella di rendere trasparente l'accesso ai dati, nascondendo all'utilizzatore le problematiche legate alla loro locazione fisica, al loro formato e alle modalità di storage, fornendo loro nel contempo gli strumenti per ottimizzare la ricerca di informazioni e l'accesso alle stesse. Lo sviluppo del VO è coordinato dall'International Virtual Observatory Alliance (IVOA, <http://www.ivoa.net>), la cui *mission* è:

*«To facilitate the international coordination and collaboration necessary for the development and deployment of the tools,*

*L'architettura basata su Web Service di VisIVO*



*systems and organizational structures necessary to enable the international utilization of astronomical archives as an integrated and interoperating virtual observatory.»*

L'IVOA si occupa dei vari aspetti legati all'organizzazione, all'accesso e alla manipolazione dei dati, dalla definizione di standard, formati e protocolli, fino allo sviluppo di applicazioni avanzate, quali portali e software di visualizzazione ed analisi, che consentano la più efficiente ed efficace fruizione del dato. È proprio quest'ultimo l'obiettivo del progetto VisIVO (<http://visivo.cineca.it>), sviluppato in collaborazione tra CINECA e Osservatorio Astrofisico di Catania - INAF e inserito nel contesto del progetto VOTech (<http://eurovotech.org/>), finanziato dalla Comunità Europea nel Sesto Programma Quadro.

Lo sviluppo di VisIVO ha l'obiettivo di fornire alla comunità scientifica uno strumento per poter visualizzare e analizzare dati astrofisici (sia osservativi che generati da simulazioni numeriche), con la possibilità di accedere ad archivi distribuiti nell'infrastruttura VO senza doversi preoccupare delle specifiche problematiche tecniche che tale accesso comporta. VisIVO costituisce il primo di una serie di strumenti forniti dalla comunità di ricerca italiana e dal CINECA, nel contesto della creazione di un servizio di archivio pubblico di dati teorici (Theoretical Virtual Observatory, TVO), finalizzato a rendere disponibili i più importanti risultati prodotti da simulazioni. Particolare attenzione verrà

dedicata ai dati prodotti nel contesto della convenzione INAF-CINECA.

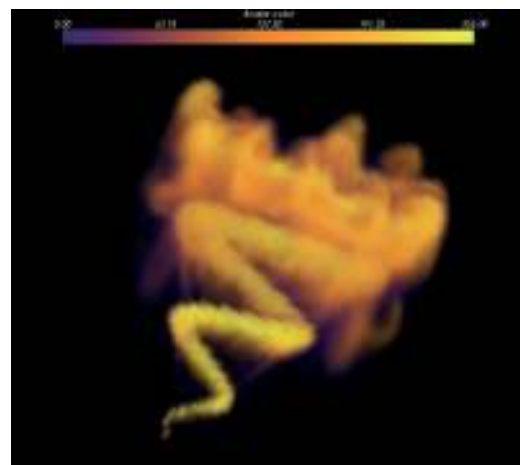
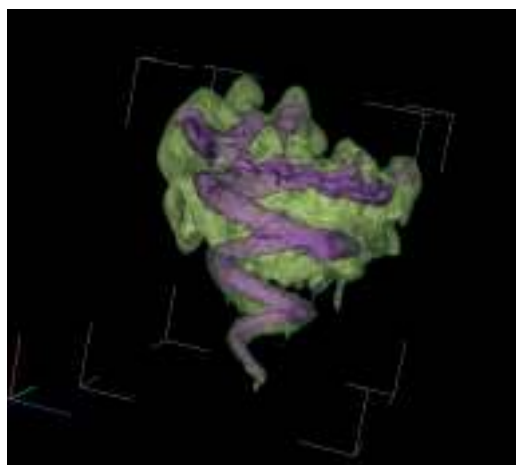
## VisIVO

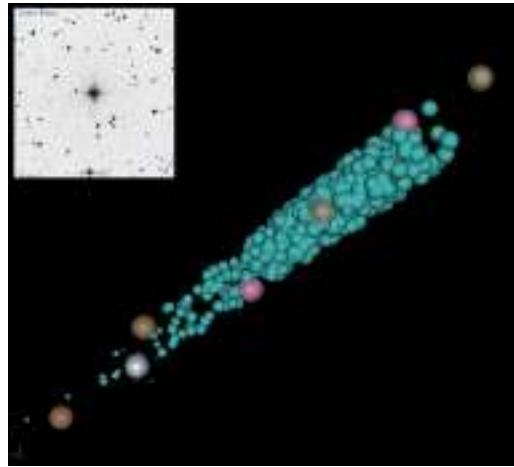
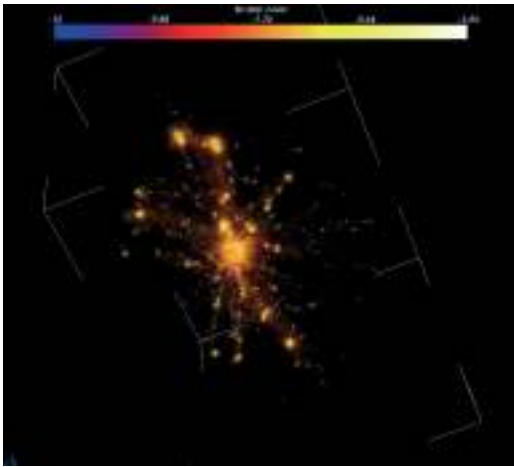
VisIVO, acronimo di “*Visualisation Interface for the Virtual Observatory*”, è un software specificamente pensato per la visualizzazione multidimensionale e l'analisi di dati astrofisici. Esso nasce dall'esperienza acquisita nel progetto Cosmo.Lab (finanziato dalla Comunità Europea nel Quinto Programma Quadro – <http://cosmolab.cineca.it>), che ha dato origine ad AstroMD, software che può esserne considerato il progenitore. La sua struttura è basata sul framework OpenMAF (<http://www.cineca.it/sap/area/maf.htm>), sviluppato e distribuito dal gruppo di visualizzazione e grafica del CINECA.

VisIVO può operare sia su dati provenienti da simulazioni, sia su dati osservativi. I dati possono essere letti da file memorizzati sul disco locale e sono supportati i formati di file più comuni: dal semplice formato *Binary Raw* sino a standard più complessi quali *HDF*, *FITS*, *Topsy* e, soprattutto, *VOTable*, lo standard sviluppato e supportato da IVOA, basato su XML. Per ciascuno di questi formati, sono stati sviluppati specifici reader dotati di interfaccia grafica che consente una gestione efficiente ed intuitiva dei relativi file di dati. Nuovi reader sono in fase di sviluppo sulla base delle richieste e delle esigenze della comunità scientifica.

La caratteristica più interessante di VisIVO è sicuramente rappresentata dalla possibilità di

*Visualizzazione di dati su griglia (simulazione di un jet rotante): nella figura di sinistra attraverso isosuperfici, nella figura di destra attraverso volume rendering*





accedere ad un database remoto, nodo del Virtual Observatory, in modo del tutto trasparente per l'utente finale. L'accesso è basato sulla tecnologia dei *Web Service* (WS). VisIVO supporta, inoltre, diverse tecniche di visualizzazione. I dati su griglia possono essere rappresentati attraverso il *Volume Rendering* (visualizzazione a "nuvola", con colori e trasparenze che cambiano in funzione del valore del campo in esame), oppure mediante superfici di livello (superfici che dividono le celle di valore più alto da quelle di valore più basso di una determinata soglia). Tuttavia, il software è più specificatamente destinato al trattamento di dati puntiformi, comuni in ambito astronomico (dai cataloghi – stelle, galassie... – alle simulazioni – N-Body, SPH...). I punti possono essere rappresentati come pixel, oppure attraverso forme geometriche tridimensionali (sfere, cubi...). Ad essi possono essere associati campi scalari (densità, temperatura), visualizzabili attraverso opportune scale di colori oppure parametrizzando ai relativi valori le dimensioni delle forme stesse. I punti possono essere selezionati o campionati in modo casuale. Possono venire associate informazioni in formato html, utilizzando il browser integrato nel software. Infine è supportata la visualizzazione stereografica, che abilita VisIVO all'utilizzo in contesti di Realtà Virtuale.

### Strumenti di Analisi

VisIVO integra una serie di strumenti di analisi delle proprietà di dati puntiformi. Il *Point Splatter* permette di distribuire una

quantità associata al singolo punto (ad esempio, la massa) su una griglia di calcolo di supporto. Il risultato può venire poi rappresentato mediante tecniche di visualizzazione dei dati su griglia.

La *Funzione di Correlazione* e il *Power Spectrum* rappresentano due strumenti standard di analisi statistica della distribuzione di punti. Entrambi descrivono il grado di correlazione tra i punti su diverse scale.

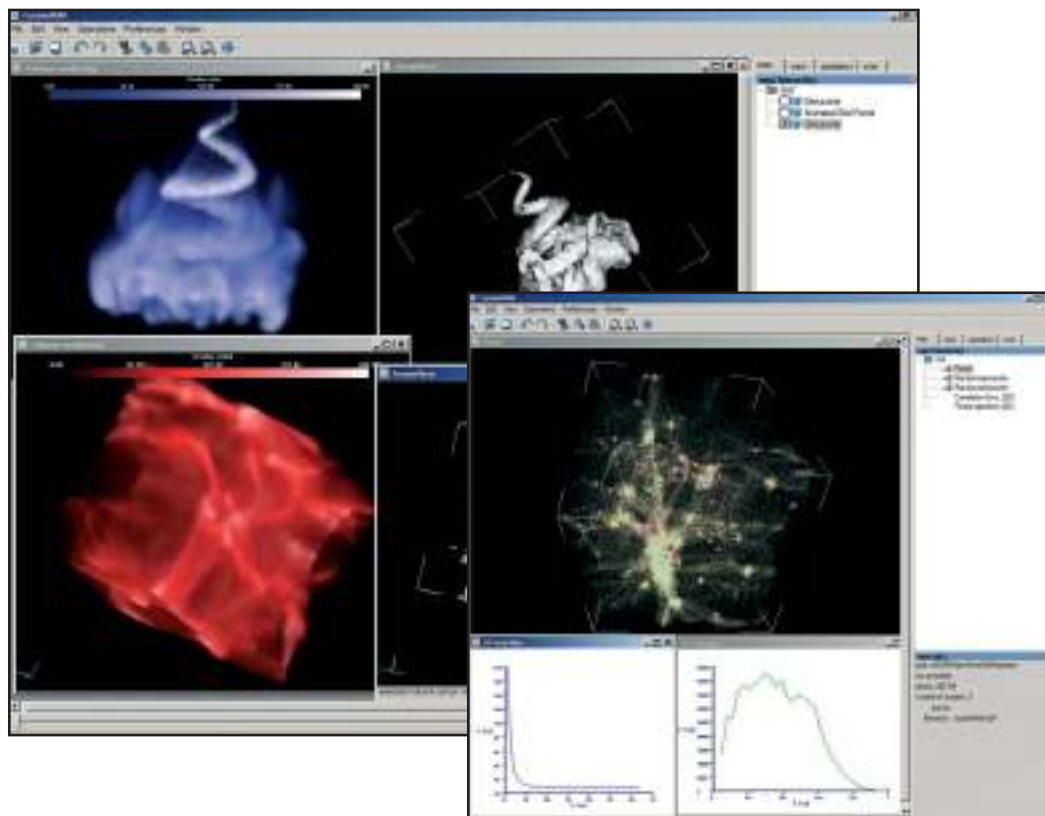
I *Funzionali di Minkowsky* forniscono stime quantitative sulla geometria (forma) e topologia (connettività) della distribuzione di punti. Queste funzioni di analisi sono state ereditate dal lavoro sviluppato per il progetto Cosmo.Lab, inoltre è attualmente in corso l'integrazione di AstroNeural, un potente strumento di *data mining* basato su reti neurali e sviluppato dalle università di Napoli e di Salerno (<http://astrofisica.na.infn.it/Primolivello/Italia/ricerche/astroneural.htm>). Funzioni più complesse, che richiedono grandi risorse di calcolo, verranno sviluppate esternamente a VisIVO, in modo da poter essere installate sui sistemi HPC del TVO e controllate attraverso la tecnologia basata su WS.

### Collaborazioni e futuri sviluppi

VisIVO, nelle sue versioni preliminari, è già stato presentato a vari congressi e workshop suscitando grande interesse da parte della comunità scientifica internazionale. È stato selezionato come strumento di visualizzazione per il *Virtual Observatory* nell'ambito del progetto *VOTech*. Inoltre, sono state attivate varie collaborazioni con enti nazionali ed

*Figura a sinistra: visualizzazione di dati puntiformi (simulazione di un ammasso di galassie) attraverso scale di colori*

*Figura a destra: sottoinsieme di stelle estratte dal catalogo di Hipparcos. Le differenti dimensioni delle sfere rappresentano la luminosità delle stelle. L'immagine nell'angolo in alto a sinistra rappresenta l'immagine reale (al telescopio) della stella evidenziata con il cubo*



*Figura a sinistra:  
la GUI di VisIVO e alcune  
modalità di visualizzazione  
dei dati*

*Figura a destra:  
strumenti di analisi dati  
integrati in VisIVO*

internazionali per favorirne lo sviluppo ed arricchirne le funzionalità: oltre alla già citata collaborazione con le Università di Napoli e Salerno, vi sono collaborazioni con l'Osservatorio Astronomico di Trieste per la creazione del prototipo di TVO e l'inclusione dei dati simulati del progetto Planck e con l'Osservatorio Astronomico di Strasburgo (CDS) per l'interfacciamento al Web Service, già menzionato in precedenza. Inoltre è in fase di sviluppo l'integrazione di VisIVO con il software di analisi e visualizzazione di immagini astronomiche "Aladin" prodotto e distribuito dallo stesso CDS. Interesse e proposte di collaborazione, in via di definizione, vengono poi dall'Osservatorio Astronomico di Edimburgo, dal MPI di Monaco nell'ambito del German Astronomical Virtual Observatory e dall'Australian Virtual Observatory.

VisIVO è realizzato completamente con strumenti Open Source e può essere scaricato gratuitamente assieme alla relativa documentazione, dal sito web o via accesso anonimo al

server CVS del CINECA. Attualmente è disponibile la versione 0.1. Tale versione è compilata esclusivamente sotto MS Windows. Tuttavia il porting in ambiente Linux (privilegiato dalla comunità scientifica) è in fase di completamento e sarà parte della nuova distribuzione, prevista prima della fine dell'estate. Un ringraziamento speciale va al gruppo di visualizzazione del CINECA che, oltre a mettere a disposizione lo strumento base, il MAF, su cui VisIVO è realizzato, fornisce il supporto necessario a realizzare uno strumento efficace ed apprezzato.

---

*Per ulteriori informazioni:*

astro@cineca.it

<http://visivo.cineca.it>

doi:10.1388/notizie-52-04



# Un TTS per l'assistenza ai clienti dei sistemi gestionali per le Università

di Luciano Canino, Stefania Caponetti

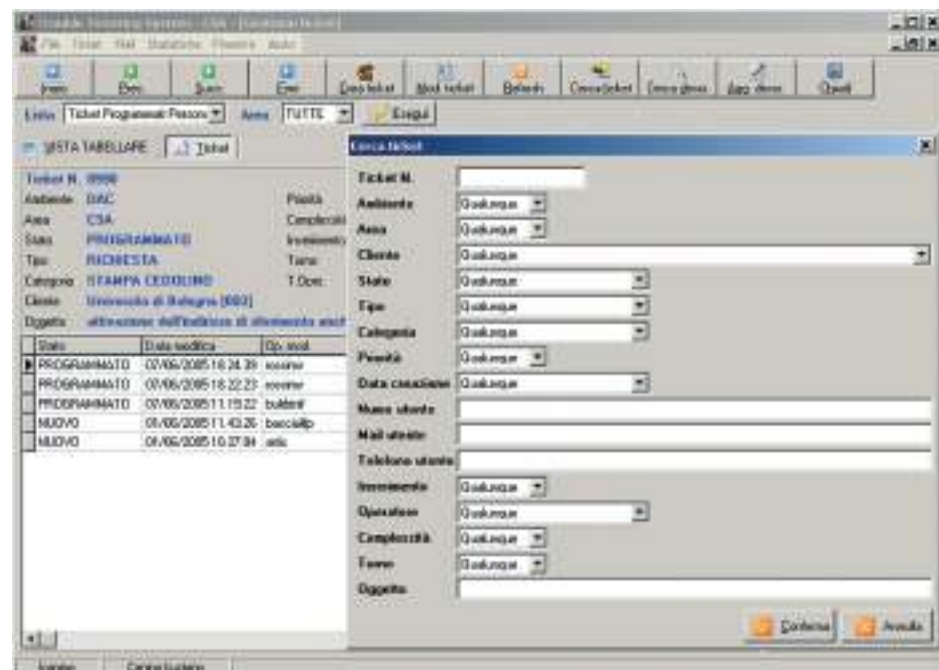
Da oltre un anno, presso il Dipartimento *Gestionali per l'Università*, è stato attivato un sistema di archiviazione delle richieste di supporto al servizio di consulenza dei Clienti dei sistemi CIA (Contabilità Integrata d'Ateneo) e CSA (Carriere e Stipendi d'Ateneo).

L'introduzione di un *Trouble Ticketing System* (TTS) che consentisse la registrazione e il trattamento delle richieste di supporto provenienti dal cliente era ormai divenuto indispensabile sia per l'incremento del numero dei clienti che usufruiscono dei servizi CIA e CSA, sia per poter disporre di uno strumento capace di misurare la qualità del servizio di assistenza rispetto gli indicatori stabiliti dalle procedure definite per il settore per ottenere la certificazione ISO 9001.

Le caratteristiche cui, in generale, un TTS deve rispondere dipendono essenzialmente dalle possibili modalità di richiesta di assistenza disponibili, e dalla scelta delle modalità di notifica ai clienti dello stato del problema. In particolare, sia per CIA sia per CSA le caratteristiche del servizio sono molto simili, i clienti possono inoltrare la segnalazione chiamando il servizio di assistenza telefonica, lasciando un messaggio nella casella vocale, inviando un fax oppure un messaggio di posta elettronica al servizio di consulenza. Le informazioni sullo stato del problema vengono comunicate al cliente via e-mail e telefono. Nonostante il mercato offrisse una ampia gamma di sistemi in grado di soddisfare le esi-

genze richieste dal servizio, si è scelto di sviluppare un sistema ad hoc. Questa scelta è stata dettata dall'importanza di disporre solamente degli strumenti necessari per il servizio ed inoltre dalla necessità di avere per questi una completa autonomia nel dettare la modalità di funzionamento, molto simili alle modalità operative prefissate dal servizio di consulenza.

Dal punto di vista operativo, ad ogni problema segnalato, indipendentemente dallo strumento di comunicazione scelto dal cliente, il servizio di consulenza fa una attenta analisi della



**La registrazione di ogni attività di consulenza consente di arricchire la knowledge base del servizio**

richiesta e la registra assegnando oltre ad un numero identificativo (ticket), anche una classificazione del problema ed in particolare:

- i riferimenti del cliente;
- il tipo di segnalazione (assistenza, errore, richiesta...);
- il livello di priorità;
- la categoria (calcolo, stampa cedolino, giuridica ...) e la risorsa umana del dipartimento che agirà sul problema, con successivo invio della notifica di generazione del nuovo ticket via e-mail sia all'assegnatario e sia al cliente che ha fatto la segnalazione.

Il problema è caratterizzato da uno specifico

stato: *nuovo, aperto, chiuso, sospeso, programmato*. La generazione di un ticket da parte del servizio di consulenza pone il ticket nello stato di *nuovo*. I collaboratori del dipartimento dispongono delle credenziali d'accesso al sistema ed ogni attività effettuata sui ticket viene registrata ed in questa fase si può decidere di notificarne il cambiamento di stato via e-mail al cliente.

Eventuali successive comunicazioni del cliente sul ticket vengono inserite a integrazione della problematica ed eventualmente possono cambiarne lo stato.

La registrazione di ogni attività di consulenza consente di arricchire la *knowledge base* del servizio.

### Scheda

di Maria Vittoria Rossi

Il CINECA si è proposto di formalizzare le procedure di svolgimento dei propri servizi al fine del rilascio della certificazione di qualità in conformità alla normativa UNI ISO 9001.

La qualità di un servizio di assistenza ai clienti si misura sostanzialmente attraverso il tempo di risposta alle richieste pervenute, è quindi indispensabile uno strumento di registrazione di tali richieste per avere una misura puntuale e affidabile.

Il sistema di TTS inserisce le segnalazioni in archivio con stato 'Nuovo'; l'operatore incaricato varia lo stato in 'Chiuso' quando risponde all'utente. La differenza dei tempi tra lo stato 'Nuovo' e 'Chiuso' costituisce il tempo di risposta.

La definizione della procedura ha stabilito gli indicatori di qualità da raggiungere per qualificare positivamente il servizio di assistenza erogato. Sono stati previsti indicatori diversi a seconda della tipologia e della priorità assegnati alla segnalazione.

Per la procedura CSA gli indicatori sono i seguenti:

- tipo Assistenza (supporto all'uso):
  - l'obiettivo è fornire risposta entro 1 giorno lavorativo.
- tipo Consulenza (verifiche, approfondimenti): l'obiettivo è fornire risposta:
  - entro 2 giorni lavorativi se la Priorità è Media/Alta;
  - entro 5 giorni lavorativi se la Priorità è Bassa.
- tipo Errore:
  - l'obiettivo è rispondere in merito ai tempi di soluzione entro 3 giorni lavorativi; la soluzione sarà realizzata in tempi da concordare a seconda della complessità e della priorità del problema.
- tipo Intervento su DB:
  - l'obiettivo è rispondere sulla fattibilità entro 5 giorni lavorativi.

Gli obiettivi si intendono raggiunti se lo sono almeno sull'80% dei casi.

Dal punto di vista organizzativo, in fase di generazione dei ticket, il TTS permette l'assegnazione di compiti alle risorse umane del dipartimento. Attraverso il TTS ogni risorsa è in grado di stabilire il carico di assistenza quotidiano mentre il Service Management è sempre in grado di "misurare" il carico d'assistenza su cui far fronte nei vari periodi dell'anno o del mese e quali categorie hanno un carico di assistenza più oneroso; ha la possibilità di sincronizzare le attività dei collaboratori, consentendo loro di essere sempre aggiornati riguardo alle problematiche e ai compiti in corso relativi ai singoli clienti, oltre che misurare la qualità del servizio rispetto agli opportuni indicatori stabiliti dalla procedura ISO secondo le norme previste dal Piano della Qualità.

### Le statistiche

Il *Service Management* ha la possibilità di interrogare il sistema informativo attraverso delle funzioni ad hoc. In particolare, per la certificazione ISO 9001 del *Servizio di Assistenza* è indispensabile disporre di statistiche che forniscano delle "misure" sui tempi di risposta del servizio. I parametri principali di riferimento per le statistiche sono il tipo di richiesta e la priorità della segnalazione. L'applicazione fornisce dei report, con le seguenti informazioni:

- misurazione della qualità del servizio. Attraverso l'analisi dei tempi di risposta al cliente dei ticket, inteso come differenza tra generazione e chiusura del ticket, vengono prodotte delle statistiche che calcolano il numero di ticket che rispettano il target imposto per rispettare gli obiettivi di qualità definiti nella procedura del manuale;
- misurazione del carico del servizio. Attraverso l'analisi dei tempi di presa in carico dei ticket, inteso come differenza tra generazione e apertura del ticket, vengono prodotti dei report che contengono il carico d'assistenza distribuito tra i vari collaboratori rispetto alle varie categorie cui il ticket è classificato;
- calcolo dei tempi reali di soluzione rispetto

TTS CSA Gennaio - Maggio 2005

Indicatore	n.ticket	% vs obiettivo
Assistenza	2900	82,79
Consulenza Media/Alta	134	88,81
Consulenza Bassa	1880	92,66
Errore	83	92,77
Intervento su DB o Richiesta	93	91,40
Priorità critica	4	100,00

ai vari attributi che caratterizzano il ticket;

- conteggio delle segnalazioni pervenute al servizio rispetto ai vari attributi che caratterizzano il ticket.

### Conclusioni

In futuro una richiesta di assistenza potrebbe arrivare dal portale del CINECA attraverso la compilazione guidata di un modulo web. La completa redazione del modulo genererà automaticamente un nuovo ticket.

La presenza del TTS ha sicuramente incrementato il livello dei servizi, ha migliorato la qualità del servizio ed ha aumentato l'efficacia operativa delle risorse umane del Dipartimento.

La soluzione adottata si inserisce nell'ottica di dotarsi di sistemi di Service Management che consentano al CINECA di automatizzare e gestire in maniera ottimale servizi interni ed esterni e procedure di supporto per l'organizzazione di prossima integrazione con il CRM (*Customer Relationship Management*) aziendale.

**Il Service Management ha la possibilità di interrogare il sistema informativo attraverso delle funzioni ad hoc. In particolare, per la certificazione ISO 9001 del Servizio di Assistenza è indispensabile disporre di statistiche che forniscano delle "misure" sui tempi di risposta del servizio**

Per ulteriori informazioni:  
info.csa@cineca.it

doi:10.1388/notizie-52-05

# OSIRIS e l'evoluzione dell'e-learning

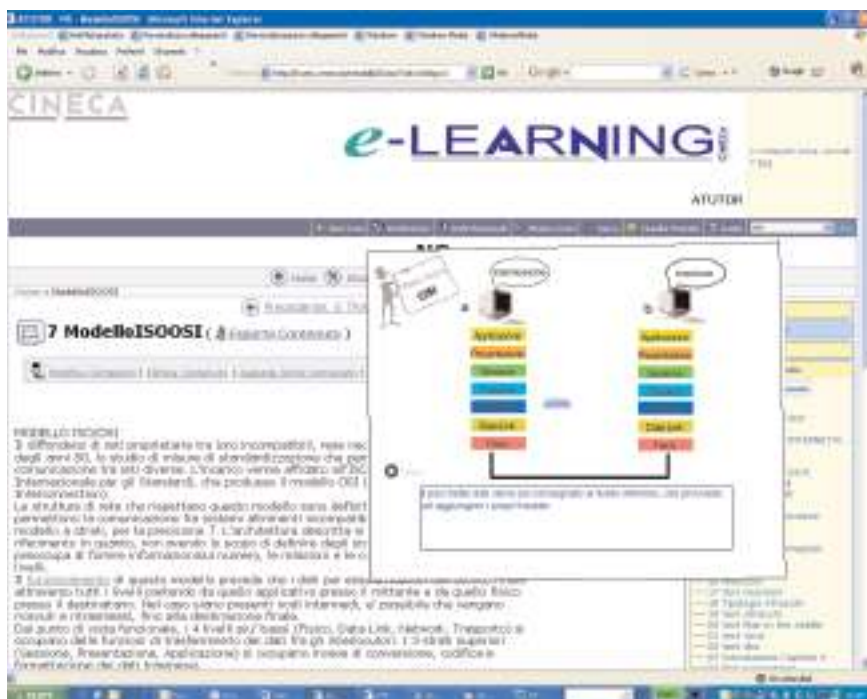
di Eugenia Rinaldi

I recenti sviluppi della tecnologia open source hanno dato origine a piattaforme di e-learning molto evolute, sempre più spesso conformi agli standard internazionali. Segnali di questa tendenza hanno iniziato ad essere evidenti da circa un paio di anni. Per questa ragione, IRIS (*Integrated Resources for Internet Students*), la soluzione CINECA nata per offrire soluzioni complete nel campo dell'e-learning, ha previsto sin dall'inizio la possibilità di integrare al proprio interno strumen-



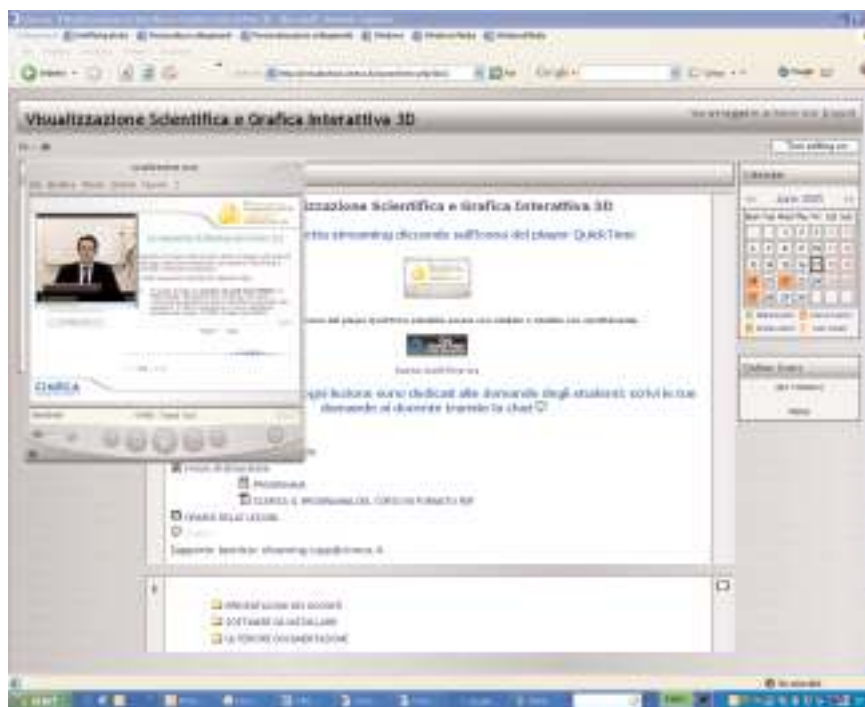
ti *open source*. IRIS è basato sull'integrazione di componenti di elevata qualità tecnologica e aderenti agli standard di e-learning internazionali. Oggi i software liberi si sono meritatamente imposti nel mondo della didattica online. CINECA ha interpretato queste

nuove esigenze e tendenze integrando nella soluzione IRIS degli LMS open source con cui eventualmente interfacciare componenti commerciali accessorie. È nato così OSIRIS (*Open Source IRIS*). L'idea su cui si fonda OSIRIS è quella di sfruttare la tecnologia open source più evoluta (annullando così i costi delle licenze software) e di concentrare gli sforzi su un'alta personalizzazione del servizio, sempre nel rispetto degli standard internazionali. La disponibilità del codice sorgente dei software fa sì che esso possa essere modificato o integrato nella maniera più consona alle esigenze del richiedente. OSIRIS preserva tutte le caratteristiche funzionali di IRIS per la realizzazione di corsi in modalità e-learning, dalla creazione del materiale fino all'erogazione e gestione dei corsi e delle classi via web. Al momento, i software open source che fanno parte di OSIRIS e su cui CINECA offre servizi di vario genere (hosting, manutenzione, formazione, personalizzazioni applicative notizie dal CINECA e grafiche) sono: Moodle; Atutor; Docebo; .LRN; Claroline. Queste piattaforme risultano fra le più diffuse a livello mondiale.





Il software utilizzato diventa la base su cui offrire servizi di qualità; la continua evoluzione della tecnologia, unita alla progressiva evoluzione e standardizzazione dei prodotti, fa sì che questa lista sia tutt'altro che chiusa o statica. Viene così a mancare la dipendenza da un'unica piattaforma o fornitore. Dal punto di vista didattico e metodologico, questo approccio open source aiuta a spostare il centro dell'attenzione dalla tecnologia al contenuto. È importante sottolineare che per potere rendere efficace la formazione online non basta una buona piattaforma, ma è necessaria una organizzazione dei contenuti mirata a sfruttare i vantaggi del formato digitale. Ciò che risulta molto interessante, inoltre, è la possibile integrazione di OSIRIS con altri sistemi informativi eventualmente già presenti all'interno di un Ente. L'importanza dell'integrazione è evidenziata in particolare dagli studi della OKI (*Open Knowledge Initiative*), un'iniziativa avviata dal MIT, *Massachusetts Institute of Technology*, nel 2001, con lo scopo di standardizzare le interfacce di comunicazione tra software e sistemi nell'ambito dell'e-learning, anche se appartenenti ad istituzioni diverse. L'OKI ha rilasciato del codice open source per l'applicazione delle OSID (*Open Service Interface Definition*) per raggiungere interoperabilità a livello di servizio. Le OSID definiscono come le componenti di un ambiente didattico possano interagire. OKI concentra dunque i propri sforzi non sullo sviluppo di singole componenti dell'ambiente didattico, bensì sullo sviluppo di interfacce specializzate che permettano l'integrazione globale dei sistemi. Al giorno d'oggi, le funzionalità interne agli LMS sono tecnologicamente molto avanzate, ma non sempre vengono interamente utilizzate dai docenti. In particolare, in Italia, soprattutto in ambito universitario, l'e-learning è spesso utilizzato a supporto e non in sostituzione della didattica tradizionale e questo fa sì che molte delle potenzialità dei sistemi di e-learning superino le reali necessità dei docenti. I docenti prediligono poche funziona-



lità, ma di facile accesso ed uso. A tal riguardo, è opportuno notare che ora gli sforzi delle comunità informatiche si dirigono verso le integrazioni dei servizi online piuttosto che su un'ulteriore evoluzione dei singoli LMS. È in questo contesto che si colloca l'idea di OSIRIS: una piattaforma e-learning sempre aperta all'evoluzione tecnologica che possa essere personalizzata ed eventualmente integrata in portali che già gestiscono altre banche dati e informazioni.

Per ulteriori informazioni:  
osiris@cenea.it

doi:10.1388/notizie-52-06

# Riabilitazione a distanza: un progetto sperimentale per la TV interattiva

di Anna Conti\*, Andrea Venturi  
\*psicologa dell'età evolutiva

L'impresa questa volta è titanica, vogliamo dare alla televisione la chance di fare del bene... Il mondo della televisione negli ultimi anni è stato "sconvolto" dall'introduzione della tecnologia digitale, ancora più dirompente, specialmente in Italia, per la presenza di una componente interattiva che sicuramente, se avrà successo, porterà a una rivoluzione del modo di rapportarsi al mezzo televisivo. Allo scopo di incentivare l'esplorazione di nuove applicazioni su questa tecnologia, lo scorso anno la Fondazione Bordini ha lanciato un bando di selezione per applicazioni ad alta interattività.

Al CINECA avevamo già cominciato a lavorare sulla definizione e sulla implementazione della strategia con cui affrontare questa "convergenza digitale" dei vari settori delle telecomunicazioni multimediali (si veda ad esempio l'articolo "*Televisione Digitale Terrestre: la televisione diventa interattiva*" e "*CreaTV la Televisione Digitale interattiva secondo il CINECA*" pubblicati rispettivamente sui numeri 50 e 51 di questo *Notiziario*) ma, a fronte di una richiesta esplicita, da parte della Fondazione Bordini, di realizzare un servizio sperimentale ad alta interattività, per dimostrare che questa tecnologia può essere vincente, si è aggiunta anche la fortuita coincidenza di poter contattare il professor Giacomo Stella, noto psicologo, già professore dell'Università di Urbino, che ha aggiunto importanti affermazioni nel campo della diagnosi e della riabilitazione dei disturbi di apprendimento del linguaggio, specialmente in relazione ai soggetti in età evolutiva. Attualmente, la riabilitazione nei soggetti a rischio dislessia viene erogata in maniera estremamente frammentata e dispendiosa per varie ragioni: in primo luogo è difficile effettuare a costi ragionevoli lo screening sulla popolazione interessata (rappresentata dai bambini durante il primo anno di scolarizzazione) e spesso, infatti vengono individuati solo i casi più eclatanti. Inoltre, il numero di centri sparsi sul territorio nazionale è esiguo, quindi, spesso, pazienti diagnosticati rimango-



no senza possibilità riabilitative efficaci in quanto il centro diagnostico risulta irraggiungibile dal luogo di residenza. Infine, le tecniche di riabilitazione utilizzano ancora modalità di somministrazione attraverso apparati specializzati (benché già basati su tecnologie informatiche) presenti solo nei suddetti centri, per cui la disponibilità del servizio è fortemente limitata da oggettivi fattori di accessibilità. Le sperimentazioni che hanno riguardato il tentativo di utilizzare a distanza queste applicazioni in un contesto *PC+Internet*, purtroppo, hanno incontrato difficoltà di varia natura quali, innanzitutto, la ancora scarsa diffusione di tali strumenti nelle case delle famiglie italiane; il complicato sistema di interfaccia di input (mouse/tastiera) per le possibilità di un bambino di sei anni con disturbi di apprendimento, la mancanza di garanzie di QoS nelle reti internet attuali, il costo della banda larga ecc...

Insieme anche ad altri due importanti Enti, INDIRE (*Istituto Nazionale di Documentazione per l'Innovazione e la Ricerca Educativa*, con sede a Firenze) e il CRESEM (*Centro Radioelettrico Sperimentale Guglielmo Marconi*, con sede a Trieste), e con la completa disponibilità di Sestarete, la più importante e diffusa emittente regionale in Emilia Romagna, che da lungo tempo collabora con CINECA per esplorare le potenzialità della tv digitale, abbiamo quindi proposto un innovativo progetto che tratta la riabilitazione dei bambini affetti da problemi di dislessia attraverso il sistema di televisione digitale interattivo.

La nostra partecipazione al bando è stata coronata da successo in quanto il progetto è stato ritenuto meritevole di cofinanziamento, raggiungendo il quinto posto in graduatoria (su diciotto progetti presentati e sei progetti finanziati). Il progetto prevede di identificare un certo numero di soggetti in età evolutiva a rischio di disturbo nelle abilità di letto-scrittura e, attraverso la nuova televisione digitale interattiva, di somministrare loro un insieme di esercizi di riabilitazione, sotto la supervisione remota di un pool di psicologi, per un periodo di quattro mesi negli agi del loro

domicilio, e verificare poi con metodo scientifico quanto statisticamente ha inciso, in positivo, questo strumento nel miglioramento della loro performance.

Nella prima fase, intorno al mese di gennaio del 2006, ci proponiamo di somministrare a un gruppo di bambini del primo anno delle scuole elementari di Bologna un test informatizzato già disponibile, realizzato da Asphi, che consente di identificare soggetti a rischio per disturbi di apprendimento della letto-scrittura (dislessia e disortografia).

Durante il successivo mese di febbraio, a questo gruppo di bambini ritenuti "a rischio", denominato gruppo sperimentale, verrà assegnato un decoder digitale terrestre interattivo e un dispositivo di identificazione personale.

Per il periodo dei quattro mesi successivi, tra marzo e giugno 2006, i bambini effettueranno esercizi giornalieri di riabilitazione utilizzando una applicazione di test erogata attraverso il decoder digitale terrestre interattivo.

L'applicazione di riabilitazione, in sostanza, si presenterà come una serie di giochi di abilità (si vedano a titolo di esempio le immagini nella pagina seguente) inseriti in un percorso narrativo coinvolgente e gratificante in modo da sollecitare la continuità dell'applicazione da parte dei pazienti.

*L'applicazione di riabilitazione, in sostanza, si presenterà come una serie di giochi di abilità inseriti in un percorso narrativo coinvolgente e gratificante in modo da sollecitare la continuità dell'applicazione da parte dei pazienti*





*Questo progetto intende sperimentare la validità della piattaforma digitale terrestre interattiva nell'erogare questa tipologia di servizi di telemedicina, attraverso la verifica dell'effettivo miglioramento dei soggetti presi in esame*

Durante l'arco della riabilitazione attraverso il digitale terrestre, ogni decoder, previa identificazione certa ma anonima del mittente, invierà le informazioni relative alla riabilitazione al centro servizi, dove un sistema informatizzato di gestione della cartella clinica provvederà ad allertare il gruppo di supervisori in presenza di eventi anomali segnalati da sistemi di allarme (*trigger*).

Al termine del periodo di riabilitazione verrà somministrato un test informatizzato simile a quello sostenuto in fase di ammissione al fine di determinare il livello di efficacia della riabilitazione attraverso questa nuova piattaforma basata su tv digitale interattiva.

Sulla base dei dati di monitoraggio della piattaforma e dei risultati ottenuti dai soggetti sperimentatori, sarà possibile trarre una serie di indicatori di performance che diano il via libera a una somministrazione di questa piattaforma su scala più ampia.

Questo progetto intende sperimentare la validità della piattaforma digitale terrestre interattiva nell'erogare questa tipologia di servizi di telemedicina, attraverso la verifica dell'effettivo miglioramento dei soggetti presi in esame e la dimostrazione che il rapporto costo-prestazioni di questa infrastruttura di esercizio è migliore della stessa attività di riabilitazione, sia nella tradizionale modalità di erogazione in studio, sia anche del caso di utilizzo di metodi alternativi avanzati come già

citato PC+Internet.

Verrà effettuata anche una analisi dei costi necessari a trasformare il prototipo in un sistema in grado di esercitare la sua attività su scala nazionale, partendo dal presupposto che una larga parte dell'infrastruttura tecnologica venga realizzata dal CINECA e dai suoi partner in modalità free software/open source (FLOSS: Free/Libre Open Source Software) in maniera da ridurre i costi di riutilizzo in altri contesti regionali, nazionali e oltre, e di permettere una potenziale estensione ed evoluzione dei servizi erogati ad altri ambiti sanitari.

A questo punto, è presto per dire se il progetto sarà coronato da successo, ma siamo in ogni caso convinti che la squadra che è scesa in campo per accettare questa sfida, è dotata della giusta dose di testardaggine ed ingenuità (caratteristiche proprie dei fanciulli) per ridare al mezzo televisivo quella dignità e quella valenza positiva nei confronti di una platea altamente sensibile come sono i minori, dignità che in tempi recenti è stata a volte offuscata da altre priorità più discutibili.

*Per ulteriori informazioni:*

[dt@cenea.it](mailto:dt@cenea.it)

<http://www.fub.it/dvb/dvbt/BandoGara/Esito/Esito.html>

*doi:10.1388/notizie-52-07*



# CoEdit: Collaborative Editing Tool

## Un applicativo web per la gestione di progetti in ambiente cooperativo distribuito

di Stefania Gardelli, Daniela Lotti, Elena Mantovani, Salvatore Rago

In questi anni, l'Area di Lavoro Cooperativo realizzata con tecnologia CINECA ha rappresentato uno strumento importantissimo nella gestione di progetti, al fine di condividere, organizzare e sincronizzare il lavoro di molte persone con compiti diversi e spesso fisicamente lontane.

Con l'evolversi dei servizi nell'ambito della Pubblica Amministrazione e dei progetti Europei, è nata l'esigenza di far crescere tale strumento per soddisfare le necessità di persone che scrivono documenti di vario tipo in modo collaborativo e che si scambiano informazioni contestuali al documento condiviso. Dalla volontà di rispondere a tali esigenze è nato *Coedit*, applicativo web il cui scopo è rendere più funzionale il flusso informativo e il processo di scrittura di un documento, evitando la frammentazione dello scambio informativo legato al documento.

Coedit è anche uno strumento eccellente per la gestione dell'homeworking nel campo della formazione a distanza.

### *A cosa serve*

CoEdit, Collaborative Editing Tool, è un applicativo web per la gestione delle attività di scrittura di documenti in ambiente cooperativo e distribuito: consente di scrivere "a più mani" il testo di un documento, di inserire

annotazioni direttamente nel documento e di navigare nella storia dei contributi, tenendo traccia di ogni possibile modifica del testo e dei relativi commenti.

### *Logica di funzionamento*

A differenza degli applicativi locali e proprietari di word processor, CoEdit è un applicativo web, supportato da qualsiasi browser, basato su tecnologie Internet standard e aperte.

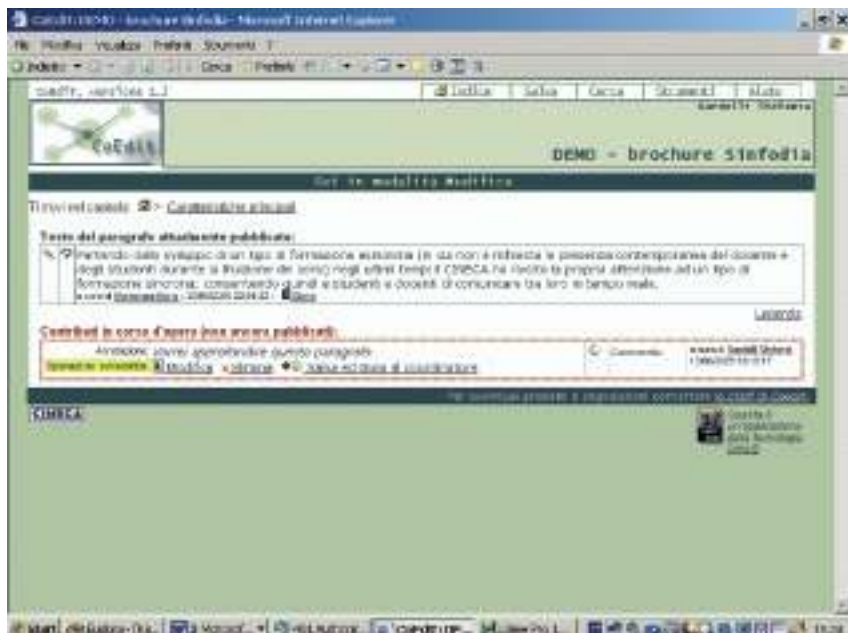
Essendo un applicativo web, gestisce il documento come un ipertesto, trattando di conseguenza il documento finale come un sito web. Ciò consente non solo la gestione di documenti multimediali, ma anche una facile e intuitiva navigazione all'interno del documento, rappresentato in una tipica struttura gerarchica di capitoli, sotto-capitoli e paragrafi. La home page del sito è l'indice del documento. L'utente può quindi navigare nel documento cliccando sulle voci dell'indice, ossia i capitoli e i sotto-capitoli del documento.

### *Le figure coinvolte*

Le figure che interagiscono con il sistema sono tre: redattore, coordinatore, supercoordinatore.

- **Il Redattore** si occupa principalmente di scrivere il testo del documento. Può aggiungere nuovi capitoli, sotto-capitoli e

*CoEdit, Collaborative Editing Tool, è un applicativo web per la gestione delle attività di scrittura di documenti in ambiente cooperativo e distribuito. Essendo un applicativo web, gestisce il documento come un ipertesto, trattando di conseguenza il documento finale come un sito web*



*Coedit è un sistema dinamico dove ogni singolo elemento del documento è aggiornato tempestivamente, per consentire la condivisione e lo scambio organizzato delle informazioni legate al documento in tempo reale*

paragrafi, può proporre revisioni al testo del documento già inserito, richiedere la cancellazione di un capitolo o di un paragrafo, fare delle annotazioni sul testo, visualizzare i contributi in corso d'opera e la storia dei contributi, salvare il documento, utilizzare il motore di ricerca. Nel campo dell'e-learning tale figura solitamente coincide con lo studente che deve svolgere il compito da casa (homeworking);

- **Il Coordinatore** è un redattore che in più si occupa di coordinare uno o più capitoli del documento, controllando l'attività svolta dai redattori che lavorano sul/i capitolo/i da lui coordinati. Il coordinatore ha il compito di approvare o rigettare le proposte di contributo che provengono dal redattore e di approvarle definitivamente per la pubblicazione finale. Il coordinatore, oltre ad avere le funzionalità di base del redattore, ha la facoltà di cancellare o spostare capitoli, sotto-capitoli e paragrafi. Nel campo dell'e-learning tale figura può coincidere con lo studente oppure con il docente, a seconda del grado di moderazione che si vuole attribuire allo strumento;
- **Il Supercoordinatore** è un coordinatore che in più si occupa di gestire il documento. E' suo compito attivare un nuovo docu-

mento, gestire la pubblicazione finale del documento e abilitare i redattori all'accesso del documento. Può inoltre abilitare uno o più coordinatori al coordinamento di un determinato capitolo o di più capitoli del documento e promuovere un redattore o un coordinatore al ruolo di supercoordinatore. Nel campo dell'e-learning tale figura coincide con il docente.

### *L'architettura tecnologica*

L'architettura tecnologica di Coedit si basa sui seguenti strati applicativi:

- **La tecnologia IANUS:** costituisce il cuore del sistema. E' una tecnologia infrastrutturale che ha il compito di assicurare il controllo e la gestione di servizi erogati via Internet;
- **BackStage Director:** rappresenta lo strumento di gestione dei contenuti e della "presentation" del sito;
- **La Piattaforma Coedit:** lo strato delle applicazioni specifiche che caratterizza lo strumento per la scrittura di documenti in ambiente cooperativo e distribuito. La piattaforma consente la gestione del documento, la gestione dei ruoli degli utenti e dei livelli di visibilità di ogni singolo elemento del documento.

Coedit è un sistema dinamico dove ogni singolo elemento del documento è aggiornato tempestivamente, per consentire la condivisione e lo scambio organizzato delle informazioni legate al documento in tempo reale.

### *Tecnologia Ianus*

La tecnologia IANUS è stata sviluppata al CINECA per consentire la realizzazione ed il mantenimento di servizi telematici basati su Internet (Extranet ed Intranet) che possano essere completamente gestiti dalla periferia con le necessarie garanzie di sicurezza.

Coedit è uno strumento riservato agli utenti abilitati, i quali possono accedere solo inserendo le proprie userid e password. Attraverso il controllo esercitato dallo strato applicativo IANUS, il sistema è così in grado di ricono-

scere in tempo reale l'utente online e di autorizzarlo ad effettuare determinate operazioni secondo il profilo a lui assegnato.

### **BackStage Director**

CoeEdit è stato realizzato con *BackStage-Director* (BS-D) la tecnologia di *web content management* sviluppata dal CINECA, ed impiegata per l'attività di produzione e manutenzione dei siti Internet.

I redattori utilizzano il sito come strumento di gestione del documento, sia per quanto riguarda il testo vero e proprio che per le annotazioni legate al testo. Lo scenografo BS-D ha predisposto un'interfaccia web che permette al redattore di visualizzare e modificare il documento in modo semplice e intuitivo. In particolare l'interfaccia web si basa sugli standard del W3C, con particolare riguardo agli standard di usabilità ed accessibilità.

Con un'installazione BS-D, si possono gestire più progetti/documenti, quindi secondo gli ambiti e le esigenze si possono costituire più gruppi di lavoro indipendenti, supervisionati da un medesimo coordinatore o da diversi coordinatori.

### **Piattaforma CoEdiT**

La piattaforma Coedit rappresenta lo strato delle applicazioni specifiche che caratterizza lo strumento per la scrittura di documenti in ambiente cooperativo e distribuito.

In particolare la piattaforma Coedit prevede:

- funzionalità associate ai ruoli degli utenti;
- gestione della visibilità degli elementi del documento. Ogni elemento del documento assume un livello diverso secondo lo stato in cui si trova; il sistema assegna un livello di partenza "0" a ciascun elemento del documento ogni qual volta si procede ad un nuovo inserimento oppure ad una nuova revisione del testo. Il Coordinatore ha il compito di far passare oppure no l'elemento del documento al livello successivo, fino all'ultimo livello che consente la pubblicazione finale;
- gestione delle annotazioni;

- salvataggio del documento. Un'applicazione specifica consente il salvataggio del documento completo, con la numerazione dei capitoli, in diversi formati: RTF (stampa), HTML (visualizzazione web) ed XML (esportazione del documento su altri sistemi).

### **Conclusioni**

Coedit può trovare utilità dovunque vi sia la necessità di scrivere documenti e progetti in un ambiente cooperativo e distribuito.

L'alto grado di usabilità e versatilità dell'applicativo fanno di Coedit uno strumento unico ed ottimale nel campo della formazione a distanza per la stesura di elaborati, tesine e documenti di ricerca in modo cooperativo e distribuito da parte di una classe di studenti dislocati su vasti territori.

**Coedit può trovare utilità dovunque vi sia la necessità di scrivere documenti e progetti in un ambiente cooperativo e distribuito**

Per ulteriori informazioni:

infogad@cenea.it

doi:10.1388/notizie-52-08



# La Banca Dati Biologica

di Luca Demattè, Marisa De Rosa, Maria Teresa Marano, Elisa Rinieri

## La ricerca traslazionale e le terapie su misura

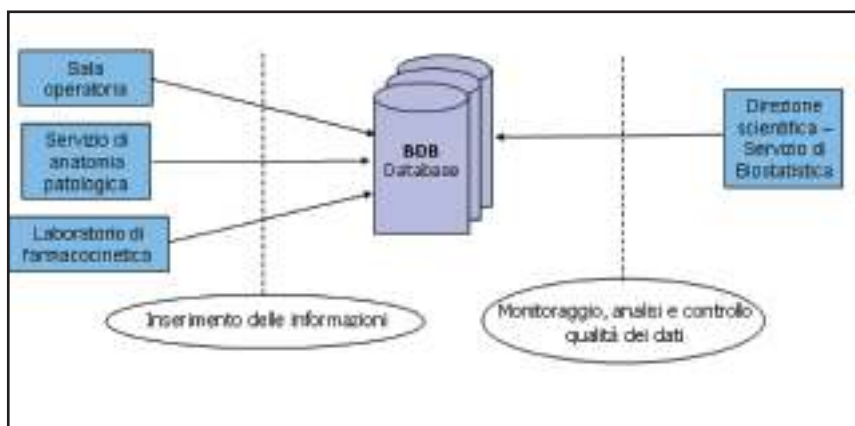
La “rivoluzione genomica” a cui stiamo assistendo negli ultimi anni ha dato origine ad un’innovazione profonda nel campo delle sperimentazioni cliniche. È così comparsa una nuova metodologia sperimentale, più adatta a caratterizzare l’efficacia delle nuove strategie terapeutiche altamente personalizzate. Il nuovo approccio sperimentale seguito da questi studi, detti *traslazionali*, cerca infatti di capire come i processi biologici influenzino l’evoluzione della malattia, e come possano essere modificati a vantaggio del paziente. È stata perciò elaborata una nuova metodologia per “traslare” le scoperte dalla clinica al laboratorio, e viceversa.

L’ambizione più alta dei ricercatori è quella di arrivare ad elaborare, più che una cura efficace contro il cancro, una cura efficace per i pazienti oncologici, giungendo a personalizzarla per ciascun individuo.

## Bisogni richiesti ai sistemi informativi e limiti dei sistemi attuali

Figura 1: flussi di informazione della Banca dati Biologica

Questo nuovo approccio sperimentale ha, ovviamente, delle ricadute importanti per



quanto riguarda i bisogni informativi degli sperimentatori. Se, da un lato, l’esigenza di avere dei flussi di informazione di alta qualità, non ambigui, sicuri da interferenze e manomissioni e aderenti a quanto stabilito nel protocollo di sperimentazione, non viene certo a cadere negli studi traslazionali, dall’altro però si impongono delle nuove necessità dovute al carattere molto più indiretto delle relazioni esistenti tra le informazioni cliniche e quelle raccolte in laboratorio. Una prima conseguenza è che uno studio traslazionale dovrà essere in grado di gestire ed analizzare una quantità di informazioni assai maggiore di quella degli studi tradizionali, ma non solo: in alcuni casi, l’analisi delle informazioni raccolte si rivela tutt’altro che banale, poichè ci si trova a dover provare o respingere una molteplicità di ipotesi che correlano insieme osservazioni e concetti appartenenti a domini di conoscenza differenti.

In linea generale, i sistemi informativi utilizzati nelle sperimentazioni cliniche o biologiche sono molto diversi tra loro: mentre i primi sono progettati su flussi informativi orientati al paziente, i secondi riguardano geni, processi intracellulari o tutt’al più intercellulari. Trovare una giusta integrazione tra i due approcci costituisce quindi un primo passo fondamentale per poter supportare la nuova ricerca in oncologia.

## La Banca Dati

### L’Istituto Regina Elena

L’Istituto Regina Elena è un Istituto di ricovero e cura a carattere scientifico (IRCCS) che opera dal 1939 nelle aree più avanzate della ricerca clinica occupandosi della cura delle patologie tumorali a più elevata incidenza e a più alto impatto sociale. Le attività dell’Istituto - inserite nell’ambito di progetti



nazionali ed internazionali - hanno come obiettivo principale la ricerca nei settori dell'epidemiologia, dell'eziologia dei tumori, della trasformazione e progressione neoplastica e delle terapie sperimentali. L'IRE svolge anche attività di ricerca applicata nella diagnostica biomolecolare e strumentale, in epidemiologia e diagnostica precoce, inoltre attua protocolli standard o innovativi per il trattamento e la riabilitazione del paziente oncologico.

### **La Banca dei Tessuti**

L'IRE gestisce, in modo opportunamente regolamentato, una banca di tessuti biologici criopreservati la cui essenziale funzione è quella di fornire materiale biologico per la ricerca scientifica, sia di base che finalizzata. L'implementazione di questa banca prevede una serie di aspetti:

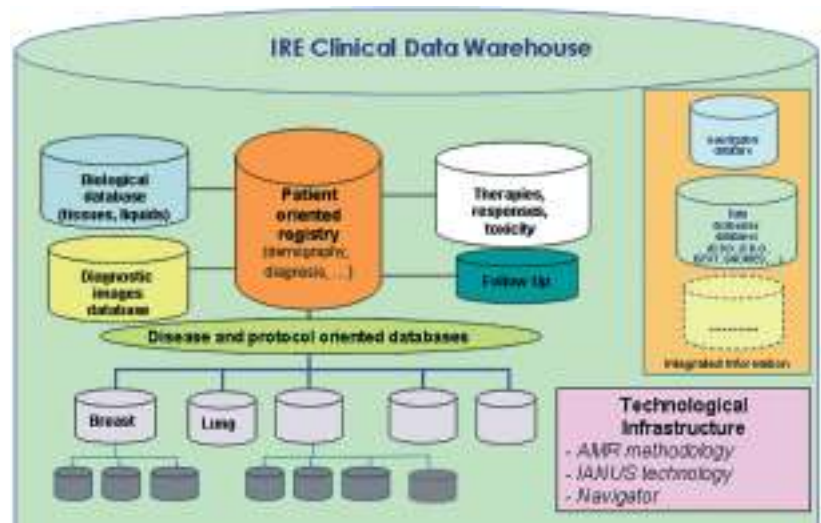
- il campionamento adeguato di tessuti (multipli, qualora sia possibile) maligni e benigni previo consenso informato del paziente;
- il prelievo e la raccolta di liquidi biologici (come sangue, urine, liquido ascitico, saliva...);
- il congelamento di tali campioni e la loro conservazione;
- l'immissione dei dati clinico-patologici in un data base.

### **Approccio AMR**

Nel 2004 è stata attivata una collaborazione tra IRE e CINECA per la creazione e il mantenimento di una banca dati, interamente basata su tecnologie standard Internet, per la gestione e il monitoraggio dei dati e di tutti i flussi informativi, per ogni singolo paziente, relativi alla banca biologica dell'IRE.

Il CINECA ha realizzato un sistema che, andando ben oltre la creazione di una banca dati biologica, consenta la registrazione di tutte le informazioni relative ai pazienti, alle loro patologie con terapie e tossicità, ai tessuti e ai liquidi criopreservati.

L'obiettivo perseguito è quello della creazione di un registro orientato al paziente, che ponga



da un lato le basi per l'attivazione di un *Clinical Data Warehouse* inteso come repository globale di tutte le informazioni cliniche e biologiche relative ai pazienti in cura presso l'Istituto Regina Elena.

La parte riservata è accessibile via Internet in modalità protetta con l'identificazione degli utenti attraverso userid e password, ed è stato realizzato sfruttando la metodologia AMR e la ventennale esperienza nella gestione delle sperimentazioni cliniche del Consorzio. Il sostrato tecnologico utilizzato per garantire l'aderenza agli standard di sicurezza del CINECA è quello della tecnologia IANUS, sul quale si appoggiano le applicazioni AMR che consentono di gestire in maniera puntuale tutte le informazioni raccolte come dati o immagini.

Un punto di forza del progetto è quello di permettere a diverse tipologie di utenti di inserire le informazioni nel momento stesso in cui vengono prodotte, realizzando in tal modo un'elevata qualità dei dati stessi.

Per questo tipo di approccio globale nella gestione dei dati dei pazienti, il progetto può essere proposto come modello nazionale per i pazienti oncologici.

### **Utilizzo di thesauri standard**

Un altro punto di forza del progetto, in ragione di quanto detto fino ad ora, sta nell'utilizzo di *thesauri* (dizionari) standard, quali *ICD-IX*,

Figura 2: lo schema del progetto

**L'IRE gestisce, in modo opportunamente regolamentato, una banca di tessuti biologici criopreservati la cui essenziale funzione è quella di fornire materiale biologico per la ricerca scientifica, sia di base che finalizzata**



Figura 3: la home page del progetto – <http://bdb.cineca.org/>

**...viene così a costituirsi un Data Warehouse Clinico- biologico dalle notevoli potenzialità, dove i problemi formulati dagli sperimentatori possono essere studiati per mezzo di un sistema di analisi multidimensionale trasversale ai singoli studi**

ICD-O, TNM e SNOMED®, con l'obiettivo non solo di migliorare la qualità dei dati e uniformare il linguaggio, ma soprattutto di aderire agli standard internazionali di codifica. Infatti, la rappresentazione dell'informazione clinica per mezzo di un vocabolario condiviso permette ai ricercatori di ridurre al minimo la possibilità di incomprensione relativamente ai singoli termini utilizzati per descrivere la patologia, la stadiazione tumorale, le caratteristiche istologiche...

### Clinical Data Warehouse

L'analisi dei dati raccolti ha come primo obiettivo quello di determinare *coorti* di pazienti in base ai loro parametri clinici, istologici e biologici, per condurre studi biologici sui tessuti e i liquidi a loro prelevati (profili genetici, presenza di marker biologici...). Contemporaneamente, però, dal momento che il sistema registra la partecipazione a questi protocolli di studio, tutte le nuove informazioni che provengono da tali studi permettono immediatamente ai ricercatori di ottenere correlazioni con i dati clinici e di orientare al meglio i protocolli terapeutici a cui i pazienti sono sottoposti. In definitiva viene così a costituirsi un Data Warehouse Clinico-biologico dalle notevoli potenzialità, dove i problemi formulati dagli sperimentatori possono essere studiati per mezzo di un sistema di analisi multidimensionale trasversale ai singoli studi, in modo che diversi punti di vista possano essere studiati per mezzo della stessa base informativa, e che si candida dunque ad essere un valido supporto alle ricerche traslazionali.

li. I report di analisi sono anche disponibili tramite *Navigator*, uno strumento di tipo OLAP realizzato specificatamente per questo tipo di analisi complesse (di cui abbiamo parlato nell'ultimo numero del Notiziario).

### Sviluppi futuri

Attualmente la Banca Dati Biologica permette un'integrazione dei dati clinici con alcuni dati di rilevanza genetica, principalmente un certo numero di bio-marker ed altri fattori biologici, destinato ad aumentare nel tempo. Uno sviluppo auspicabile del sistema, ed un'autentica scommessa tecnologica, potrebbe essere tentato integrando nel Data Warehouse un numero maggiore di informazioni genetiche, quali ad esempio i profili di espressione genica rilevati direttamente dai microarray, un campo in cui CINECA ha già acquisito una competenza rilevante (vedi articolo di pagina 10). Le difficoltà metodologiche imposte all'analisi da uno sviluppo di questo tipo potrebbero essere affrontate con maggior efficacia dagli approcci innovativi che sono stati sviluppati nell'ambito del *Knowledge Management*. In effetti, già da alcuni anni è stata sviluppata una *Gene-Ontology* (GO) con la collaborazione dell'intera comunità dei ricercatori coinvolti a vario titolo nella genetica. Ciò che manca ad un utilizzo pieno della GO nel tipo di studi di cui abbiamo finora parlato è la sua estensione a domini di conoscenza più vicini all'aspetto clinico.

La sfida finale sarà allora quella di integrare in un unico sistema di analisi un complesso eterogeneo di informazioni e conoscenze di alto livello.

Per ulteriori informazioni:  
[infosiss@cinca.it](mailto:infosiss@cinca.it)

doi:10.1388/notizie-52-09

# La “sicurezza ambientale” al CINECA

di Sergio Biagi Foresti, Alessio Mauri

L'attività del CINECA si evolve sempre più verso applicazioni ad alta criticità.

Tutto questo rende necessario l'adozione di provvedimenti atti ad eliminare o a contenere al minimo possibili disservizi per gli utenti del Centro di Calcolo.

A questo scopo sono state realizzate diverse opere di protezione della struttura che ospita il Consorzio :

- recinzione esterna lato via Magnanelli di un metro di altezza;
- vasca volano di raccolta per le acque meteoriche utilizzata anche per il gruppo di pompaggio antincendio di 80 mc;



- paratia mobile davanti all'ingresso principale del CINECA e rifacimento della rete fognaria.

Gli impianti elettrici e di condizionamento sono stati potenziati per far fronte alle nuove esigenze operative e ai possibili disservizi provocati dalle interruzioni ENEL. Pertanto si sono realizzati:

- cabina elettrica di trasformazione con due trasformatori da 1000 kVA e la possibilità

di inserimento di un terzo trasformatore. Questa cabina andrà ad affiancare l'esistente dotata di tre trasformatori da 400 kVA;



- generatore elettrico da 1700 kVA che affianca il preesistente di 250 kVA, per sopperire ai power failure di fornitura elettrica anche per lunghi periodi e alimentare tutte le attuali utenze CINECA, compresi gli impianti di condizionamento delle sale elaboratori e comunicazioni;
- gruppo di 3 ulteriori UPS da 80 kVA in affiancamento ai gruppi dei 3 UPS da 80 kVA e dei 2 UPS da 250 kVA già in servizio.



Accanto al potenziamento di impianti elettrici e a protezioni anti-allagamento, contempora-

*L'attività del CINECA si evolve sempre più verso applicazioni ad alta criticità. Questo rende necessaria l'adozione di provvedimenti atti ad eliminare o a contenere al minimo possibili disservizi per gli utenti del Centro di Calcolo.*

neamente alla realizzazione come integrazione dei vari ampliamenti (sala macchine, inter-rati ...), non sono stati trascurati i sistemi più sofisticati e continuamente aggiornati di:

- rilevazione incendi, in tutto il complesso CINECA;
- spegnimento nelle sale elaboratori tramite gas inerti (argon);
- antintrusione;
- videosorveglianza;
- controllo accessi.

**Tutti i sistemi di sicurezza sono gestiti e supervisionati da addetti presenti al CINECA per 365 giorni all'anno e per 24 ore al giorno**

C'è infine da ricordare che tutti questi sistemi sono gestiti e supervisionati da addetti presenti al CINECA per 365 giorni all'anno e per 24 ore al giorno.

### **Il prossimo futuro**

Le risorse richieste dai centri di calcolo o da quelli che si occupano delle comunicazioni, sono oggi molto diverse da quelle di pochi anni fa. Si assiste ad una continua concentrazione, in spazi ridotti, della potenza di calcolo. Questo comporta un aumento di potenza elettrica e conseguentemente la produzione di calore, con valori che al momento possono raggiungere i 20 kW e più, per rack. Al momento quindi coesistono rack con potenze dissipate di pochi kW e rack di 20-24 kW. I sistemi di condizionamento tradizionali si stanno rilevando inadeguati, poco flessibili ed ingombranti per risolvere i problemi indotti da queste grandi quantità di calore concentrate e localizzate in spazi ridotti e definiti. Di qui la necessità di avere sistemi di condizionamento in grado di condizionare non solo l'ambiente del centro di calcolo, ma anche, in particolare localmente, la fonte principale di calore; questo comporta, oltre ad un notevole risparmio energetico, un maggior confort per le persone operanti nei locali (D. Lgs. 626/94), non essendo necessario raggiungere valori di temperatura troppo bassi con l'aria diffusa da griglie poste sul pavimento galleggiante. Sono attualmente allo studio alcuni sistemi di condizionamento ed in particolare sistemi localizzati, piazzati a soffitto e/o direttamente sul top del rack, e sistemi localizzati a circuito freddo chiuso e ci stiamo attrezzando per eventuali

futuri sviluppi nel campo delle necessità di aumenti di energia elettrica mediante potenziamento delle cabine di trasformazione, degli UPS ed infine dei GE.

In conformità con quanto prescritto dal D. Lgs 626/94, il CINECA si è dotato di un "Servizio di prevenzione e Protezione" (SPP) che provvede alla:

- Analisi dei rischi all'interno dell'Azienda;
- Predisposizione del Piano di emergenza;
- Organizzazione del servizio antincendio;
- Organizzazione del servizio di primo soccorso;
- Gestione dei Dispositivi di Protezione Individuale (DPI);
- Verifica annuale dei rischi e delle azioni.

Il CINECA provvede inoltre all'informazione e formazione dei lavoratori, infatti all'atto dell'assunzione o dell'inizio del rapporto di collaborazione vengono forniti i seguenti documenti:

- Operare con sicurezza in azienda;
- Opuscolo sui rischi connessi all'impiego delle apparecchiature VDT;
- Piano di emergenza.

Copia aggiornata dei suddetti documenti è conservata sulla bacheca elettronica del CINECA; annualmente è previsto un corso di informazione ed aggiornamento tenuto dal medico competente e, secondo necessità, vengono organizzati corsi di formazione per la squadra antincendio ed il servizio di primo soccorso.

Secondo quanto previsto dall'art. 18 del D. Lgs 626/94 sono inoltre presenti al CINECA dei Rappresentanti per la sicurezza, eletti dal personale CINECA. I due organismi interagiscono e collaborano fattivamente per garantire condizioni di lavoro ottimali per il personale.

*Per ulteriori informazioni:*  
gestionesistemi@cineca.it

doi:10.1388/notizie-52-10



## GESTIONE E ANALISI DELL'INFORMAZIONE

**Plug&Play: Galleria di Immagini di Sabina Parmeggiani\****\*Interactive Marketing***Di cosa si tratta**

La *Galleria Immagini* consente di caricare in tempo reale immagini e dati ricavati da un database prestabilito. Per realizzare l'interfaccia è stato utilizzato il formato Flash di Macromedia.

Flash è programmato per caricare le informazioni tramite un documento XML che viene aggiornato dinamicamente ad ogni richiesta della pagina, garantendo che il contenuto della banca dati venga riflesso in tempo reale nell'applicazione.

Diverse sono le possibilità offerte da questo strumento: ad esempio le immagini possono essere condivise da più gallerie nello stesso tempo, che implementano in modo magari assai differente la visualizzazione delle immagini e dei titoli, sia esteticamente a causa di differenti layout della maschera grafica, sia per modalità di trattamento, come ad esempio la sequenza di presentazione.

È inoltre possibile raggruppare le immagini in categorie, e presentarle tramite un menù a scelta dal quale l'utente seleziona il gruppo, ottenendo la sequenza di miniature cliccabili.

**Vantaggi**

*Flessibilità di visualizzazione e alta qualità dell'immagine*

Il Plug&Play (P&P) *Galleria Immagini* è uno strumento molto utile per chi voglia effettuare una presentazione fotografica web in cui alle singole immagini vengono affiancati dati testuali, utilizzando un'interfaccia di facile comprensione e ottenendo un rapido aggiornamento dei contenuti.

Per questo motivo è lo strumento ideale per presentazioni sia a scopo didattico che altamente professionalizzanti, come ad esempio in campo medico-biologico.

L'alto grado di flessibilità dell'applicazione garantisce inoltre la possibilità di personalizzare non soltanto il contenuto della presentazione stessa, ma anche il layout grafico e la modalità di sequenza di immagini e testo.

*Integrato con il software di content management BackStage Director*

Il P&P è completamente integrato con il software di *web content management BackStage Director* sviluppato da CINECA e Interactive Marketing.

**RUBRICA**

doi:10.1388/notizie-52-rubrica

## EINS: EUROPEAN INFORMATION NETWORK SERVICES

**ULTIME NOTIZIE DA EINS GEM  
di Paola Manni, Roberta Turra****Eins e Data/Text Mining**

Dalla Home Page di EINS è ora disponibile il link: *Data and Text Mining*. Da questo link possono essere raggiunti due servizi, *PatMole* e *MedMole*, sviluppati con tecnologia di Text Mining (di cui forniamo una descrizione di seguito) nell'ambito rispettivamente dell'informazione brevettuale e dell'informazione medica della banca dati Medline. Per accedervi e per avere maggiori dettagli è necessario contattare il proprio Centro Nazionale EINS. È auspicabile prevedere in futuro un'integrazione di tali tecnologie di analisi dei dati e dei testi con il servizio banche dati EINS GEM.

**Cos'è il TEXT MINING**

Per Text Mining si intende il processo automatico di estrazione di informazione significativa, e quindi di conoscenza, da grandi quantità di documenti testuali non strutturati (banche dati testuali, e-mail, file di word-processing, raccolte di news, pagine web ed altre ancora). Le tecniche di Text Mining possono essere utilizzate principalmente per due diverse finalità: *classificazione automatica* o *raggruppamento tematico*.

Nel primo caso consentono di raggruppare documenti in base a categorie predefinite (*categorization*) e trovano applicazione, per esempio, nei sistemi anti-spamming, nei sistemi di automatic routing/dispatching di documenti, ecc...

Nel secondo caso, consentono di raggruppare i documenti in base al contenuto (*clustering*) e quindi scoprire argomenti "nascosti" e interessanti interrelazioni fra argomenti.

**Documentazione cartacea MOLE - Minig On Line Expert**

MOLE è la tecnologia specificamente sviluppata dal CINECA per l'utilizzo delle tecniche di Text Mining online. Questo strumento con-

sente la rapida individuazione delle informazioni più interessanti all'interno di grandi quantità di documenti, anche di fonte eterogenea, operando un raggruppamento dinamico (online) dei documenti selezionati dall'utente in categorie significative.

L'appartenenza ad un gruppo (cluster) è determinata da misure matematiche e linguistiche della "content similarity": l'uguaglianza di contenuti.

I benefici di questo approccio sono duplici: fornire velocemente la visione di una grande quantità di documenti e delle relazioni che intercorrono fra essi (attraverso una mappa tematica); aiutare la navigazione nella ricerca di ulteriori documenti

**PatMole: l'analisi di documenti brevettati**

PatMole è la versione di MOLE personalizzata per l'analisi di documenti brevettuali di fonte EPO (Espacenet). Attualmente consente di analizzare tutti i brevetti a partire dal 1990, per un totale di circa 1.800.000 documenti. Scopo principale è l'individuazione di nuovi orientamenti del mercato, nuovi potenziali concorrenti, conoscere le strategie della concorrenza ed individuare possibili sinergie e opportunità.

**MedMole: l'analisi degli abstract di Medline**

MedMole consente di analizzare 6 milioni di abstract di fonte Medline. Inizialmente sviluppato per i bioinformatici, ha moltissimi campi di applicazione, tra cui l'identificazione di gruppi di geni funzionalmente correlati.

**Informazioni:**

EINS GEM HelpDesk  
e-mail: [eins@ Cineca.it](mailto:eins@ Cineca.it)  
telefono 051 6171497  
<http://www.einsgem.org/>



## RUBRICA

doi:10.1388/notizie-52-rubrica

## COMUNICAZIONI E SISTEMI DISTRIBUITI

### Virtualschool@cineca

La lunga tradizione delle Scuole Estive-CINECA è stata arricchita quest'anno da un servizio di formazione a distanza basato sul sistema di e-learning OSIRIS (descritto nell'articolo di pagina 20). Una decina di studenti ha potuto seguire in questa modalità i corsi che si sono tenuti in aula, tramite un sistema appositamente predisposto. L'esperienza, molto apprezzata dagli studenti, verrà descritta nel prossimo numero del notiziario. Per prendere visione degli strumenti messi a disposizione degli studenti è possibile collegarsi all'indirizzo:  
<http://virtualschool.cineca.it/>



### Presentazione del Presidente dell'Autorità per l'energia elettrica e il gas

Il 23 giugno 2005, il CINECA ha trasmesso in diretta su Internet la presentazione del Presidente dell'Autorità per l'energia elettrica e il gas relativa alla "Relazione annuale sullo stato dei servizi e sull'attività svolta". <http://streaming.cineca.it/report2005/>

### eSTREAM, Leuven 26-28 Maggio 2005

Il 26 maggio 2005 a Leuven Belgio, CINECA è intervenuto al workshop "Video Streaming in Education in Europe, a Best Practise Seminar". Il workshop, incentrato sull'uso della tecnologia streaming nel-

l'ambito dell'e-learning, è stato organizzato dai partner di eSTREAM, un progetto finanziato all'interno del programma SOCRATES/MINERVA.

### Conferenza GARR\_05, Pisa 10 - 13 Maggio 2005

In occasione del convegno GARR\_05 CINECA ha curato il servizio di streaming live in formato Real.  
[http://www.cineca.it/csd/news/Convegno\\_GARR05.htm](http://www.cineca.it/csd/news/Convegno_GARR05.htm)

### Televisione Digitale Interattiva: le attività di divulgazione.

*Java Conference 2005, 22-23 giugno Centro Congressi Milanofiori - Assago*

Proseguono le attività di divulgazione del laboratorio per la Televisione Digitale Terrestre del CINECA, con la partecipazione alla "Java Conference 2005". Andrea Venturi, responsabile del progetto di tv digitale interattiva del Consorzio, è intervenuto al convegno con una presentazione dal titolo "Software libero per la tv digitale interattiva: la scelta è possibile".

[http://www.cineca.it/csd/news/Java\\_Conf\\_05.htm](http://www.cineca.it/csd/news/Java_Conf_05.htm)

### Forum Europeo sulla Televisione Digitale Terrestre, di Lucca

Il 10 giugno 2005 il Consorzio ha partecipato al "Forum Europeo sulla Televisione Digitale Terrestre", presentando la propria esperienza in questo ambito.

<http://www.cineca.it/csd/news/convlucca2005.htm>

### La formazione

Tra le attività del Laboratorio per il Digitale Terrestre rientra l'organizzazione di corsi di introduzione alla rete della TV Digitale Interattiva e all'utilizzo della piattaforma Just DVB-IT. Maggiori informazioni su <http://www.cineca.it/corsi/>

## RUBRICA

doi:10.1388/notizie-52-rubrica

## GESTIONE SISTEMI

### Evoluzione in ambito Open Source dell'infrastruttura DBMS di Giovanna Piscitelli

Nel panorama dei sistemi di gestione di basi di dati (DBMS), le soluzioni Open Source hanno ormai acquisito una posizione di primo piano. Diverse sono le motivazioni che ne hanno decretato il successo, tra cui principalmente il supporto di una numerosa comunità di utenti e sviluppatori, l'integrazione con altri prodotti Open Source, la disponibilità nel tempo di funzionalità sempre più evolute, la riduzione dei costi che l'adozione di queste soluzioni comporta.

La diffusione di applicazioni basate su database server Open Source e l'interesse sempre crescente verso queste tecnologie portano alla decisione di ampliare in questa direzione l'offerta CINECA di soluzioni DBMS per la fornitura di servizi ad elevata criticità.

In questa ottica è stata pertanto realizzata una infrastruttura per il supporto di servizi basati su database MySQL.

MySQL è tra i database server Open Source il più popolare. Deve la sua diffusione tra l'altro alle sue caratteristiche di portabilità (è disponibile sui principali sistemi operativi ed esistono interfacce per i più diffusi linguaggi di programmazione) e di flessibilità (mette a disposizione diverse architetture per la memorizzazione dati, o *storage engine*, le cui specifiche funzionalità consentono la convivenza sul medesimo sistema di soluzioni applicative dai diversi requisiti). A questo si aggiunge la costante evoluzione che ha portato nel tempo a garantire funzionalità e soluzioni tipicamente offerte dai più noti DBMS commerciali.

L'infrastruttura predisposta al CINECA ha l'obiettivo di supportare l'intero ciclo di vita di un servizio basato su MySQL, dalla fase di sviluppo e test, alla messa in produzione su un sistema che garantisca i necessari requisiti di affidabilità e disponibilità.

Sono stati predisposti pertanto un server MySQL di sviluppo e uno di produzione.

Il server di produzione è ospitato su un cluster di due nodi Linux in alta affidabilità. I dati sono memorizzati sulla Storage Area Network CINECA. Il server MySQL è configurato come un servizio del cluster, ovvero ad esso è associato un insieme di risorse, l'indirizzo IP su cui ascolta per le richieste da remoto e lo spazio disco che contiene i file di dati, che normalmente sono assegnate ad un nodo del cluster e che, qualora questo smetta di funzionare, vengono migrate in maniera automatica sul restante nodo del cluster. Questa configurazione assicura la disponibilità dei dati in caso di fermo, imprevisto o programmato, di uno dei nodi del cluster. Per garantire inoltre la protezione dei dati da perdite accidentali e da errori software/hardware sul sistema di produzione, è stata configurata una replica master to slave tra il server di produzione e un secondo server MySQL.

La replica è stata implementata interamente con funzionalità integrate nel server MySQL. Per la gestione centralizzata dei server MySQL dell'infrastruttura è stato infine messo a disposizione il tool web PhpMyAdmin. Sono incluse in questa interfaccia funzionalità prettamente amministrative (creazione di utenti e database) e funzionalità di carattere applicativo (creazione e gestione di tabelle e indici all'interno di un database) che ne consentono pertanto l'utilizzo a vari livelli. L'accesso alle diverse funzionalità è regolato dai privilegi di cui dispone l'utenza usata per autenticarsi al singolo server MySQL da gestire.

L'offerta CINECA in fatto di soluzioni DBMS Open Source è destinata a evolvere ulteriormente. È attualmente in fase di realizzazione una infrastruttura analoga a quella fin qui descritta destinata a supportare servizi in produzione basati su DBMS PostgreSQL.

## RUBRICA

doi:10.1388/notizie-52-rubrica