



Consiglio Nazionale delle Ricerche

Programma Nazionale di Ricerche in Antartide



Unità Operativa ISTI-CNR

Geo-Metadata Server: Metadata Entry Facility

Maria Grazia Di Bono, Massimo Martinelli, Ovidio Salvetti

**Istituto di Scienza e Tecnologia dell' Informazione
Consiglio Nazionale delle Ricerche
Area della Ricerca di Pisa
Via G. Moruzzi, 1 - 56124 PISA**

Indice

1. Struttura generale	4
2. Struttura generale del sistema d'interfaccia.....	7
3. Struttura generale dell'interfaccia	8
4. Uso dell'interfaccia	10
4.1. L'accesso al sistema	10
4.2. La navigazione	11
4.3. Uso generale	12
4.4. Thesaurus e parole chiave	13
4.5. Selezione di un'area geografica	15
4.6. Invio delle informazioni per la registrazione (Submit)	16
4.7. Uso delle immissioni parziali.....	18
5. Note tecniche.....	20
5.1. Requisiti per l'uso	20
5.2. Testing	20

Premessa

Nell'ambito del progetto S.I.R.I.A. , Sistema Informativo della Ricerca Italiana in Antartide, l'unità operativa dell'Istituto di Scienza e Tecnologie dell'Informazione ha il compito di studiare e realizzare un sistema per la raccolta via web e la gestione dei metadati relativi alle campagne antartiche.

Scopo di questa nota è quello di illustrare la realizzazione e le possibilità offerte dal modulo di Metadata Entry del sistema Geo-Metadata Server (GMDS).

1. Struttura generale

Il Geo-Metadata Server (GMDS) è un sistema per la conservazione e la diffusione di dati provenienti dalle campagne italiane in Antartide.

L'Istituto di Scienza e Tecnologia della Informazione del CNR di Pisa (ISTI-CNR) si occupa di progettare e realizzare il sistema di raccolta, conservazione e consultazione delle meta-informazioni.

La struttura generale del sistema è riportata in Figura 1.

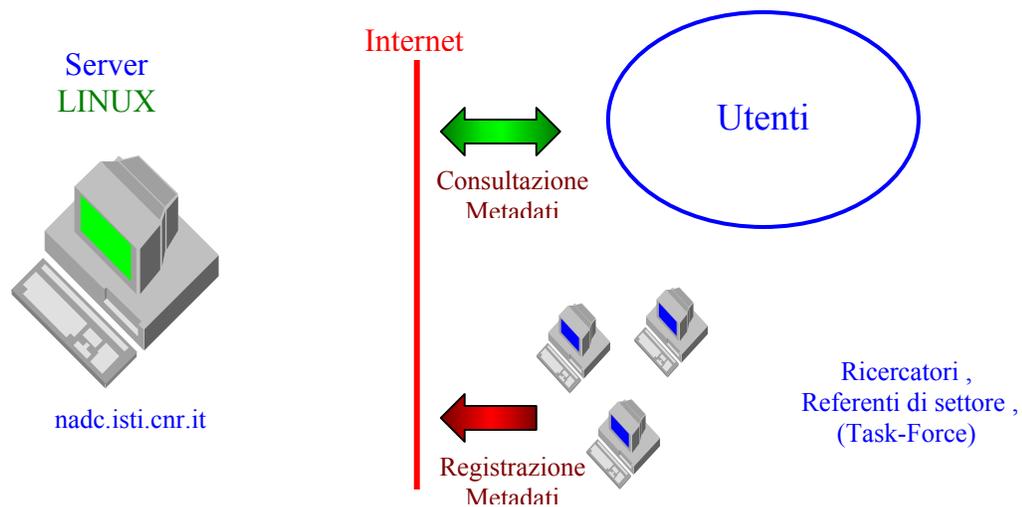


Figura 1. Architettura generale del sistema.

La Task-Force, composta da tre membri, ha il compito di recuperare tutte le meta-informazioni relative alle campagne italiane di ricerca in Antartide, di compilare le schede di metadato e, previa certificazione di qualità da parte dei referenti di settore, di inviarle all'ISTI che si occupa della loro registrazione. A tale proposito l'ISTI ha messo a disposizione il software per la compilazione via web delle schede, che segue lo standard di metadato CEN/TC 287.

Nella macchina `nadc.isti.cnr.it`, risiede il catalogo dei metadati dedicato alla conservazione delle meta-informazioni ufficialmente approvate dai supervisori competenti e l'interfaccia per la sua interrogazione.

Gli accessi esterni alla macchina sono quindi di due tipi:

1. Accesso per la registrazione delle meta-informazioni;
2. Accesso per la consultazione delle meta-informazioni.

Per gli accessi del primo tipo è stato previsto un meccanismo di registrazione ed autenticazione degli utenti che consente l'utilizzo del tool di Metadata Entry. Questo tipo di accesso è consentito solo ad un gruppo ristretto di persone, che comprende i tre membri della Task-Force e tutti i ricercatori o i referenti di settore registrati. Una volta registrato, specificando un username, una password ed il settore PNRA di appartenenza (nel caso degli utenti della Task-force si può selezionare la voce "All sector"), l'utente può accedere al tool web specificando l'interfaccia che si vuole utilizzare. Le interfacce web disponibili comprendono una interfaccia standard che ricalca in maniera completa lo standard CEN e delle interfacce personalizzate, realizzate seguendo le specifiche concordate con i referenti dei singoli settori. Tutti gli utenti registrati possono compilare le schede di metadato e salvarle in una specifica area del server (area A), organizzata ricalcando la struttura dei settori del PNRA. La compilazione delle schede di metadato può essere parziale, ossia il tool consente di salvare tali schede secondo la modalità "patial filling" ed di recuperarle in un secondo momento, completarle e salvarle con la modalità "final submit". In una successiva fase di validazione i metadati completi potranno essere verificati dai referenti di ciascun settore. Dopo la fase di validazione, le schede che dovranno essere memorizzate nel database saranno salvate in una seconda area del server (area B). La registrazione effettiva nel catalogo delle meta-informazioni avviene sotto la diretta responsabilità dell'ISTI attraverso una procedura automatica di "mapping" appositamente realizzata.

Gli accessi del secondo tipo sono consentiti a tutti gli utenti interessati alla consultazione via web del catalogo delle meta-informazioni, attraverso una interfaccia di interrogazione del database e non richiede nessuna registrazione.

La

Figura 2 mostra il diagramma funzionale del sistema in riferimento all'utilizzo dei dati e dei metadati da parte di due diversi gruppi di utenti.

Il primo gruppo di utenti, che comprende la Task-Force, i ricercatori ed i referenti di ciascun settore, accede ai metadati in scrittura e lettura, attraverso il processo di salvataggio parziale o definitivo delle schede compilate in un'area temporanea. L'altro tipo di accesso in lettura avviene attraverso la consultazione on-line del catalogo.

Per tutti gli altri utenti l'accesso ai dati avviene solo in lettura per la consultazione via web del catalogo dei metadati.

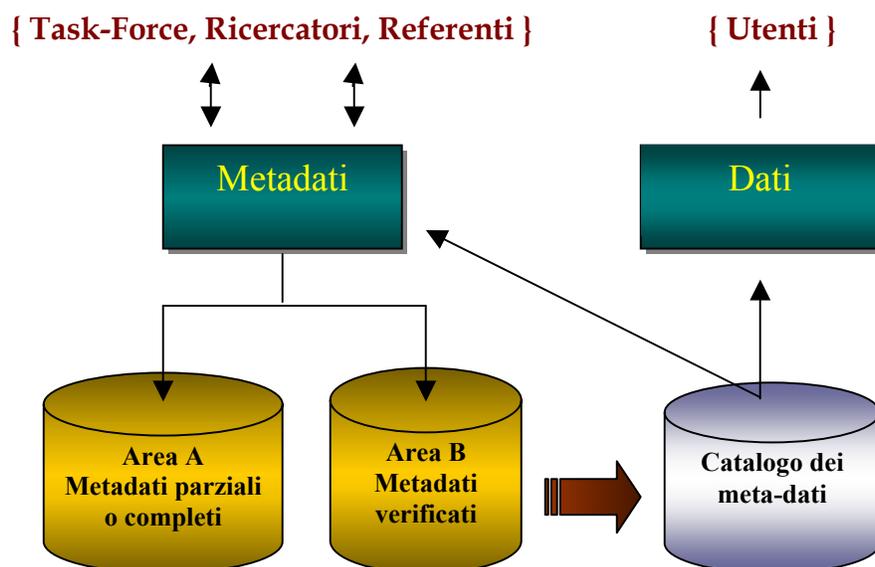


Figura 2. Diagramma funzionale.

2. Struttura generale del sistema d'interfaccia

L'interfaccia per l'inserimento delle meta-informazioni è stata progettata per adattarsi alle esigenze specifiche dei vari settori della ricerca in ambito antartico. La sua struttura di base permette lo sfruttamento generale dello standard di metadati CEN/TC 287, che costituisce anche la parte comune a tutte le versioni personalizzate.

Le versioni personalizzate sono state create infatti partendo da questa base comune e quindi

- eliminando i campi non necessari (con l'introduzione eventuale di valori particolari da inserire nei campi obbligatori in modo trasparente all'utente);
- utilizzando insiemi di valori concordati con i vari settori di ricerca per quei campi che lo hanno permesso;
- fornendo strumenti ad hoc per semplificare le operazioni più ripetitive o tipiche di ciascun settore.

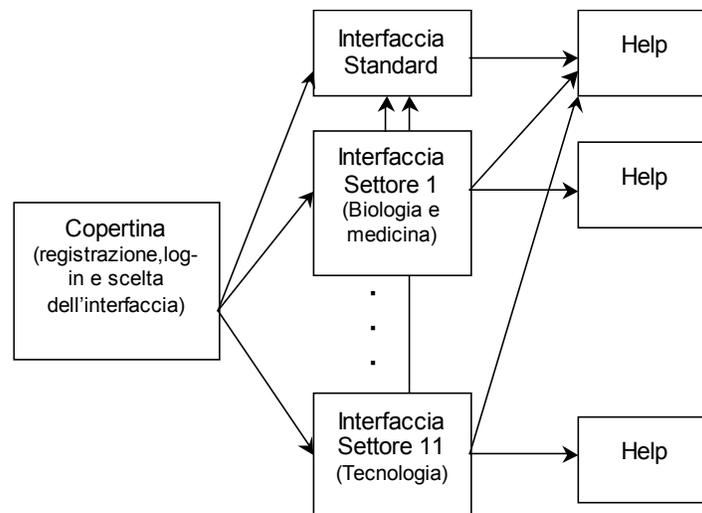


Figura 3. Struttura generale del sistema d'interfaccia

3. Struttura generale dell'interfaccia

L'interfaccia per inserimento delle meta-informazioni è stata strutturata in blocchi d'informazione per agevolarne l'utilizzo e semplificarne la comprensione e la gestione da parte dell'utente.

I blocchi d'informazione sono stati ricavati dalla struttura dello standard adottato per le meta-informazioni, il CEN/TC 287. Esso prevede che i campi che lo compongono siano raggruppati secondo il genere di informazione fornita (informazioni globali, di amministrazione, di distribuzione, ecc.). Tale struttura (Figura 4) è l'ideale anche per l'interfaccia.

L'interfaccia per l'inserimento della meta-informazione è costituita da una serie di pagine, ognuna delle quali è associata ad uno dei blocchi d'informazione elencati in Figura 4.

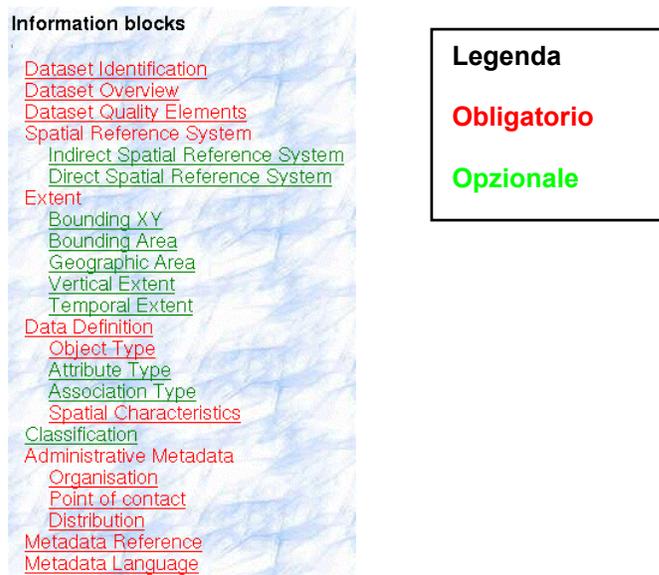


Figura 4. Struttura a blocchi della meta-informazione

Come si vede nella Figura 5 , l'aspetto dell'interfaccia si presenta suddiviso in tre *frame*.

Il primo, a sinistra nel display e più piccolo, elenca la lista dei blocchi d'informazione in un modo analogo a quello visto in Figura 4; in questo modo è consentito l'accesso alle varie pagine che raccolgono le informazioni per i vari blocchi, nei quali sono presenti anche i pulsanti per il salvataggio delle informazioni delle schede compilate (Submit), quello per il caricamento delle schede parzialmente compilate (Partial Fillings) e quello per l'azzeramento delle informazioni raccolte (Reset).

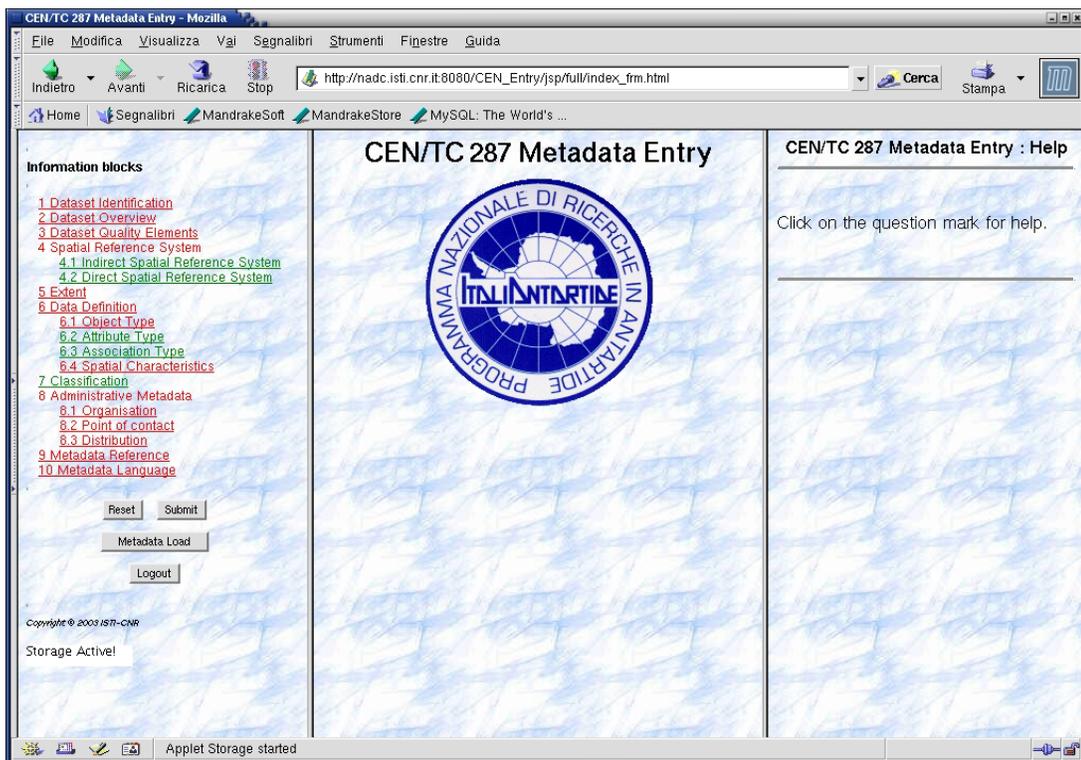


Figura 5. Aspetto del Metadata Entry tool

Il secondo *frame*, visualizzato al centro, riporta man mano le pagine relative ai vari blocchi d'informazione che permettono la compilazione delle schede dei metadati. I vari blocchi sono composti da campi di testo, liste di valori o quant'altro necessario.

Il terzo *frame*, visualizzato sulla destra, consente di visualizzare la pagina di help relativa a ciascun campo da compilare, semplicemente cliccando sul punto interrogativo che si trova accanto all'intestazione di ciascuno.

4. Uso dell'interfaccia

4.1. L'accesso al sistema

L'accesso al sistema avviene previa autenticazione degli utenti che devono effettuare una registrazione iniziale. La figura 6 e 7 mostrano rispettivamente la pagina di registrazione degli utenti e quella iniziale di accesso al Metadata Entry tool.

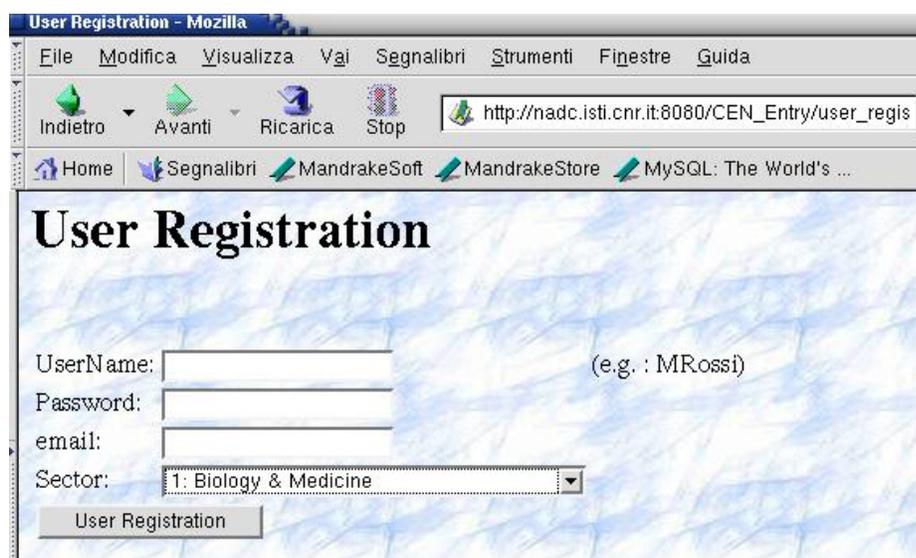


Figura 6. Pagina di registrazione.

Attraverso la specifica di un nome utente, una password e l'indirizzo e-mail e la selezione del settore di appartenenza è possibile effettuare la registrazione nel database per accedere al web tool di Metada Entry. Nel caso in cui l'utente sia uno dei membri della Task Force, è possibile selezionare dalla lista dei settori la voce "All sectors".

Una volta effettuata la registrazione è possibile accedere al sistema.

La pagina principale dell'interfaccia del sistema per l'inserimento dei metadati, consente all'utente di accedere a informazioni di carattere generale sulla funzionalità del modulo di Metadata Entry e sull'organizzazione che ne regola lo sviluppo (link *Details*).

Inoltre consente di effettuare la scelta, facilitata dalla presentazione di un'apposita lista, dell'eventuale interfaccia personalizzata, relativa ad un preciso settore di ricerca, che si desidera utilizzare.



Figura 7. Pagina d'accesso al Metadata Entry tool..

Operata la scelta dell'interfaccia l'accesso al sito è consentito attraverso un meccanismo di autenticazione (login e password) degli utenti registrati.

4.2. La navigazione

L'interfaccia permette una navigazione tra le sue varie pagine semplice ed intuitiva.

Ogni pagina, infatti, può essere acceduta direttamente cliccando sul nome del blocco d'informazione che interessa, in modo che in ogni momento sia possibile inserire una qualunque delle informazioni necessarie al corretto completamento dell'operazione.

In alternativa è possibile visionare ed eventualmente compilare tutte le pagine in sequenza, selezionandone una dalla lista e poi passando alla successiva o alla precedente, semplicemente agendo sui pulsanti presenti all'inizio e alla fine di ogni pagina.

La seconda possibilità è preferibile in quanto permette in un unico passaggio di inserire tutte le informazioni necessarie al completamento dell'operazione, evitando il

rischio di dimenticarne alcune e dover ritornare sui propri passi per inserire ciò che manca.

Per rendere più semplice l'orientamento all'interno dell'interfaccia, ogni volta che viene attivata una pagina, nel frame di sinistra compare una piccola freccia verde in prossimità del nome del blocco corrispondente.

4.3. Uso generale

L'uso dell'interfaccia si basa su alcuni concetti base e su dei modi di procedere standard che sono utilizzati in tutte le pagine che la compongono.

1. *Tutti i nomi*, sia dei blocchi d'informazione sia dei campi, sono colorati in rosso per indicare quelli obbligatori da compilarsi in ogni caso (talvolta inserendo diciture del tipo "not applicable" o "unknown"), o in verde per quelli opzionali.

2. *Tutti i nomi dei campi* (tranne quelli più ovvi, quali nome, cognome, ecc.) hanno a sinistra un punto interrogativo bianco. Questo rappresenta un link che permette l'accesso diretto alla pagina della guida in linea sulla quale sono riportate informazioni sul significato del campo e su come deve essere riempito (talvolta anche con esempi). Queste pagine sono personalizzate secondo il settore scelto e forniscono quindi informazioni mirate al proprio campo di ricerca.

3. *I campi ripetibili* si presentano inizialmente con il solo nome affiancato a destra da un pulsante identificato da un simbolo "+". La selezione di questo tasto permette di aggiungere un nuovo campo, vuoto, dove inserire le informazioni richieste. Ovviamente l'inserimento di un nuovo campo avviene senza perdere le informazioni già inserite nei campi precedenti.

4. *I blocchi d'informazione ripetibili* si presentano con affianco il pulsante "New " che permette l'aggiunta di un nuovo blocco. Quando uno o più blocchi

sono stati aggiunti, a destra di questo pulsante appare una lista di numeri relativi a ciascun oggetto che permettono di visualizzare le informazioni precedentemente inserite nei vari blocchi. I nuovi blocchi vengono sempre inseriti in fondo.

5. *La lunghezza dei campi a testo libero* è stata fissata a valori diversi secondo il tipo di campo. Per i campi che devono contenere titoli, nomi o codici la lunghezza massima è stata fissata a 60 caratteri. Per i campi a testo libero che contengono brevi descrizioni la lunghezza è stata fissata a 3500 caratteri. Infine, per quei campi, come ad esempio l'abstract, per cui è richiesta una quantità di spazio superiore per una descrizione maggiormente dettagliata la lunghezza massima del contenuto è stata portata a 65000 caratteri. Queste lunghezze vengono controllate al momento dell'immissione impedendo l'inserimento di nuovi caratteri quando la lunghezza supera il limite fissato. Per evidenziare quale è la lunghezza massima consentita e quella corrente del contenuto viene mostrato un messaggio sulla barra di stato del browser.

4.4. Thesaurus e parole chiave

Il blocco d'informazioni denominato "classification" permette la selezione di una o più parole chiave da associare al dataset che si sta descrivendo. Queste parole chiave possono provenire da diversi *thesaurus* definiti secondo standard internazionali o all'interno del settore di ricerca interessato. Come visualizzato in Figura 6, l'interfaccia si presenta divisa in due gruppi di controlli:

- il primo per la selezione del thesaurus e delle parole chiave da utilizzare
- il secondo che elenca e permette di rimuovere le parole chiave scelte finora.

Operativamente l'utente deve innanzitutto scegliere dall'apposita lista quale thesaurus intende utilizzare. Fatto ciò, alla pressione del pulsante "Add keyword(s)

from the selected thesaurus”, viene aperta una nuova finestra che permette di navigare tra i termini appartenenti al thesaurus scelto.

Questa finestra è personalizzata per ciascun thesaurus, per meglio sfruttare le caratteristiche del singolo thesaurus, come ad esempio la possibilità di scegliere la categoria di termini e quindi accedervi attraverso un link.

La scelta di uno qualunque dei termini della categoria aggiunge tale parola chiave a quelle relative al metadato che si sta compilando, per cui la lista delle parole chiave dell’interfaccia principale viene aggiornata comprendendo anche il nuovo elemento.

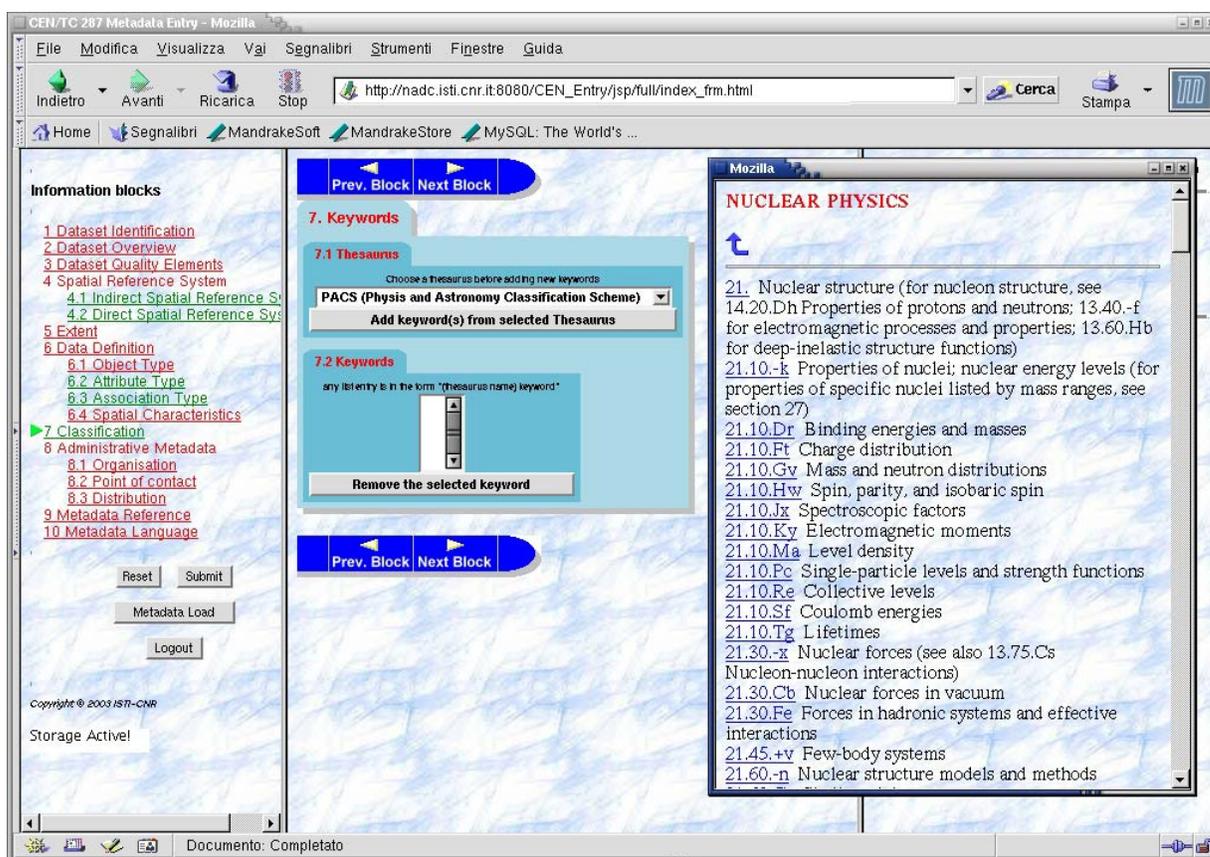


Figura 8. Selezione delle keyword dai Thesaurus.

Dalla lista delle parole scelte è possibile eliminare un termine semplicemente selezionandolo e cliccando sul pulsante “Remove the selected keyword”.

I thesaurus attualmente disponibili sono il PACS (Physics and Astronomy Classification Scheme), un thesaurus specifico per il settore 1 (Biology and medicine) denominato "Sector 1 custom" ed uno specifico per il settore 9 (Chemical contamination) con il nome "Sector 9 custom".

4.5. Selezione di un'area geografica

Nel blocco d'informazioni "Bounding XY" è presente un pulsante denominato "Visual selection" che apre una nuova finestra (vedi Figura 7) in cui un applet permette la selezione visuale di un'area geografica. La selezione avviene sfruttando una mappa dell'Antartide su cui, agendo col mouse, l'utente può far apparire una porzione di anello attorno al polo rappresentante l'area selezionata.

La selezione di tale area avviene puntando il cursore del mouse su un angolo dell'area d'interesse; premendo poi il pulsante sinistro e spostando il cursore senza rilasciare il pulsante l'area selezionata si allarga in senso antiorario considerando la posizione corrente del cursore come l'estremo opposto dell'area selezionata.

Una volta che il pulsante del mouse viene rilasciato l'area corrente viene accettata e le sue coordinate sono mostrate nelle apposite caselle di testo.

La selezione può anche avvenire inserendo direttamente nelle caselle di testo le coordinate dell'area da selezionare: così facendo sulla mappa appare la rappresentazione dell'area indicata.

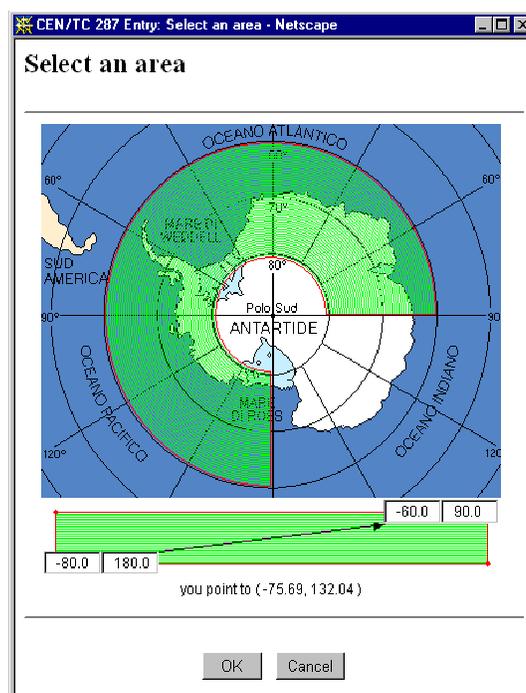


Figura 7. Esempio dell'aspetto del tool per la selezione di un'area geografica

Per trasferire le coordinate dell'area selezionata nelle caselle del blocco d'informazioni "Bounding XY" è sufficiente cliccare sul pulsante "Ok". Se si agisce, invece, sul pulsante "Cancel" la finestra viene chiusa lasciando invariate le coordinate presenti nel blocco d'informazioni "Bounding XY".

4.6. Invio delle informazioni per la registrazione (Submit)

Dopo aver completato l'inserimento di tutte le informazioni necessarie e dopo una doverosa verifica di correttezza e completezza, queste possono essere inviate al server per la loro registrazione nel catalogo. La registrazione può essere di due tipi distinti: una registrazione di un'immissione parziale del dataset, effettuata per sospendere tale operazione e riprenderla in un secondo momento, oppure una registrazione effettiva.

La registrazione di un'immissione parziale di un metadato consiste nel memorizzare le informazioni inserite in un file conservato in un'apposita area locale del server, così da poter essere riprese in un secondo momento per essere completate o corrette. In questo caso, i controlli effettuati durante questo tipo di registrazione sono minimi (si limitano al controllo delle lunghezze dei campi), poiché molti campi possono risultare incompleti.

A questo scopo è presente nel frame di sinistra il pulsante "Submit" (cfr. pag. 5). Agendo su questo pulsante viene aperta una nuova finestra (detta "di registrazione") che mostra tutti i campi del metadato con i valori assegnati dall'utente.

La Figura 8 mostra la struttura della pagina per la registrazione. Questa finestra è divisa in tre frame: quello più in alto per la selezione dei file allegati, quella centrale per la visualizzazione del contenuto dei campi riempiti dall'utente e quella inferiore contenente il controllo sull'effettiva esistenza dei file allegati.

Operativamente l'utente per registrare un metadato deve: verificare la correttezza delle informazioni inserite, riempire tutti i campi relativi ai file allegati. Al momento della registrazione le due parti più in alto nella finestra vengono aggiornate con l'esito delle operazioni parziali (l'invio dei file allegati, nella parte alta, e la registrazione dei

vari campi del metadato, in quella centrale) indicando la loro buona riuscita o l'avvenimento di un qualche problema che ne ha impedito il completamento.

La registrazione effettiva del metadato consiste invece nella sua registrazione nel database del sistema, il che richiede che tali file siano prima sottoposti all'attenzione di un supervisore competente che provvederà alla verifica di correttezza e completezza.

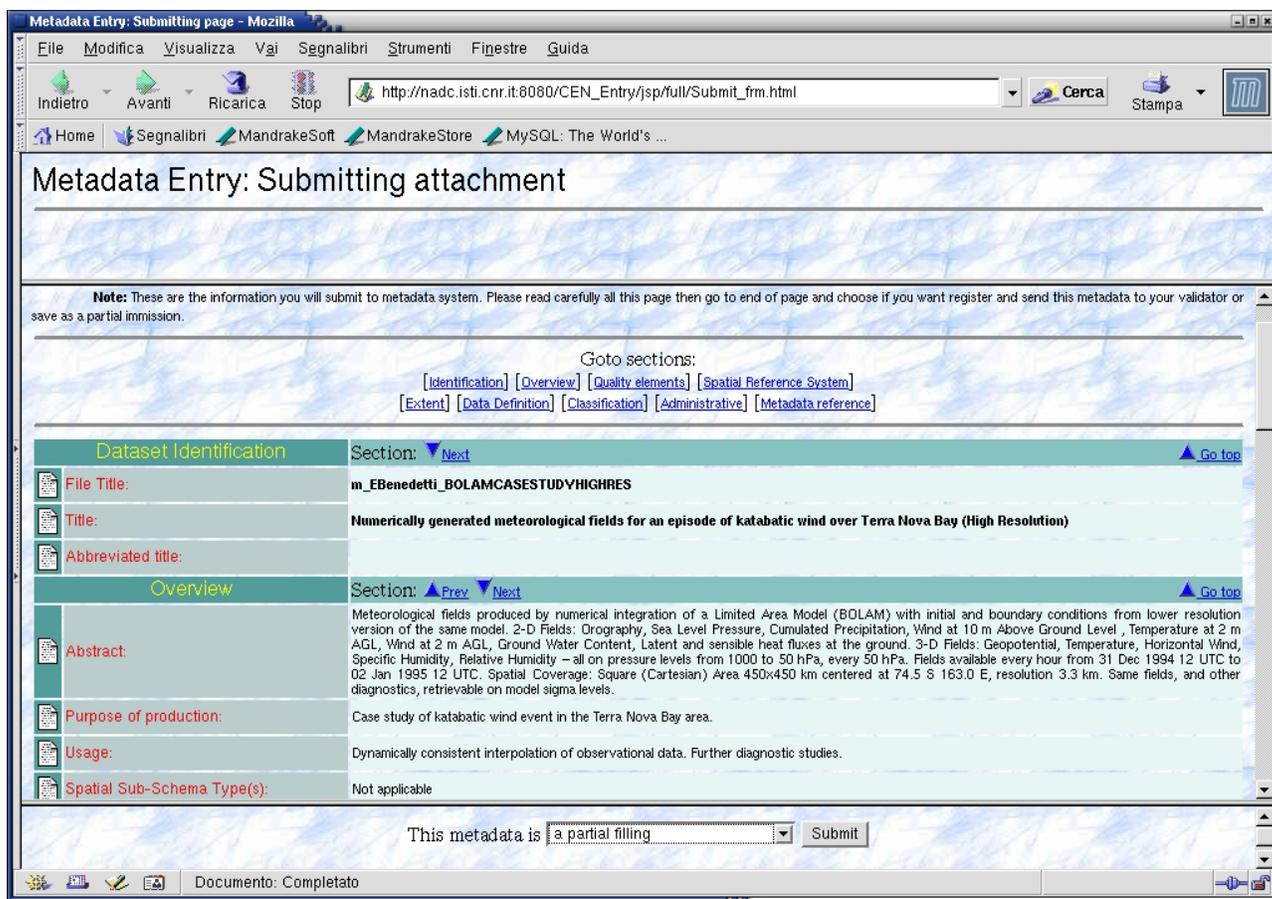


Figura 8. Aspetto della finestra per la registrazione delle schede di metadato.

Questo comporta che il metadato subisca un controllo accurato per verificare che tutti i campi obbligatori siano in qualche modo riempiti (anche se con termini del tipo "Not applicable"). Tale registrazione avviene ad opera di ciascun referente di settore, il quale accedendo via web al modulo di Metadata Entry, ha la possibilità di vedere tutte le schede di metadato compilate per il proprio settore e di validarle qualora tali schede

siano state completate in via definitiva dai vari ricercatori. Il referente, dopo un controllo accurato di ciascuna scheda, può decidere di inviarla in modo definitivo nell'apposita area del server. Tale operazione comporta la cancellazione dall'area temporanea del file relativo. La spedizione del file di metadato compilato comprende tutti gli altri eventuali file allegati (immagini di esempio, campioni dei dati, file di documentazione, bibliografie ecc.). La successiva fase di registrazione effettiva nel database avviene a cura dell'ISTI.

4.7. Uso delle immissioni parziali

Una volta salvate, le immissioni parziali possono essere riprese in un secondo tempo per essere completate. Per far questo il sistema prevede una finestra per la gestione delle immissioni parziali che può essere aperta agendo sul pulsante "Partial filling" presente sotto il pulsante di "Submit".

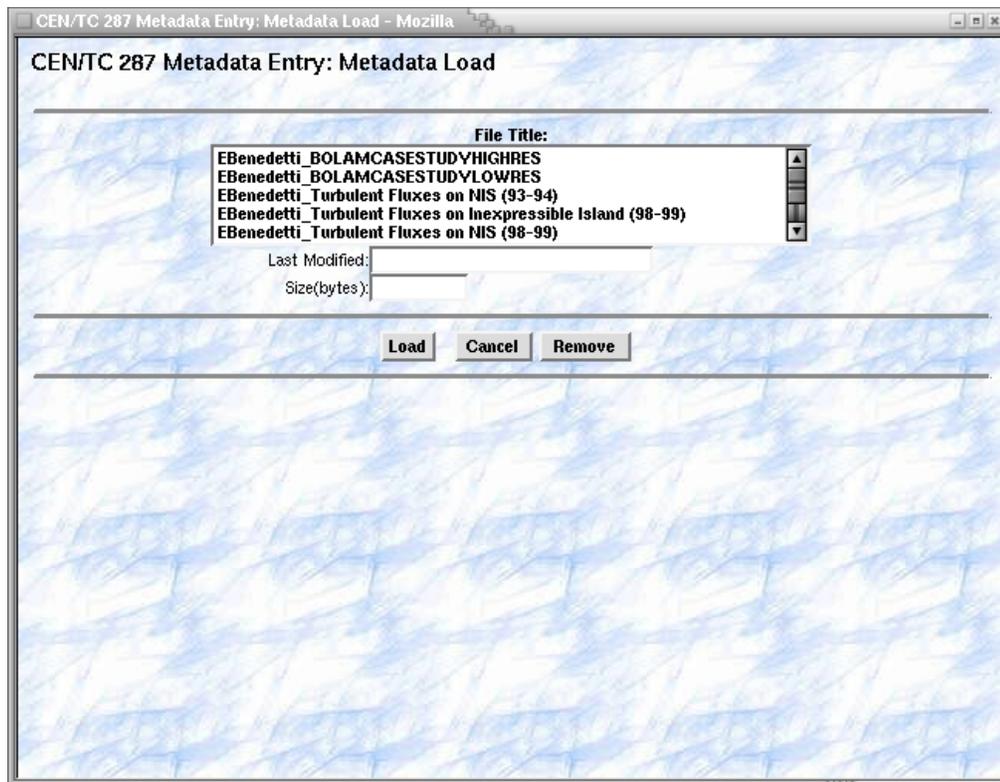


Figura 9. Aspetto della finestra per il caricamento delle schede di metadato.

Come si può vedere nella Figura 9, questa finestra mostra la lista delle immissioni parziali registrate permettendo il loro caricamento o la loro rimozione.

Per effettuare il caricamento delle informazioni occorre selezionare l'elemento che si desidera caricare (il nome del file) e premere il pulsante "Load". In questo modo la pagina viene ricaricata e fornisce l'esito dell'operazione.

Dopo il caricamento tutte le informazioni precedentemente inserite vengono cancellate e sostituite con quelle presenti del file caricato, tutti i campi non utilizzati vengono svuotati e tutti i campi e i blocchi ripetibili, se non utilizzati, vengono rimossi.

La finestra per la gestione delle immissioni parziali permette anche la cancellazione delle precedenti immissioni. Questa scelta è stata preferita a quella della cancellazione automatica dopo la registrazione effettiva di un metadato poiché queste immissioni parziali possono essere utilizzate come modelli base.

5. Note tecniche

5.1. Requisiti per l'uso

Il sistema basa il suo funzionamento su applet Java per la funzionalità di selezione dell'area geografica e per memorizzare alcune informazioni, codice JavaScript per controlli di inserimento di informazioni, servlet java e pagine java-server-pages (jsp) per la generazione dinamica delle pagine e lo sviluppo della logica business. Per la resa estetica si utilizzano fundamentalmente tabelle e i Cascading Style Sheets (CSS). Per un pieno utilizzo del sistema Metadata Entry è quindi necessario un browser Web in grado di sfruttare tutte queste tecnologie, verificando inoltre che sia configurato in modo che esse siano abilitate.

5.2. Testing

Il sistema è stato testato con vari sistemi operativi e browser:

- con Windows XP/2000/NT utilizzando il browser Netscape Communicator (NS) rel. 4.5 o superiore., e Internet Explorer 5.0 (IE) e 6.0 scaricando il relativo Java Plugin all'indirizzo <http://java.sun.com/j2se/1.4.2/download.html>;
- con Linux, dopo aver scaricato il Java Plugin per i browser Netscape, Mozilla e Galeon disponibili allo stesso indirizzo;
- con Mac OS X, che include la versione completa di Java 2 Standard Edition (versione 1.4.2) e la Virtual Machine HotSpot, utilizzando Safari 1.2.