

MetaPub : il management e il recovery del sistema

marzo 2007

Giuseppe A. Romano, e-mail: giuseppe.romano@isti.cnr.it
Roberto Salis, e-mail: roberto.salis@isti.cnr.it

Istituto di Scienza e Tecnologie dell'Informazione "A.Faedo"
Area della Ricerca del CNR, Pisa, Via Moruzzi 1 56123

Indice

Il sistema MetaPub	4
Il caricamento dei dati e il loro aggiornamento	6
La procedura get_loadCNRpubbs	7
La procedura loadCNRpubbs	8
La procedura loadaix	10
Come aggiungere un database al sistema.	10
L'aggiornamento delle interfacce utente	12
Definire le interfacce utente per un nuovo database	12
Aggiornare le interfacce utente per Metapub	15
L'esportazione dei dati contenuti nei database di MetaPub	18
La struttura del documento bibliografico	18
L'esportazione dei dati verso il database del minisito del CNR	21
Come scaricare i dati di un nuovo Istituto	23
Il recovery del sistema	25
<u>Bibliografia</u>	27

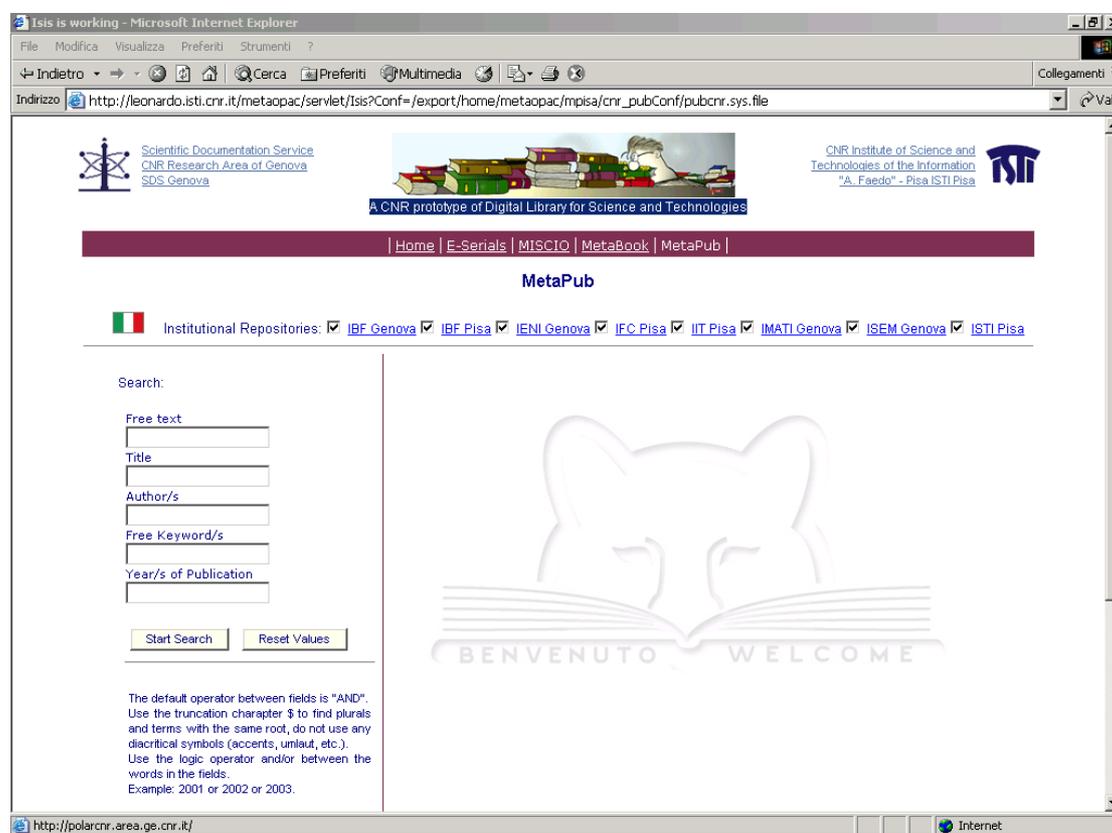
Il sistema MetaPub

Il MetaPub è un sistema distribuito per l'accesso sia ai metadati che alle pubblicazioni stesse degli Istituti del CNR che hanno partecipato/partecipano al progetto per la costituzione di "A CNR prototype of Digital Library for Science and Technologies".

Il servizio MetaPub è accessibile alla URL:

http://leonardo.isti.cnr.it/metaopac/servlet/Isis?Conf=/export/home/metaopac/mpisa/cnr_pubConf/pubcnr.sys.file

Di seguito la Welcome Page del servizio:



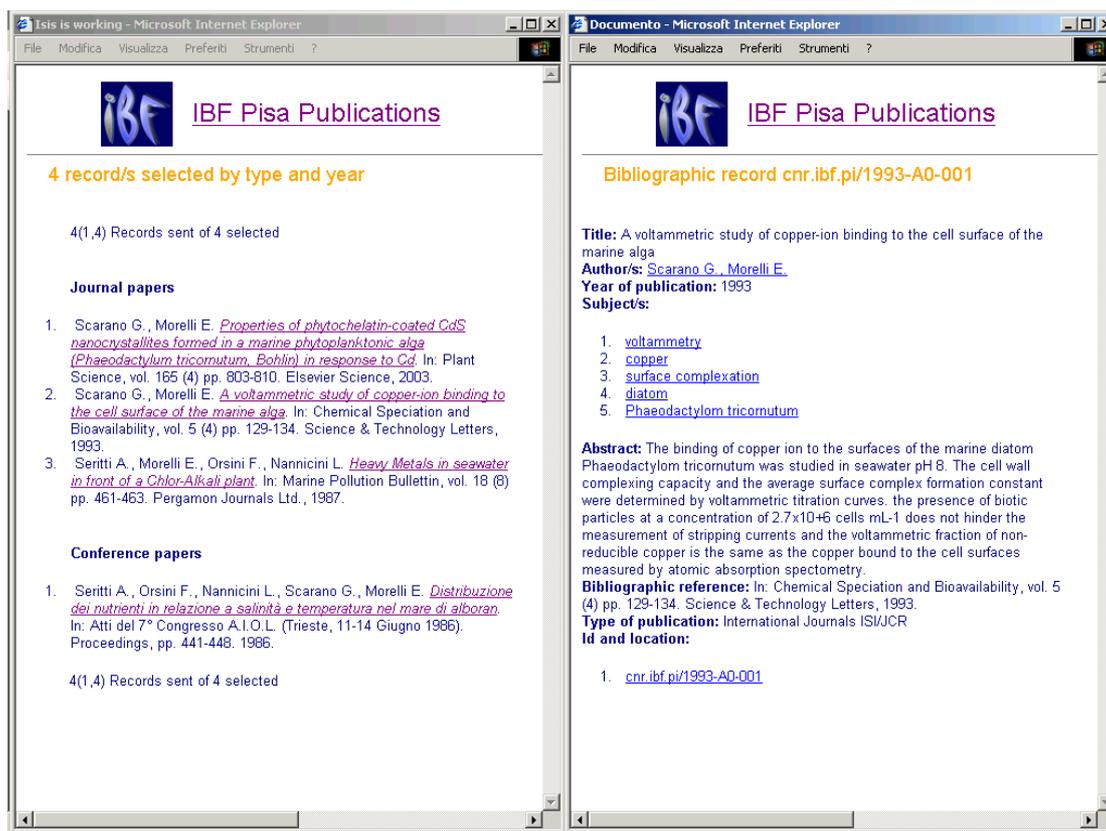
Il sistema è scalabile e, anche se al momento i cataloghi delle pubblicazioni sono installati su un unico server, possono essere distribuiti capillarmente sui server degli Istituti afferenti al progetto.

Il sistema utilizza, per la ricerca e la presentazione dei documenti il software ExtGAS; vedi:

1. Isis/GAS: un sistema multi piattaforma per l'accesso a basi di dati testuali via Internet

2. Isis/GAS: manuale di installazione e d'uso
3. ExtGAS: un sistema multiplatforma per l'accesso ai cataloghi di sistemi biblioteconomici eterogenei
4. ExtGAS: manuale di installazione e d'uso
5. ExtGas/IsisGAS revisione 2.0 - manuale di installazione e d'uso

Il sistema MetaPub permette l'accesso alle informazioni sopra menzionate restituendo liste organizzate Istituto per Istituto e ordinate per tipologia di pubblicazione in ordine decrescente di anno.



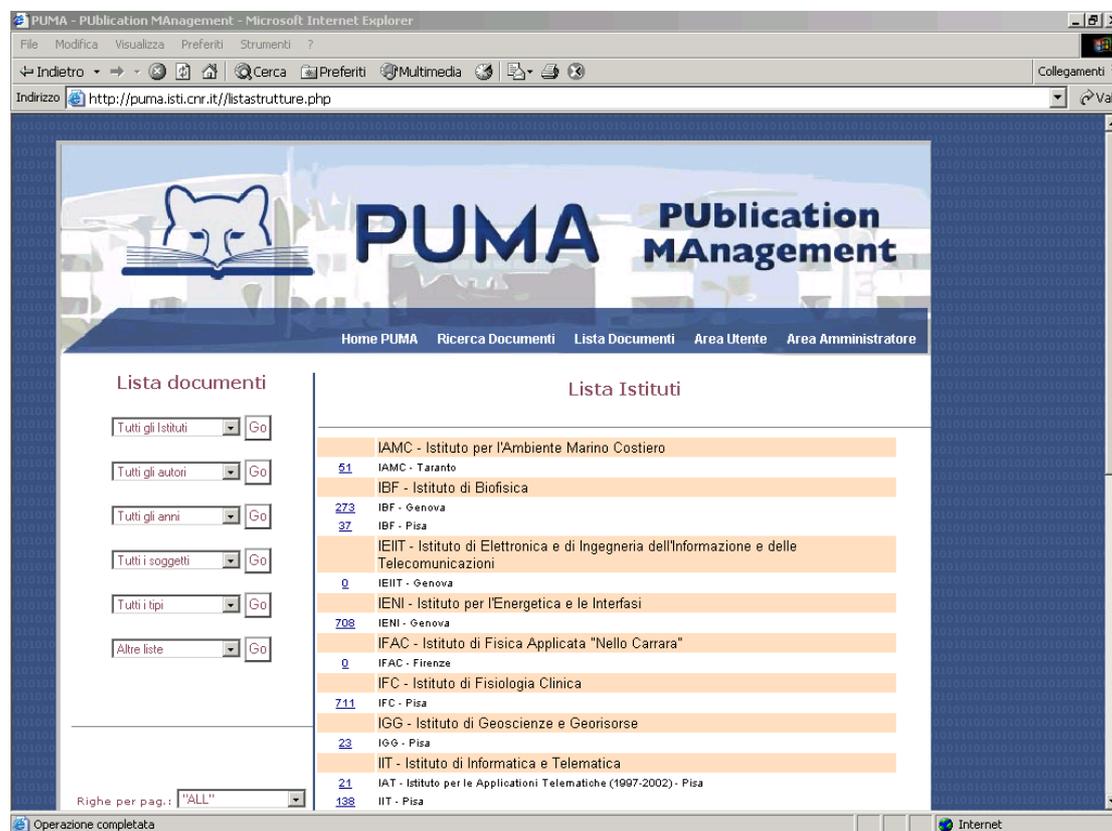
E' inoltre possibile interrogare direttamente il catalogo delle pubblicazioni di uno specifico Istituto selezionandolo dall'area "Institutional Repositories" o dall'intestazione del display del contenuto della ricerca così come evidenziato nell'immagine qui sopra.

I dati vengono caricati sul database delle pubblicazioni direttamente dal repository tramite una serie di procedure automatiche che garantiscono l'aggiornameto giornaliero delle informazioni.

Scopo di questo manuale è spiegare come aggiornare, modificare, distribuire l'informazione e recuperare le funzionalità del server in caso di crash del sistema.

Il caricamento dei dati e il loro aggiornamento

Il sistema MetaPub acquisisce i metadati delle pubblicazioni dal repository ufficiale del sistema PUMA (Publication Management) accessibile alla URL <http://puma.isti.cnr.it/>



I dati vengono resi accessibili tutte le notti dall'http server del sistema PUMA alla URL

<http://puma.isti.cnr.it/PumaDB.bib/>

Il server principale del sistema MetaPub è collocato su **metabsrv.isti.cnr.it**. Esso contiene tutte le procedure per lo scaricamento e l'aggiornamento del sistema oltre ai database del sistema. I server secondari, invece, sono collocati su **dell11.isti.cnr.it** e **orione.isti.cnr.it**. I server secondari possiedono i database del servizio MetaPub, vengono aggiornati al termine della procedura di aggiornamento del server principale e sono i metaopac MetaPub server operativi.

L'aggiornamento dei database avviene ogni mattina dal martedì al sabato alle 05.30 attivando sul server principale la procedura:

```
/export/home/metaopac/mpisa/proc/get_loadCNR_pubbs
```

che carica selettivamente i file disponibili sul server.



La procedura **get_loadCNRpubbs**:

1. Si posiziona sulla directory `/home/metaopac/mpisa/IsisDbdir/cnr_ienige`.
2. Salva, se esiste, il precedente file di un Istituto in un file temporaneo.

```
mv -f /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.ibf.ge_bib.txt \  
/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.ibf.ge_bib.txt_saved
```

3. Copia il nuovo file del relativo Istituto tramite il comando "wget".

```
/usr/bin/wget -nv -O  
/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.isib.pd_bib.txt  
"http://puma.isti.cnr.it/PumaDB.bib/cnr.isib.pd_bib.txt"
```

4. Se non esiste il nuovo file scaricato (qualche problema di connettività o altro) ripristina il file salvato al punto 2.

```
if ( [ ! -s /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.ibf.ge_bib.txt ] ) then  
cp -f -p -d /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.ibf.ge_bib.txt_saved  
/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/cnr.ibf.ge_bib.txt  
fi
```

5. Manda in esecuzione la procedura:

```
/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/loadCNRpubbs
```

6. Termina l'esecuzione dell'aggiornamento del sistema.

Volendo aggiungere un nuovo Istituto acquisendo il relativo file dei metadati è necessario aggiungere il punto 2. , 3. , 4. per il relativo file.

La procedura **loadCNRpubbs** è collocata nella directory

`/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/`

e è casuale che il nome della directory corrisponda, almeno approssimativamente, al nome di uno degli Istituti che partecipano a servizio MetaPub: per motivi di continuità, anche la directory di configurazione di Metapub ha circa lo stesso nome `cnr_ienigeConf`.

La procedura `loadCNRpubbs` esegue la conversione dei dati ricevuti con la procedura precedente e li memorizza in una unica directory mantenendo separati i file generati. In questo modo ogni database potrebbe essere generato separatamente e su computer separati; nel caso particolare i dati vengono assemblati in un unico file che costituisce il file di caricamento del database. In tutti i casi i database caricati nello stesso oggetto database cumulato sono sempre logicamente separati perché per ogni documento appartenente a ognuno di essi viene generato un codice univoco che lo identifica nel contesto del database cumulato. La procedura:

1. Si posiziona sulla directory
`/export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/`
2. Esegue il programma `FromXMLtoIsis` che converte i dati da formato XML-PUMA a CDS/ISIS iso2709. Il programma accetta 3 parametri:
 - a - il nome del file che descrive i file di input
 - b - la directory di output
 - c - `trace/notrace` se necessario mettere in traccia il programma e i dati relativi

Inoltre, il programma carica nel tag/field 900 l'identificazione univoca della authority cui si riferisce il database. Se l'authority fosse `cnr.ibf.pi` verrebbe caricato il valore `IBFPI` cioè eliminando la specifica generica "cnr". Questo permette di avere sullo stesso database più collezioni separate.

3. Salva i precedenti dati da caricare da `test_pub.iso` a `test_pub.iso_saved`.
4. Copia tutti i file generati nel file `test_pub.iso`
5. Cancella il database corrente (`delheader`)
6. Carica il nuovo database (`loadiso.cnr`)
7. Genera l'inverted file (`genproc.cnr`)
8. Genera gli indici alternativi per l'accesso (titoli, autori, soggetti)
9. Copia i file dei dati da caricare (`test_pub.iso*`) sul computer `orione.isti.cnr.it` e `dell11.isti.cnr.it`

10. invia a orione.isti.cnr.it e dell11.isti .cnr.it i comandi per eseguire sui computer remoti quanto eseguito dal punto 5 al punto 8.
11. Esegue la procedura di salvataggio dei dati e dei database generati (taring_cnr_prj_vc)

Il file che descrive i dati di input, punto 2.a, può contenere un massimo di 50 descrizioni di database separati. Ogni riga descrive un database:

```
cnr.ieni.ge_bib.txt ienige <dc:begin> </dc:end> 39
```

Il primo parametro definisce il file di input

Il secondo il nome del file out che conterrà i dati convertiti seguito dal suffisso ".iso"

Il terzo parametro indica il tag XML che determina l'inizio di una descrizione bibliografica

Il quarto indica il tag XML che definisce la fine di una descrizione bibliografica

Il quinto parametro è delegato a uso futuro

Per i punti 5-8 si veda:

"Il MetaOpac Pisano: l'organizzazione logica e il mapping fisico" - ISTI-B4-02, 2003

"Il Metaopac del CNR di Pisa e la sua collocazione nel Metaopac pisano" - ISTI-B4-14, 2004.

Il punto 8 in particolare richiama la procedura di generazione degli indici di supporto alla ricerca (autori, titoli, soggetti). Ogni indice è un database di supporto e come tale i dati vanno estratti dai database dei Cataloghi, generati in formato iso2709, caricati sui database di servizio e costruiti gli inverted file. Ogni indice risiede su una directory separata che è genericamente collocata nella directory:

```
/export/home/metaoapc/mpisa/AIX/cnrpub/
```

ogni directory contiene il database che costituisce l'indice di appoggio, le procedure per generare i dati e caricare il database e , inoltre, una procedura specifica che permette di copiare i dati estratti ordinatamente nelle directory:

1. titolo.total
2. soggetto.total
3. autore.total

Tali directory contengono i database degli indici totali per autore, soggetto, titolo.

La procedura che estrae, genera, carica il database e genera l'inverted file è per ogni directory loadaix, ed è specializzata per lo specifico indice.

La procedura **loadaix**:

Di seguito viene trattata, genericamente, la procedura loadaix e alcune implicazioni che la caratterizzano. Non è specificatamente necessario conoscere il funzionamento delle procedure di caricamento e generazione dell'inverted file dei database CDS/ISIS.

1. Cancellazione dei file che costituiscono il database e del file di input iso2709 che generalmente si chiama <dbaix>.iso
2. Estrazione dei dati, che costituiranno l'indice di supporto, dal database specifico utilizzando la API **wwwisis** che accetta in input il contenuto del file **in.query**. Vedi: Il MetaOpac Pisano: l'organizzazione logica e il mapping fisico - ISTI-B4-02, 2003 - Gli indici globali e locali pag. 35-36.
3. Ordinamento alfabetico del file estratto.
4. Generazione del file iso2709 che contiene i dati dell'indice da caricarsi su un database CDS/ISIS. Vedi: Il MetaOpac Pisano: l'organizzazione logica e il mapping fisico - ISTI-B4-02, 2003 - Il programma di conversione **aixexloadiso**/l'accesso agli indici e il loro utilizzo pag. 36-42
5. caricamento del database con la procedura aix.loadiso
6. generazione dell'inverted file con la procedura aix.genproc
7. cancellazione dei file temporanei e salvataggio del file di input per aixexloadiso nella directory <type>.total in modo da poter generare alla fine un indice cumulato specifico per il tipo (autore, titolo, soggetto).

Come aggiungere un database al sistema.

Volendo aggiungere un nuovo database al sistema occorre:

1. Aggiornare la procedura **get_loadCNRpubbs** nei suoi punti 2, 3, 4
2. Aggiornare il file descritto in 2.a di **loadCNRpubbs**
3. Aggiornare la procedura **genindexes**, al punto 8 di **loadCNRpubbs**, per aggiungere le directory che contengono i nuovi indici e mettere a punto le procedure di estrazione, conversione, e generazione di nuovi indici.
4. Aggiornare la procedura **/home/metaopac/taring_cnr_prj_vc** per il recovery del sistema.

In modo particolare aggiungere il nome della directory di configurazione del nuovo database al comando `tar -czf cnr_prj_vc.tgz <list>`

Naturalmente il tutto va replicato sui server dell11.isti.cnr.it e orione.isti.cnr.it escludendo le procedure di acquisizione dei dati e la loro conversione e la procedura **taring_cnr_prj_vc**. Si guardino la procedura **loadCNRpubbs** dei due

computer:

```
cd /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige
rm -f -r tmp
sh loadiso.cnr
/bin/bash genproc.cnr
sh genindexes
rm -f -r tmp
cd
```

che sono attivate dai comandi

```
ssh dell11.isti.cnr.it sh /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/loadCNRpubbs
e
ssh orione.isti.cnr.it sh /export/home/metaopac/mpisa/IsisDbdir.cnr_ienige/loadCNRpubbs
```

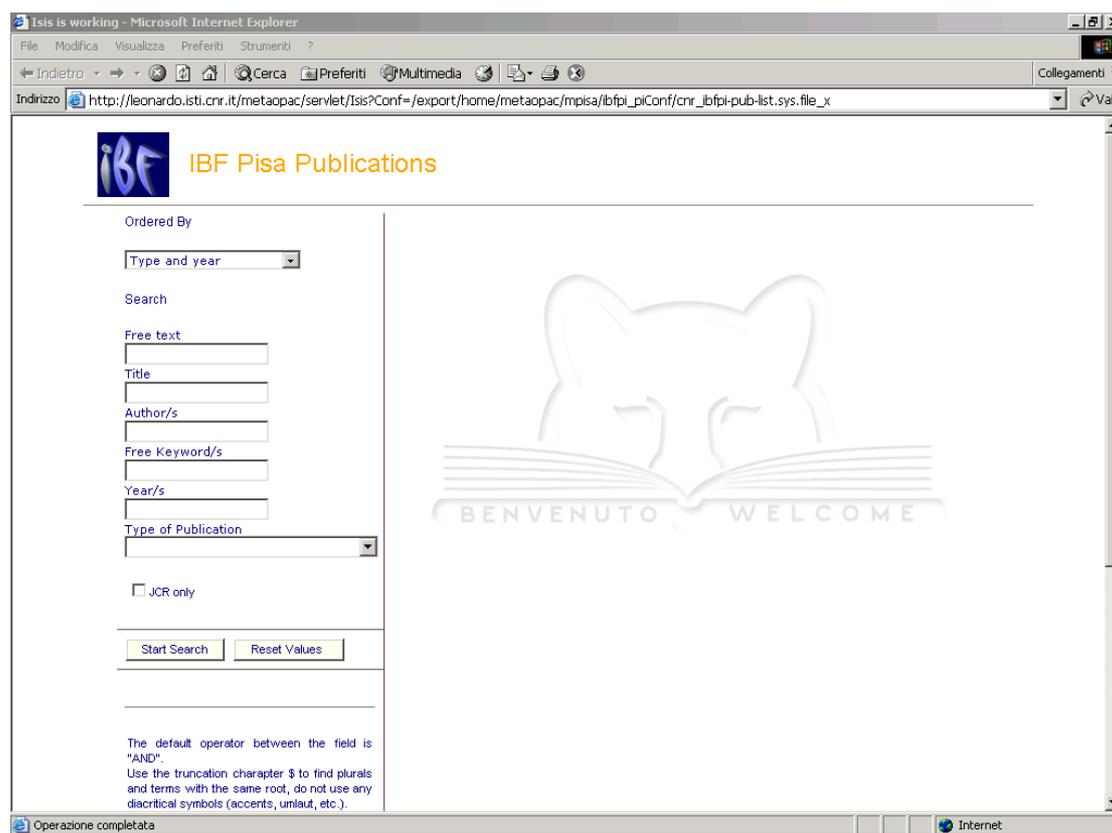
Tale procedura contiene solo i punti 5,6,7,8 della stessa procedura residente su metabsrv.isti.cnr.it

L'aggiornamento delle interfacce utente

Oltre alla interfaccia utente generale di MetaPub descritta a pag. 4 esistono, per ogni database del servizio 2 interfacce utenti generiche che possono essere descritte in più Database System File a seconda del multilinguismo richiesto (nel senso che dovrebbero essere multipli di 2 interfacce utente)

1. Un Database System File descritto genericamente come `cnr_<dbname>-pub-list.system.file` che viene utilizzato durante l'accesso ai dati, dei database componenti, dall'interfaccia multidatabase (quella descritta a pagina 4).
2. Un Database System File descritto genericamente come `cnr_<dbname>-pub-list.system.file_x` che viene utilizzato durante l'accesso a uno specifico database.

La prima si differenzia dalla seconda perché essendo utilizzata dal metamotore ha la possibilità di attivare la pagina dello specifico database.



Si confronti la predente immagine con quella a pagina 5.

Per aggiungere al servizio un nuovo database bisogna:

1. Definire le interfacce utenti generali per il database

2. Aggiornare l'interfaccia generica di MetaPub

Definire le interfacce utenti per un nuovo database

Per la definizione delle interfacce utenti per uno specifico database conviene copiare e modificare una interfaccia già operativa. Normalmente le interfacce utente sono collocate in una specifica directory per ogni database del servizio. Al momento esse sono collocate in:

```
/home/metaopac/mpisa/isti_piConf per ISTI
/home/metaopac/mpisa/ibfpi_piConf per IBF-PI
/home/metaopac/mpisa/ifc_piConf per IFC
/home/metaopac/mpisa/iit_piConf per IIT
/home/metaopac/mpisa/ienige_geConf per IENI-GE
/home/metaopac/mpisa/imatige_geConf per IMATI-GE
/home/metaopac/mpisa/ibfge_geConf per IBF-GE
/home/metaopac/mpisa/isemge_geConf per ISEM-GE
```

Se analizziamo il contenuto di /home/metaopac/mpisa/ienige_geConf

1. cnr1TiLcount.pft: è il formato di stampa per la restituzione del numero di record selezionati se la richiesta arriva dal metamatore MetaPub.

Nel formato:

```
/* THISCOUNTPFT
*/
mhl,if val(v1001) = val(v1031) then '<Prolog>' ,/, 'Pubblicazioni IENI-
GE' ,/, '</Prolog>' ,/, '<NumberOfSelectedRecords>' ,/, F((val(v1032)-
val(v1031)+1),3,0), ':', v1002, /, '</NumberOfSelectedRecords>' ,fi, /,
```

va modificato il valore IENI-GE nel valore relativo al nuovo database.

2. cnr_ienige-pub-list.sys.file: è la definizione dell'interfaccia in inglese per l'accesso ai database tramite il metamatore.

Va modificato il nome del file in relazione al nuovo database e vanno modificate le linee:

```
QueryText <br><p class="menu_query"><input type="checkbox" name=Field9
value="JCR/(910) not (B5/(190))">JCR only<input type=hidden name=Field5
value="IENIGE"><input type=hidden name=Rqar value=y</p>
```

```
GeneralQuerySuffix AND ap/(904) and IENIGE/(900)
```

```
IsisHeaderArea <table border=0 cellpadding=0 cellspacing=0 width=500
```

```
ALIGN=""><tr><td ALIGN="left"><a href="">
IENI Genova Publications</a></font></td></tr></table>
```

In modo da allineare il valore del nome del database IENIGE/IENI, i nomi dei file immagine e il nome della directory di configurazione al nuovo nome del database.

Mentre le linee:

```
DisplayAreaContentFile
/export/home/metaopac/mpisa/ienige_geConf/ieni_ge.Welcome.file
```

```
IsisCiparFileName /export/home/metaopac/mpisa/ienige_geConf/cnr.txt
```

```
LocalJavascriptFunctions
/export/home/metaopac/mpisa/ienige_geConf/cnr-pJavascript.file
```

Vanno modificate per allinearle al nuovo directory di configurazione del database e al nuovo nome del database

3. cnr_ienige-pub-list.sys.file-it: è la definizione dell'interfaccia in italiano per l'accesso dal metamotore

Come per il punto 2.

4. cnr_ienige-pub-list.sys.file-it_x: è la definizione dell'interfaccia in italiano per l'accesso allo specifico database

Come per il punto 2.

5. cnr_ienige-pub-list.sys.file_x: è la definizione dell'interfaccia in inglese per l'accesso allo specifico database

Come per il punto 2.

6. cnr-pJavascript.file: è il Javascript file

Deve essere, se è il caso, aggiornato

7. cnr-pListtibib_r.pft: è un formato di stampa
8. cnr.txt: è il file che descrive la locazione dei database, indici, formati di stampa e di ordinamento.

Deve essere fatto un change globale tra la vecchia directory di configurazione "ienige_geConf" nella nuova che descrive il nuovo database

Inoltre vanno cambiate le tre righe:

```
aixau.*=/export/home/metaopac/mpisa/AIX/cnrpub/autore.ienige/aix.*
aixsu.*=/export/home/metaopac/mpisa/AIX/cnrpub/soggetto.ienige/aix.*
aixti.*=/export/home/metaopac/mpisa/AIX/cnrpub/titolo.ienige/aix.*
```

In modo che gli indici di autori, titoli, soggetti puntino ai nuovi database.

9. epillog.cnr-pub: è il formato di stampa utilizzato alla fine di ogni ricerca per interpretare i risultati.

La stringa "Pubblicazioni IENI-GE" contenuta nel file va sostituita con una compatibile con il nuovo nome del database

10. ieni_ge.Welcome.file: Welcome file in inglese

Deve essere rinominato coerentemente con il nuovo database

11. ieni_ge.Welcome.file-it: Welcome file in italiano

Deve essere rinominato coerentemente con il nuovo database

12. ieni_ge.Welcome.file.orig: Welcome file di salvataggio

13. unipiAixTi_x_.pft: formato di stampa dell'indice dei titoli

14. unipiE_x_ti.pft: formato di estrazione dei dati per la costruzione dell'indice dei titoli.

Aggiornare le interfacce utente per Metapub

L'aggiornamento delle interfacce utente per Metapub è una operazione più veloce poiché è necessario fare conoscere a metapub solo la locazione del nuovo database.

La directory in cui è collocata la definizione di MetaPub è:

```
/export/home/metaopac/mpisa/cnr_pubConf
```

e tale directory esiste sui tre computer:

1. dell11.isti.cnr.it

2. metabrv.isti.cnr.it
3. orione.isti.cnr.it

Si tenga presente che metabrv.isti.cnr.it è il server operativo per l'aggiornamento dei database mentre non lo è come server per l'accesso ai database; per cui si consiglia di aggiornare l'interfaccia MetaPub su metabrv.isti.cnr.it, provarla e poi esportare i relativi file sui rimanenti 2 computer.

Nella directory su descritta sono da modificare i seguenti file:

1. pubcnr.sys.file-en, pubcnr.sys.file-it: si tenga presente che pubcnr.sys.file è un link simbolico a pubcnr.sys.file-en

Aggiungere al file una riga identica a:

```
IsisHeaderArea <input type=checkbox name="db" value=2 CHECKED> <a href="http://leonardo.isti.cnr.it/metaopac/servlet/Isis?Conf=/export/home/metaopac/mpisa/ienige_geConf/cnr_ienige-pub-list.sys.file_x" target="db2" class="collection">IENI Genova</a>
```

modificando:

1. l'href relativo al nuovo database
 2. il nome del database contenuto nell'anchor point,
 3. value=2 e target="db2" in modo da allinearli alla posizione che il nuovo database troverà nella lista dei database, essendo tale lista ordinata alfabeticamente (Attenzione: la lista va completamente aggiornata).
2. pubcnr.LogicDbFile che contiene la locazione dei database da accedere

Inserire nel file, al posto di sua competenza rispetto all'ordinamento alfabetico, una riga identica a:

```
http://leonardo.isti.cnr.it/metaopac/servlet/Isis
/export/home/metaopac/mpisa/ienige_geConf/cnr_ienige-pub-list.sys.file
10000 100 . .
```

Modificando la directory di configurazione e il relativo file di configurazione. Se la nuova directory fosse aaaConf e il relativo file di configurazione aaa-pub-list.sys.file si avrebbe:

```
http://leonardo.isti.cnr.it/metaopac/servlet/Isis
/export/home/metaopac/mpisa/aaaConf/aaa-pub-list.sys.file
10000 100 . .
```

Si tenga presente che il database potrebbe essere collocato su un computer diverso e remoto e quindi la definizione potrebbe essere più complessa e richiedere la modifica del primo parametro nella definizione del server nel Logic DbFile.

Per maggiori dettagli per la configurazione del Logic DbFile si veda:
"ExtGas/IsisGAS revisione 2.0: manuale di installazione e uso" - ISTI-B4-16, 2006 - pagine 5-6.

L'esportazione dei dati contenuti nei database di MetaPub

L'esportazione dei dati di MetaPub deve essere fatta coerentemente database per database. Infatti i database, anche se memorizzati in un unico contenitore, sono considerati oggetti a se stanti e il loro utilizzo può essere differenziato.

Prima di discutere del problema dello scaricamento dei dati è bene analizzare il formato del documento bibliografico. Premesso che CDS/ISIS permette di definire fino a 999 campi identificati dai codici numerici da 1 a 999 detti tag, che ogni campo può essere istanziato fino a 255 volte e che ogni campo può contenere fino a un massimo di 32 sottocampi identificabili dal carattere "^" seguito da una lettera/numero: "^a" o "^c"...., a struttura del documento bibliografico è la seguente:

1. tag 001 : Internal identification come concatenazione univoca di collezione/authority/identification : per esempio
cnr.ieni/cnr.ieni.ge/2006-A0-021
2. tag 002 : Original Identifier che attenendosi al tag precedente è
/authority/identification: cnr.ieni.ge/2006-A0-021
3. tag 004 : Data di creazione nel seguente formato
^a<OriginalDateFormat>^byyyymmdd nell'esempio
^a26/1/2007^b20070126
4. tag 005 : Data di modifica nel seguente formato
^a<OriginalDateFormat>^byyyymmdd nell'esempio
^a26/1/2007^b20070126
5. tag 030 : titolo originale
6. tag 031 : titolo normalizzato dei caratteri diacritici
7. tag 050 : soggetti (free keywords)
8. tag 070 : Autori
9. tag 090 : Altri Autori/Autori
10. tag 120 : data di pubblicazione/anno di pubblicazione
11. tag 160 : Referenza Bibliografica
12. tag 161 : Note

13. tag 170 ISSN/ISBN nella seguente maniera:
 - ^aISBN
 - ^bISSN
14. tag 190 : Identificazione e collocazione per esempio: cnr.ieni.ge/2006-AO-021
15. tag 210 : classificazione ACM
16. tag 211 : classificazione MSC
17. tag 212 : classificazione ASFA
18. tag 213 : classificazione LC
19. tag 214 : classificazione CA
20. tag 215 : classificazione MESH
21. tag 216 : classificazione NAL
22. tag 217 : classificazione PACS
23. tag 218 : classificazione CNR
24. tag 250 : accesso al documento originale via internet (se è possibile)
25. tag 330 : abstract
26. tag 900 : codice dell'Istituto di appartenenza: estratto dall'authority
27. tag 904 : tipo di documento bibliografico: sempre "AP". Viene qui definito per compatibilità con gli altri dati bibliografici dell'ISTI e dell' Area di Ricerca di Pisa
28. tag 910 : tipologia del documento nel contesto MetaPub
29. tag 911 : tipologia del documento nel contesto ministro CNR
30. tag 990 : codice e nome della commessa CNR
31. tag 991 : classificazione della disciplina di appartenenza
32. tag 998 : BigField. Campo che contiene. identificati dal codice di

sottocampo. i campi componenti la Referenza Bibliografica. In questa maniera è possibile costruire i dati del minisito CNR con notevole precisione.

1. ^a : Journal Title (<dc_relation.journal>)
2. ^b : Volume Number (<dc:relation.volume>)
3. ^c : Issue Identifier (<dc:relation.issue>)
4. ^d : Numero di pagine (<dc:source.pages>)
5. ^e : Social Issue Identifier (<dc:relation.specialissue>)
6. ^f : Contributors (<dc:contributors>)
7. ^g : Publisher (<dc:publisher>)
8. ^h : Date of Publication (<dc:date.publication>)
9. ^i : Book Title (<dc:relation.book>)
10. ^j : Edition (<dc:relation.edition>)
11. ^k : Series Title(<dc:relation.series>)
12. ^l : Series Contributors (<dc:contributors.series>)
13. ^m : Publication Place (<dc:relation.place>)
14. ^n : Conference Title(<dc:relation.conference>)
15. ^o : Date of Event (<dc:relation.eventdate>)
16. ^p : Volume of Conference (<dc:source.volume>)
17. ^q : Project (<dc:relation.project>)
18. ^r : Various Title (<dc:relation>)
19. ^s : Date Available (<dc:date.available>)
20. ^t : Event (<dc:relation.event>)
21. ^u : Journal Code (<dc:relation.journal.code>)
22. ^v : CIVR Code (<dc:relation.civr>)
23. ^w : Affiliation (<dc:affiliation>)

Si confronti la struttura del documento bibliografico qui descritto con quello più generale del Metaopac del Catalogo della della biblioteca dell' Area di ricerca di Pisa in " Il Metaopac del CNR di Pisa e la sua collocazione nel Metaopac pisano" - ISTI-B4-14, 2004 pag. 9-12.

In maniera molto elementare si potrebbero scaricare tutti i dati appartenenti a un'unica authority eseguendo la seguente procedura che produce in out un file XML well formed:

```
/home/metaopac/mpisa/BiremePrograms/bin/wwwisis  
in=/home/metaopac/metapubCNR/ISTI/test1.query
```

dove test1.query è:

```
db=cnr-pub  
bool=( ISTI/(900) and 2006/(120) )  
h1=1
```

```
h2=8192
pft=@bicexml_ns.xmlpft
cipar=/export/home/metaopac/mpisa/cnr_ienigeConf/cnr.txt
out=./test1.out
```

e il file ottenibile è:

```
<?xml version="1.0" encoding="iso-8859-1"?>
<BibliographicList>
<BibliographicRecord>
<Id>cnr.isti/2006-A2-16</Id>
<title>Efficient mining of temporally annotated sequences</title>
<AuthorList>
  <Author>Giannotti F.</Author>
  <Author>Nanni M.</Author>
  <Author>Pedreschi D.</Author>
</AuthorList>
<SubjectList>
  <Subject>Temporal Data Mining</Subject>
  <Subject>Sequential Pattern</Subject>
</SubjectList>
<YearOfPublication>2006</YearOfPublication>
<acmClass>
  <Classification>I.2.6 Learning</Classification>
</acmClass>
<Abstract>Sequential patterns mining received much attention.....
.....
.....experiments performed on synthetic datasets.</Abstract>
<Notes>In: 2006 SIAM Conference on Data Mining (Bethesda, .....
.....Skillicorn and Jaideep Srivastava (eds.). SIAM, 2006.</Notes>
<InternetAccess>
  "http://www.siam.org/meetings/sdm06/proceedings/032giannottif.pdf"
</InternetAccess>
</BibliographicRecord>
<BibliographicRecord>
.....
.....
<BibliographicList>
```

Adattando il formato di stampa e la query stessa è possibile ottenere svariati modelli di risultato.

L'esportazione dei dati verso il database del minisito del CNR

Un altro modo di esportare i dati è quello di produrre file sequenziali contenenti i dati delle pubblicazioni di ogni Istituto del CNR partecipante a MetaPub nel formato caricabile nel minisito del CNR.

Nella directory /home/metaopac/MetapubCNR di metabrv.isti.cnr.it sono contenuti:

1. I programmi di conversione dei dati (Directory JavaPGM)
2. I file che descrivono le discipline e le tipologie delle Pubblicazioni
3. Le directory degli Istituti

In ogni directory di Istituto sono contenuti:

1. I link ai file generali delle discipline e delle tipologie (se omessi verranno ignorati)
2. Il file dei nomi dei dipendenti con le relative matricole: matricole.cnr (se omesso verrà ignorato)
3. Il file dei nomi dei dipendenti **non strutturati** (borsisti, dottorandi, contrattisti etc.) con le relative matricole: matricole.no_cnr (se omesso verrà ignorato)
4. I file delle query (*.query) che permettono di selezionare il database dell'Istituto a seconda dei tipi di Pubblicazione che si vogliono/devono accorpate per produrre i dati che tramite i programmi di conversione costruiranno i file caricabili sul minisito
5. La procedura getPubFromIsis che manda in esecuzione lo scaricamento dal database MetaPub dei dati utilizzando le specifiche descritte nei file *.query:

```
../wwwisis in=jcr.query  
../wwwisis in=nojcr.query  
../wwwisis in=bv.query  
../wwwisis in=a1b1ed.query  
../wwwisis in=a2b2.query  
../wwwisis in=a3b3a6b6a7b7.query  
../wwwisis in=b4b5ecppprtrmaiswp.query
```

I dati prodotti verranno scaricati in file aventi gli stessi nomi dei file *.query e come suffisso ".out").

E' abbastanza evidente il modo di aggiungere procedure nuove di scaricamento dei dati.

Ogni "query" è specializzata a scaricare particolari tipi di pubblicazioni. Il file b4b5ecppprtrmaiswp.query che contiene:

```
db=cnr-pub  
bool=( b4/(190) or b5/(190) or ec/(190) or pp/(190) or pr/(190) or ma/(190) or  
tr/(190) or is/(190) or wp/(190) or ed/(190)) AND ( 2006/(120) ) AND  
( IСТИ/(900) ) AND ( AP/(904) )
```

```
h1=1
h2=8192
freq=@cnrSSyeti.pft
prolog='#Prolog : ', 'Anno 2006: rapporti',
epilog=@epilog.cnr-pub
pft=@cnr-outpub.pft
cipar=/export/home/metaopac/mpisa/cnr_ienigeConf/cnr.txt
out=./b4b5ecppprtrmaiswp.out
```

scarica tutte le pubblicazioni dell'ISTI per l'anno 2006 che sono di tipo b4, b5, ec, pp, pr, ma, tr, is, wp, ed

6. La procedura gencnrpub produce una directory dal nome del tipo out.yyyy-mm-dd.hh:mm:ss (out.2007-03-22.15:27:43) che contiene i file dati caricabili sul minisito.

```
date
rm *.logfile
p=out.`date +%Y-%m-%d.%H:%M:%S`
echo $p
mkdir $p
java -classpath ../JavaPGM cnrpub list.file $p trace 074
date
```

Il programma cnrpub ha in input 4 parametri

1. list.file che contiene a lista dei file da elaborare e la loro tipologia locale

```
a1b1ed.out libri
a2b2.out pubblicazioni
a3b3a6b6a7b7.out abstract
b4b5ecppprtrmaiswp.out rapporti
jcr.out jcr
nojcr.out nojcr
bv.out brevetti
```

2. Il nome della directory di out
3. trace/notrace se si vogliono tracciare le operazioni sul logfile
4. CDS o Centro Di Spesa che in realtà corrisponde a un codice di tre numeri per identificare l'Istituto.

Come scaricare i dati di un nuovo/vecchio Istituto

Lo scaricamento dei dati di un nuovo Istituto implica la sua definizione/esistenza all'interno del sistema e la selezione dei dati appartenenti all'Istituto stesso specificando il valore che il codice di Istituto deve assumere nel tag 900.

Nel caso di nuovo istituto eseguire i punti 1.-4. , mentre i successivi sono applicabili anche agli Istituti già definiti.

1. Definire una nuova directory, con il codice dell'Istituto, nella directory del sistema /home/metaopac/metapubCNRpub e posizionarsi al suo interno
2. Copiare dalla directory di un Istituto già esistente:
 1. la procedura getPubFromISIS
 2. la procedura gencnrpub e assegnare il CDS giusto all'Istituto.
 3. il file list.file che contiene i nomi dei file di out prodotti da getPubFromIsis e la loro tipologia.
 4. I file query che dovranno essere aggiornati per quello che riguarda il tag 900 (Se il nuovo codice Istituto è IIPPR allora deve essere IIPPR/(900)), e eventualmente per differenti accorpamenti di tipologie di pubblicazioni (a0, b0,)
3. Eseguire i due link per rendere operativi i file discipline.cnr e tipologie.cnr:
 1. In -s ../discipline.cnr discipline.cnr
 2. In -s ../tipologie.cnr tipologie.cnrI due file possono anche non esistere o essere locali all'Istituto.
4. Ottenere dall'Istituto il file dei nomi e delle matricole dei dipendenti e caricarli sulla relativa directory come matricole.cnr
5. Ottenere dall'Istituto il file dei nomi dei dipendenti **non strutturati** (borsisti, dottorandi, contrattisti etc.) con le relative matricole: matricole.no_cnr
6. Eseguire la procedura getPubFromIsis e verificarne la produzione dei file di out. Attenzione: se non viene selezionato nessun documento il file di out non è vuoto ma contiene qualcosa di simile:

```
#Prolog : Anno 2006: brevetti
<error-msg>
0 record retrieved
</error-msg>
```
7. Eseguire la procedura gencnrpub.
8. Eseguire un tar -czf della directory di out.

Il recovery del sistema

Il sistema Metapub è installato sui seguenti 3 computer:

1. **metabsrv.isti.cnr.it** - come server per l'aggiornamento dei dati, produzione dei file di recovery e accesso all'informazione
2. **dell11.isti.cnr.it** - come server per l'accesso all'informazione
3. **orione.isti.cnr.it** - come server per l'accesso all'informazione

Il server operativo per l'accesso all'informazione viene determinato dalla configurazione del proxy server. Verificare su `/etc/apache2/sites-available` sul proxy server `metabsrv.isti.cnr.it`.

Prescindendo dalla collocazione del proxy server potrebbe essere necessario spostare il server che opera l'aggiornamento dei dati da `metabsrv.isti.cnr.it` a uno degli altri 2 server. Questo potrebbe avvenire anche a fronte di un crash del proxy server e suo ripristino (vedi: *Il MetaOpac Pisano e gli altri componenti del Servizio : organizzazione, controllo e recovery del sistema per l'anno 2006 - ISTI-B4-17, 2006 - L'organizzazione del MOP e degli altri servizi pag 19-21*)

Tenendo conto che:

1. La procedura `/home/metaopac/taring_cnr_prj_vc` salva sulla directory `/home/metaopac/mpisa/web/MeataopacDirectorySaver/solaris9` il file **`cnr_prj_vc.tgz`** (vedi punto 11. della procedura **`loadCNRpubbs`**)
2. La procedura `/home/metaopac/taring_FromXMLtoIsis` salva sulla directory `/home/metaopac/mpisa/web/MeataopacDirectorySaver/solaris9` il file **`FromXMLtoIsis.tgz`** (vedi punto 2. della procedura **`loadCNRpubbs`**)
3. Manualmente viene salvata la directory `/home/metaopac/metapubCNRpub` tramite la procedura `/home/metaopac/taring_metapubCNRpub` in `/home/metaopac/mpisa/web/MeataopacDirectorySaver/solaris9` come file **`metapubCNRpub.tgz`** (vedi: ***L'esportazione dei dati verso il database del minisito del CNR***)
4. Manualmente il file `/home/metaopac/taring.taring` viene salvato in `/home/metaopac/mpisa/web/MeataopacDirectorySaver/solaris9` come `taring_taring.tgz` (esso contiene tutte le procedure di compressione utilizzate da `metabsrv.isti.cnr.it`)

E che su ogni server è presente la directory e i relativi file menzionati nei 4 punti precedenti. Sarà sufficiente copiare i quattro file:

1. cnr_prj_vc.tgz, su una directory temporanea e successivamente decomprimere il file
2. taring_taring.tgz su una directory temporanea e decomprimerlo
3. copiare FromXMLtoIsis.tgz sulla directory /home/metaopac e decomprimerlo
4. copiare metapubCNRpub.tgz sulla directory /home/metaopac e decomprimerlo

Gli ultimi 2 punti soddisfano il ripristino dei due software in quanto non esistono tali directory sul server.

Per il punto 1. è necessario procedere manualmente a:

1. copiare il file /<tmp>/mpisa/proc/get_loadCNRpubbs in /home/metaopac/mpisa/proc/get_loadCNRpubbs
2. rinominare la directory /home/metaopac/mpisa/IsisDbdir.cnr_ienige in /home/metaopac/mpisa/IsisDbdir.cnr_ienige_saved
3. copiare la directory /<tmp>/mpisa/IsisDbdir.cnr_ienige sotto la directory /home/metaopac/mpisa
4. editare la procedura /home/metaopac/mpisa/IsisDbdir.cnr_ienige/loadCNRpubbs e commentare i comandi scp e ssh relativi al server stesso (per impedire la copia su se stesso)

Per il punto 2. copiare le procedure

1. taring_cnr_prj_vc
2. taring_metapubCNRpub
3. taring_FromXMLtoIsis

nella directory /home/metaopac

Infine aggiornare il crontab di user(metaopac) per attivare giornalmente le procedure di aggiornamento e salvataggio del sistema (eseguire la procedura /home/metaopac/mpisa/proc get_loadCNRpubbs)

Bibliografia

1. Giuseppe A. Romano - Isis/GAS : un sistema multi piattaforma per l'accesso a basi di dati testuali via Internet - ISTI-B4-28, 2003 - URL: <http://leonardo.isti.cnr.it/metaopac/IsisGas.pdf>
2. Giuseppe A. Romano - Isis/GAS : manuale di installazione e d'uso - ISTI-B4-31, 2003 - URL: <http://leonardo.isti.cnr.it/metaopac/IsisGasInstall.pdf>
3. Giuseppe A. Romano - Il MetaOpac Pisano: l'organizzazione logica e il mapping fisico - ISTI-B4-02, 2003 - URL: <http://leonardo.isti.cnr.it/metaopac/MetaOpacLogic.pdf>
4. Giuseppe A. Romano - Il MetaOpac Pisano: descrizione tecnica e procedure di recovery - ISTI-B4-30, 2003 - URL: <http://leonardo.isti.cnr.it/metaopac/MetaopacPisano.pdf>
5. Giuseppe A. Romano - Il metaopac: controllo e recovery del sistema - ISTI-B4-07, 2004 - URL: http://leonardo.isti.cnr.it/metaopac/IsisGAS_BriefMan.pdf
6. Biagioni S., Giannini S., Giuseppe A. Romano - Il Metaopac del CNR di Pisa e la sua collocazione nel Metaopac pisano - ISTI-B4-14, 2004 - URL: <http://leonardo.isti.cnr.it/metaopac/OpacCNR.pdf>
7. Pierlugi Niccolai, Enrico Venturini - TECA: Applicazione di CDS/ISIS per i dati bibliografici - Titivillus , Firenze 1993
8. Giuseppe A. Romano - Il MetaOpac Pisano: organizzazione, controllo e recovery del sistema per l'anno 2005 - ISTI-B4-17, 2005 - URL: http://leonardo.isti.cnr.it/metaopac/MOP_BriefMan.pdf
9. Giuseppe A. Romano - ExtGAS: un sistema multipiattaforma per l'accesso ai cataloghi di sistemi biblioteconomici eterogenei - ISTI-B4-15, 2005 - URL: <http://leonardo.isti.cnr.it/metaopac/ExtGas.pdf>
10. Giuseppe A. Romano - ExtGAS: manuale di installazione e d'uso - ISTI-B4-16, 2005 - URL: <http://leonardo.isti.cnr.it/metaopac/ExtGasInstall.pdf>
11. Giuseppe A. Romano - ExtGas/IsisGAS revisione 2.0: manuale di installazione e uso" - ISTI-B4-16, 2006 - URL: <http://leonardo.isti.cnr.it/metaopac/ExtGASOmoDbInstall.pdf>
12. Giuseppe A. Romano - "Il MetaOpac Pisano e gli altri componenti del Servizio : organizzazione, controllo e recovery del sistema per l'anno 2006 " - ISTI-B4-17, 2006 - URL: http://leonardo.isti.cnr.it/metaopac/MOP_BriefMan_2006.pdf
13. Stefania Biagioni, Carlo Carlesi, Daniela Musa - Sistema Atlas:

catalogazione derivata da registrazioni in formato UNIMARC -
IEI-B4-30 : Istituto di Elaborazione della Informazione,
settembre 1994

14. CDS/ISIS - software, tools, descriptions - URL:
[http://portal.unesco.org/ci/ev.php?URL_ID=2071&URL_DO=DO_T
OPIC&URL_SECTION=201&reload=1082372136&PHPSESSID=746
2f4450cf4d320877976df1ae94292](http://portal.unesco.org/ci/ev.php?URL_ID=2071&URL_DO=DO_T
OPIC&URL_SECTION=201&reload=1082372136&PHPSESSID=746
2f4450cf4d320877976df1ae94292)
15. BIREME - CISIS reference manual - URL:
<http://productos.bvsalud.org/html/en/home.html>
16. BIREME - wwwisis version 3.0 - URL:
<http://productos.bvsalud.org/html/en/home.html>
17. LIBERO - software - URL:
<http://www.infologic.it/ita/libero.htm>
18. Antonio Scolari - Unimarc, Roma, AIB, 2000 Enciclopedia Tascabile
ETIS.