

## Interface Aleph500/Su-Doc

<b>Référence</b>	Interface Aleph500/Su-Doc – 1.2
<b>Date de révision du document</b>	07/02/2003
<b>Version du document</b>	1.2
<b>Etat</b>	Diffusable
<b>Auteur</b>	Yves Grange (Responsable développement)
<b>Diffusion</b>	Client

### **Objet du document**

Ce document décrit le fonctionnement de l'interface entre Aleph500 et l'Abes. Il permet au correspondant informatique Aleph500 ou au responsable des échanges Su-Doc (ayant préalablement reçu une formation système Aleph500) de paramétrer, contrôler et superviser l'ensemble du processus. Pour plus de détails concernant les échanges proprement dit, se référer au document de l'Abes : «Dossier décrivant les spécifications pour les échanges de données entre le Su-Doc et les SIGB ».



## Sommaire

<b>1. PREALABLE</b>	<b>4</b>
<b>2. MODE DE FONCTIONNEMENT DES TRANSFERTS REGULIERS</b>	<b>5</b>
2.1 DEROULEMENT COMPLET D'UN CYCLE	5
2.2 PARAMETRAGE DES ECHANGES DE MAILS	6
2.3 DEROULEMENT DE L'INTEGRATION PROPREMENT DITE	6
2.4 SAUVEGARDES DE CHAMPS LOCAUX	6
2.5 SUPPRESSION DE CHAMPS SU-DOC	6
2.6 CONSTRUCTION DE L'INDEX IDN	6
<b>3. ANALYSE DES LOGS ET ERREURS</b>	<b>6</b>
3.1 CONTROLE DE ROUTINES (PARTIE BIBLIOTHECONOMIE)	6
3.2 GRAPHE DE DEDOUBLONNAGE DES NOTICES	6
3.3 ANALYSE EN CAS DE DYSFONCTIONNEMENT	6
<b>4. ANNEXE</b>	<b>6</b>
4.1 MAIL « MESSAGE DE FIN DE TRAITEMENT EN CENTRAL »	6
4.2 MAIL « GET TITLE DATA »	6
4.3 MAIL « GTD STATUS »	6

## 1. Préalable

Avant de mettre en place les transferts réguliers : les notices locales au SIGB doivent être mises à jour. Cette étape est appelée transfert initial : elle consiste à mettre à jour le champ 001 (numéro source) de vos notices à partir d'une liste de correspondances fournie par l'Abes.

La liste fournie par l'Abes à la structure suivante :

Numéro\_PPN : Numéro\_source\_actuel

```
PPN 000000779 : frBN00000208X
PPN 000000779 : sib0526023
PPN 000000809 : frBN000002179
PPN 000000809 : lm8214
PPN 000000809 : ocm08468009
PPN 000000809 : ocm38029748
PPN 000000809 : ocm38029786
PPN 000000809 : sib0183227
PPN 000000809 : sib0183233
PPN 000000817 : frBN000002186
PPN 000000817 : ocm05307070
PPN 000000817 : sib0728694
PPN 00000085X : frBN000002268
PPN 00000085X : ocm01294212
PPN 00000085X : sib0145233
PPN 000000892 : frBN000002332
PPN 000000892 : frBN000796412
```

Le «Numéro\_source\_actuel» est préfixé suivant qu'il provient de SIBIL (sib), de la BNF (frBN), d'OCLC (oclc), du CCN (cnn) ... **Il est impératif que les notices du SIGB soient préfixées de la sorte avant le transfert initial.**

### Description détaillée du transfert initial :

Les notices du SIGB sont exportées de la base et analysées une par une :

- Cas 1 : le champ 001 de la notice n'est pas préfixé (sib, frBN, oclc, cnn ...) et ne contient que des chiffres : le traitement ajoute un préfixe ALP au numéro source.
- Cas 2 : le champ 001 de la notice est préfixé (il contient au moins un caractère alphabétique) :
  - Si le numéro source se trouve dans la liste des correspondances PPN, alors le champ 001 est remplacé par le numéro PPN correspondant et l'ancien numéro source est placé en 035 préfixé par ALP.
  - Si le numéro source ne se trouve pas dans la liste des correspondances PPN, alors aucune modification n'est appliquée (la notice pourra être mise à jour ultérieurement par adjonction d'un numéro ALP en 035).

Les notices ainsi modifiées sont rechargées dans la base locale et ré indexées (seul une ré indexation de type « index directs » est nécessaire).

## 2. Mode de fonctionnement des transferts réguliers

Les transferts réguliers permettent de mettre à jour de façon automatique les catalogues de la bibliothèque (bibliographique et autorité) à partir d'un catalogue commun (Su-Doc catalogue), ils s'appuient sur deux outils : les mails et les ftp.

Les mails sont utilisés aux fins suivantes :

- avertir de la disponibilité de fichiers de notices sur le serveur de l'Abes,
- demander l'envoi par ftp des fichiers présents sur le serveur de l'Abes sur le serveur Aleph500,
- vérifier le déroulement correcte des transferts ftp.

Les transferts de fichiers (ftp) :

- se font uniquement du serveur Abes vers le serveur Aleph500,
- sont déclenchés à la demande par mail d'une personne (ou d'un serveur) accréditée par l'Abes,
- Sont déclenchés automatiquement par Aleph500

### 2.1 Déroulement complet d'un cycle

- De façon régulière, l'Abes crée des fichiers de notices modifiées ou ajoutées sur son serveur. Elle envoie ensuite un mail à l'utilisateur m505 lui indiquant la mise à disposition de ces fichiers.
- Un robot situé sur le serveur Aleph500 relève et analyse les messages reçus sur la boîte mail de l'utilisateur m505 (cf. Paramétrage des échanges de mail).
- Si le message reçu est du type «Message de fin de traitement en central » : le robot envoie automatiquement à l'Abes un mail «GET TITLE DATA » demandant le transfert des fichiers.
- Après réception du message «GET TITLE DATA », l'Abes envoie par ftp sur le serveur Aleph500 tous les fichiers en attente sur son propre serveur. A la fin de ces transferts, un nouveau mail «GTD STATUS » est envoyé à l'utilisateur m505 indiquant le résultat du transfert.
- Si le mail «GTD STATUS » indique que le transfert s'est déroulé correctement, les noms des fichiers à charger sont placés dans le fichier TO\_LOAD pour être traités.

## 2.2 Paramétrage des échanges de mails

Avant-propos : un client de messagerie classique peut être paramétré pour consulter la boîte mail de l'utilisateur m505. Il faut cependant toujours laisser l'option « garder une copie sur le serveur » activée et utiliser exclusivement cette boîte mail pour visualiser les messages.

Les lignes suivantes ont été ajoutées dans le job daemon d'Aleph500 (util e 5) :

```

=====
!
!           job_list
!
=====
! Note that this list can define either programs or procedures.
! If programs, col. 5 lists the program commands.
! If procedures, col. 5 lists the library code, and cols. 6 and 7
! the name parameters of the procedure.
!
! Column definition for PROGRAMS
! Col. 1 : day
!       2 : hour
!       3 : queue (y/n)
!       4 : log name OR target name (see following)
!       5 : program to run OR library code (see following)
!
! Formatting of parameters/log file name
!
! %D - day (00-06) [Sunday=00 Monday=01 ... Saturday=06]
! %H - hour (00-23)
! %M - minutes (00-59)
!
! xxxx,print_%D_%H_%M -> if run at 09:00 on Sunday xxxx,print_00_09_00
!
! %DATE[+/-][D/W/M]nnn - create date relative to the current date
!                       '+' - after current date '-' - before current date
!                       D - days W - weeks M - month
!
! xxxx_%DATE+D007      -> set parameter to 7 days from now
! yyyy_%DATE+M001     -> set parameter to 1 month from now
!
! Note that macros, or templates can be used based on definitions in UTIL E/5/5:
!
! W1 22:00:00 Y                USM50 p_cir_01                USM50
! 02 D1      Y                USM50 p_cir_01                USM50
!
! After editing this file restart jobd
!   2       3       4       5
!!-!!!!!!-!-!!!!!!-!!!!!!-!!!!!!-!!!!!!-!!!!!!>
01 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh
02 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh
03 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh
04 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh
05 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh
06 D2      N mail.log          /aleph/a50_5/aleph/proc/su-bot.csh

```

### Paramétrage du job\_list (util e 5 7)

```

=====
!
!           job_list.conf
!Scheduling of regular programs that run on the job daemon
!
=====
! Define weekly procedure slots:
!
! 1. Code
! 2. W(eekly)
! 3. Flags (Sun - Sat) Y/N
!
!1   2   3   4   5
!!----!-!!!!-!!!!-!!!!
D2   D 09:00 23:00 01:00

```

Paramétré comme ci-dessus, le robot (**su-bot.csh**) vérifie l'arrivé de nouveaux messages chaque jour de la semaine (sauf le dimanche) et chaque heure entre 9h et 23h. Attention : afin de ne pas générer d'erreurs, il est vivement conseillé de faire tourner le robot entre 9h et 23h, soit en dehors des heures d'intégration des notices du Su-Doc dans Aleph500.

Le fichier pop3.pl (consultation des messages de la boîte) doit être paramétré de la sorte :

```
#!/usr/bin/perl

use Mail::POP3Client;

$pop = new Mail::POP3Client( USER => "m505",
                             PASSWORD => "*****",
                             HOST => "*****",
                             PORT => "110" );
for ($i = 1; $i <= $pop->Count(); $i++)
{
    foreach ( $pop->HeadAndBody( $i ) )
    {
        print $_, "\n";
    }
    $pop->Delete( $i);
    print "_____END_OF_MESSAGE_____ \n";
}
}
```

ou :

**USER** : nom d'utilisateur de la boîte mail (normalement : m505)

**PASSWORD** : mot de passe de la messagerie.

**HOST** : nom du serveur pop (doit être accessible depuis le serveur Aleph500)

**PORT** : port du serveur pop (généralement 110)

Le fichier smtp.pl (envoi du message GET TITLE DATA) doit être paramétré comme ci-dessous :

```
#!/usr/bin/perl

use Mail::Sendmail;

%mail = ( To => 'abes_ftp@carmin.Su-Doc.abes.fr',
          From => 'm505@univ-paris3.fr',
          Subject => 'GET TITLE DATA'
        );

$mail{Smtp} = 'lendit.univ-paris3.fr';

$mail{Message} = "GTD_ILN=74\n";
$mail{Message} .= "GTD_YEAR=2002\n";
$mail{Message} .= "GTD_FILE_TO=195.221.79.2\n";
$mail{Message} .= "GTD_ORDER=TR124*\n";
$mail{Message} .= "GTD_REMOTE_DIR=/home/sud\n";

sendmail(%mail) or die $Mail::Sendmail::error;

print "OK. Log says:\n", $Mail::Sendmail::log;
```

ou :

TO : destinataire du message

FROM : expéditeur du message

SUBJECT : GET TITLE DATA

\$mail{smtp} : Nom du serveur smtp (serveur de mail) par lequel le message est envoyé.

\$mail{message} : corps du message, chaque ligne doit être terminée par '\n' (qui signifie retour à la ligne).

- GTD\_ILN = numéro ILN du site
- GTD\_YEAR = année des fichiers demandés
- GTD\_FILE\_TO = Adresse ip du serveur Aleph500
- GTD\_ORDER = numéro de traitement attribué par l'Abes (toujours le même)
- GTD\_REMOTE\_DIR = emplacement de destination des fichiers Su-Doc (doit être accessible par ftp pour l'Abes)

## 2.3 Déroulement de l'intégration proprement dite

Le fichier TO\_LOAD contient les noms des fichiers à intégrer dans le SIGB. Le robot (su-bot.csh) ajoute un nom de fichier dans TO\_LOAD si l'Abes envoie un message GTD STATUS 0 (signifiant que le fichier est bien arrivé sur le serveur Aleph500). Le robot trie ensuite ces noms de fichiers par ordre croissant du numéro de traitement attribué par l'Abes

Toute modification manuelle du contenu du fichier TO\_LOAD doit être signalée à Ex Libris.

Le déclenchement de la procédure d'intégration des notices du Su-doc se fait par l'intermédiaire du job daemon : (util e 5):

```
!=====  
!          job_list  
!=====  
! Note that this list can define either programs or procedures.  
! If programs, col. 5 lists the program commands.  
! If procedures, col. 5 lists the library code, and cols. 6 and 7  
! the name parameters of the procedure.  
!  
! Column definition for PROGRAMS  
! Col. 1 : day  
!       2 : hour  
!       3 : queue (y/n)  
!       4 : log name OR target name (see following)  
!       5 : program to run OR library code (see following)  
!  
! Formatting of parameters/log file name  
!  
! %D - day (00-06) [Sunday=00 Monday=01 ... Saturday=06]  
! %H - hour (00-23)  
! %M - minutes (00-59)  
!  
! xxxx,print_%D_%H_%M -> if run at 09:00 on Sunday xxxx,print_00_09_00  
!  
! %DATE[+/-][D/W/M]nnn - create date relative to the current date  
!                       '+' - after current date '-' - before current date
```



```
!                               D - days W - weeks M - month
!
! xxxx_%DATE+D007             -> set parameter to 7 days from now
! yyyy_%DATE+M001            -> set parameter to 1 month from now
!
! Note that macros, or templates can be used based on definitions in UTIL E/5/5:
!
! W1 22:00:00 Y                USM50 p_cir_01                USM50
! 02 D1      Y                USM50 p_cir_01                USM50
!
! After editing this file restart jobd
!
!      2      3      4      5
!!-!!!!!!!!-!-!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!>
s1 01:30:00 N log_su_doc      lance_su_doc.csh
```

La procédure « lance\_su\_doc.sh » commence par récupérer les informations contenues dans le fichier TO\_LOAD. Elle vérifie ensuite que le premier fichier de la liste correspond au numéro (ou numéro + 1) du dernier fichier intégré dans Aleph500. Le numéro du dernier fichier intégré est enregistré dans le répertoire « bin » du Su-Doc sous forme de nom de fichier (ex. TR134R24A001.RAW).

Ensuite la taille du fichier à intégrer va être analysée : si le fichier est vide, le traitement passe directement au fichier suivant (il arrive que l'Abes envoie des fichiers vides si aucune modification n'a été apporté au catalogue).

L'étape suivante va vérifier la nature du fichier :

- Le nom du fichier contient un A ou un B : notices bibliographiques
- Le nom du fichier contient un C : notices autorités

Une fois le type de fichier déterminé, le traitement contrôle l'état de la base :

- La base est déjà verrouillée, le traitement est mis en pause jusqu'à ce que la base soit de nouveau accessible,
- La base est accessible. Le traitement s'exécute après en avoir en verrouillé l'accès afin d'éviter toute modification accidentelle durant l'intégration du fichier

L'étape suivante s'assure que la connexion à la base est possible puis crée un répertoire pour le stockage des informations relatives au chargement (fichier d'origine, logs ...) (sous la forme : NomDeLaBase\_TypeDeFichier\_AAAAMMJJ\_hhmmss)

Exemple :

SCD01\_A\_20030124\_031259/

SCD10\_C\_20030124\_035512/

....

Les étapes suivantes de la procédure sont :

- le passage en Unicode des notices,
- le remplacement des indicateurs d'article de l'Abes par ceux d'Aleph500,
- la destruction des sous-champs \$5 du 035,
- la destruction des champs Su-Doc non désirés,
- la création des liens (le sous-champ \$0 pour les champs 4XX est remplacé par le sous-champ \$1 (contenant la chaîne «001PPN» suivi de la valeur initiale du sous-champ \$0),
- la mention «PPN» est ajoutée en début des sous-champs \$3,
- les espaces du champ 105 sont remplacés par des tirets,
- Filtre CCN : si la notice contient exclusivement des champs 035 commençant par ISSN ou CCN, la notice ne sera ni ajoutée ni mise à jour dans Aleph500.

Une fois ces étapes passées, le programme prend les notices une à une et les répartit dans les 3 fichiers suivants (cf. « GRAPHE DE DEDOUBLONNAGE DES NOTICES ») :

- fichier des mises à jour (notices déjà présentes dans Aleph500) : modif.seq,
- fichier des ajouts (notices ajoutées au catalogue Aleph500) : ajout.seq,
- fichier des incohérences (notices non intégrées dans Aleph500) : inco.seq.

**Le programme s'appuie sur un index direct (IDN) pour établir les correspondances entre les numéros de notices Abes et Aleph500. Il est donc impératif de ne pas modifier cet index.** Si aucun index direct IDN n'existe dans la base, le traitement est aussitôt stoppé (cf. point 2.6).

Deux fichiers de type log sont générés au cours de cette exécution : debug.txt & debug\_fin.txt (cf. Contrôle de routines).

Dès que les trois fichiers de notices sont constitués, les notices de la base Aleph500 correspondantes à celles décrites dans le fichier modif.seq sont exportées vers un fichier plat. Celui-ci sera ensuite exploité afin de récupérer les données locales au SIGB suivant le paramétrage du fichier chp\_local.ini décrit ci-dessous.

Les notices des fichiers modif.seq et ajout.seq sont ensuite chargées dans le catalogue local puis ré-indexées. Selon la configuration en place (serveur, volumétrie, ...), cette étape peut prendre plusieurs heures car cette opération doit être effectuée en tâche de fonds (ue\_01). Chaque minute, le traitement global contrôle le nombre de notices restant à indexer.

Une fois l'indexation terminée, la base objet du traitement est déverrouillée, les logs sont mis à disposition dans le répertoire « print » et le nom du fichier traité est indiqué dans le répertoire « bin » (en lieu et place de celui figurant précédemment).

## 2.4 Sauvegardes de champs locaux

Il est possible de sauvegarder des champs locaux du SIGB. Le fichier de paramétrage se nomme : `chp_local.ini`.

```
! fichier : chp_local.ini
!
!
!   col. 1 : contient toujours ^[0-9]\{9\}
!
!   col. 2 : étiquette du champ à conserver + indicateur
!             le caractère « . » (point) signifie n'importe quel caractère
!             exemple : 7..10 signifie tous les champs 700 avec indicateurs 10
!
!   col. 3 : jeux de caractères : L
!
!   col. 4 : optionnel : ajoute une condition de sélection
!
!
!Exemple :
!
!^[0-9]\{9\} 606.. L .*$\$2rameau
!^[0-9]\{9\} 606.. L .*$\$2lc
!^[0-9]\{9\} 3.... L
!
! Signifie que les champs locaux 606 (quelque soient les indicateurs)
! contenant un $$2rameau ou un $$2lc seront repris des notices locales.
! Les champs 3XX seront repris des notices locales.
!
!
!      1          2      3                                4
!!!!!!!!!!!!!!-!!!!!!-!-!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
^[0-9]\{9\} 606.. L .*$\$2rameau
^[0-9]\{9\} 606.. L .*$\$2lc
^[0-9]\{9\} 3.... L
```

le point « . » : signifie n'importe quel caractère, y compris l'espace.

L'étoile « \* » : caractère de répétition, agit sur le caractère ou le caractère spécial précédent.

Exemple : .\* : signifie n'importe quel chaîne de caractères.

Le backslash « \ » : fait perdre sa signification spéciale au caractère suivant.

## 2.5 Suppression de champs Su-Doc

Il est possible de supprimer des champs provenant du Su-Doc. Le fichier de paramétrage se nomme : chp\_supp.ini

```
! fichier : chp_supp.ini
!
!
! col. 1 : contient toujours ^[0-9]\{9\}
!
! col. 2 : étiquette du champ à supprimer + indicateurs
!           le caractère « . » (point) signifie n'importe quel caractère
!           exemple : 7..10 signifie tous les champs 700 avec indicateurs 10
!
! col. 3 : jeux de caractères : L
!
! col. 4 : optionnel
!           ajoute une condition de sélection
!
!
!Exemple :
!
!^[0-9]\{9\} 999.. L
!^[0-9]\{9\} 606.. L .*\$\$2lc
!
! Signifie que les champs Su-Doc 606 contenant un $$2lc ne seront pas repris.
! Les champs Su-Doc 999 ne seront pas repris
!
!
!           1           2   3                               4
!!!!!!!!!!!!!!-!!!!!!-!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
^[0-9]\{9\} 999.. L
```

## 2.6 Construction de l'index IDN

L'index IDN est primordial pour le bon fonctionnement du dédoublement des notices. Il est impératif qu'il soit construit de la sorte :

- Dans tab11 :

```

=====
!
!      tab11 - Assignment of MARC fields to indexes
!
=====
! Last header change: 03/10/99
!
!-----
! B) THE FOLLOWING POSITIONS ARE FOR NON-WORD INDEXES (HEADINGS, "DIRECT" INDEX
! AND LINK BACK FROM AUT)
! col. 2   : A (headings)/I (Direct index)/R (link back from AUT)
! col. 3   : document record field code. # can be used for the third to
!           fifth positions to indicate truncation of numeric additions to
!           the field code (eg. 245## for 2451, 2452, 24501).
! col. 4&5 : subfields and subfield contents to filter for (ex. if $$2 usm50
!           were in this column only records with usm50 in subfield 2
!           would have this field indexed). !! NOTE!! Filter text must be
!           entered in lower case.
! col. 6   : Index code
! col. 7   : Subfield(s) of the field (blank indicates entire field)
! col. 8   : Group. Used only for type A (headings). Values are: A, B or C.
!           This column may be used to define a particular group of
!           heading indexes for updating. This option is only available
!           when the program is run from the prompt command line. It is not
!           available from the Web Services.
! col. 9   : Non-filing indicator for non_filing procedure in tab_filing
! col. 10  : Not in use
!-----
!B)2  3  4   5           6           7           8 9 10
!!-!-!!!!!!-!-!!!!!!-!!!!!!-!!!!!!-!!!!!!-!!!!!!-!!!!!!-!!!!!!-!!!!!!

```

11 I 001## IDN

- Dans tab00.fre et tab00.eng :

```

=====
!
!      tab00.eng
!Field codes and names, for indexes (WRD, ACC & IND)
!
=====
! Last header change:27/07/1999
!
! TABLE_KEY 2,5,10,11
!
! HELP
!
! ALEPH allows various definitions of access paths (indexes) to document
! records, to suit the needs of each application. Indexes can be based
! on specific fields, a combination of fields, and individual words from
! specific document fields. For example, the Titles table can contain
! titles, additional titles, uniform titles, etc. The Authors table may
! contain authors, co-authors, additional authors, translators, etc.
! to suit the needs of each application. Indexes can be based on specific
! fields, a combination of fields, and individual words from specific
! document fields.
! For example, the Titles table can contain titles, additional titles,
! uniform title translators, etc
!
! Users must define tab00 lines for the general Word group (W-001)
! and for each s pecific Word group (W-nnn).
!
! Sort file
! In addition to definitions for indexing, tab00 also includes
! definitions of filling procedures for sort keys (z101 table of sort keys.)

```

```

!
! END_HELP
!
!
! COL 1. 1; ALPHA {H}; ;
!           H;
!           H;
! COL 2. 5; ALPHA_NUM, UPPER; ;
!           Heading code;
!           Code of heading/word/index file (for word
!           files, must start counting from W-001 and
!           should not jump large range of numbers);
! COL 3. 5; ALPHA_NUM, UPPER; -;
!           Index type;
!           Index name;
! COL 4. 1; NUM {0,1,2, }; ;
!           Suppress end punctuation ;
!           0 - no suppression
!           1 or space - suppress all end punctuation
! COL 5. 2; NUM; ;
!           Procedure;
!           Filing procedure;
! COL 6. 2; ALPHA_NUM; ;
!           00 (not implemented);
!           00 (not implemented);
! COL 7. 1; ALPHA_NUM; ;
!           Not implemented;
!           Not implemented;
! COL 8. 1; ALPHA_NUM; ;
!           Not implemented;
!           Not implemented;
! COL 9. 4; NUM; ;
!           Location;
!           Location weight (last 2 digits in the W section
!           are for weighting as used in the Rank command
!           of the WWW OPAC);
! COL 10. 5; ALPHA_NUM, UPPER; ;
!           Heading code;
!           Code of heading/word/index file;
! COL 11. 20; TEXT; ;
!           Heading Name;
!           Name of heading/word/index file;
!
! 2      3      4 5 6 7 8 9      10      11
!-!!!!!!-!!!!!!-!-!!-!!-!!-!-!!!!-!!!!!!-!!!!!!!!!!!!!!!!!!!!!!
H IDN   IND      22 00      0000 IDN   No source

```

- Dans tab\_filing :

```

!=====
!           tab_filing - definition of filing routines for headings
!=====
! Last header change: 09/08/2000
!
! HELP
! This table is used for defining filing procedures that can be
! used when building the filing key for Headings (Z01), Index
! (Z11) entries, and Sort keys (Z101). The filing procedure
! identifier for these is set in col.5 of tab00. The filing
! procedure identifier is also used in col.3 of tab01 for
! the WEB Services b-ret-21 (Sort Catalog Records) and
! b-ret-01 (Retrieve Catalog Records) as well as for FILTER
! in the OPAC (i.e. checking records in a set for presence
! of particular text). If no procedure is listed in col.3 of
! tab01, the system defaults to procedure 99.
! Therefore, this tab_filing table should always include
! procedure 99, with at least "to_lower", "pack_spaces" and
! "del_subfield".
! The table is also used for the WEB Services Shelf Reading
! Reports (b-item-04, -05 and -06), using the location type.
! .
! This table is limited to 500 lines.
! .

```

```

! A single filing procedure is made up of a number of individual
! procedures, which are explained following.
! Note that the order in which the procedures is listed is important.
! For example, "numbers" (which compresses comma or dot in numbers
! to change 100,000 and 100.000 to 100000) must come before listing
! comma or dot in "change-to-blank".
! .
! List of filing procedures:
! - abbreviation: compress a dot between single characters (e.g. I. B. M.
! changes to I B M, I.B.M. changes to IBM)
! - bbk: special procedure for Russian filing standards, in which sorting
! sequence is special characters, followed by Cyrillic
! characters, followed by Latin characters, followed by numbers.
! - build_filing_key_lc_call_no: special procedure for correct sequencing
! of LC call numbers
! - char_conv: perform character conversion procedure according
! to the procedure name listed in col.4. This name must
! match procedure identification in col.1 of
! /alephe/unicode/tab_character_conversion_line
! - compress: compress (i.e. strip) the characters listed in col.4
! - compress_blank: delete blanks
! - del_subfield: delete subfield sign ($$x)
! - expand_num: expand number (add leading zeroes to fill numbers
! to 7 digits, for numeric filing)
! - mc_to_mac: change initial mc to mac
! - non_filing: drop initial text using non-filing indicator
! - non_numeric: delete non-numeric characters
! - numbers: compress a comma and a dot between numbers (e.g. 2,153
! changes to 2153)
! - suppress: suppress using <<...>> (drop all text contained
! within the signs, and the signs themselves)
! - to_blank: change characters listed in col.4 to blank
! - to_lower: change case to lower
! - to_carat: change subfield sign to ^^ (for hierarchical sorting
! of headings)
! END_HELP
! COL 1. 2; NUM; ;
! ID;
! Two-digit identifier of the filing routine. This identifier
! is used in column 5 of tab00 and column 3 of tab01;
! COL 2. 1; ALPHA {L,H,A,R,S} UPPER; #;
! ALPHA;
! ALPHA (character set indicator) of the text;
! COL 3. 20; ALPHA_NUM; ;
! Procedure;
! Name of the filing procedure;
! COL 4. 20; ALPHA_NUM; ;
! Parameters;
! Parameters for the filing procedure (when relevant);
! 1 2 3 4
!!-!-!!!!!!!!!!!!!!!!!!!!!!!!!!!!-!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!
!rem Numero avec prefixe
22 # del_subfield
22 # compress -
22 # compress_blank

```

Pour vérifier que l'index IDN est construit correctement, vous pouvez visualiser son contenu par la commande suivante :

```

s+ nom_de_la_base
SELECT Z11_DOC_NUMBER, UPPER(Z11_TEXT) FROM Z11 WHERE SUBSTR(UPPER(Z11_REC_KEY),1,8) = 'IDN
PPN';

```

Si aucune réponse n'est obtenue, il y a un problème. Vérifiez le paramétrage !

### 3. Analyse des logs et erreurs

#### 3.1 Contrôle de routines (partie bibliothéconomie)

Tout traitement d'un fichier provenant de l'Abes est tracé sous forme de Log (dans le répertoire print alias dp de la base administrative). Ces traces sont consultables directement par le module Edition d'Aleph500. Les différents fichiers sont :

- Un fichier TR\*\*\*R\*\*\*001\_debug.txt : qui détaille les opérations sur les notices (ajouts, modifications).

```

Traitement démarré le 16/12/2002 à 03:21:38.           ← Heure de début de traitement
Connexion à la base m505 ...
Utilisateur m505 connecté.
Compteur Last-doc-number (Table z52) : 000278955      ← Numéro de la dernière notice de la base
Nombre de correspondances Aleph <=> PPN : 237550      ← Nombre de notices avec un n° PPN en 001
1
Traitement des notices...

====> Notice n° 000000001                             ← Lecture de la première notice du fichier
|-> 001 : PPN000033146 -->n° Aleph : 000051012       ← 2
|-> 035 : ALP?????????                               ← Pas de n° ALP en 035 dans cette notice
|-> 035 : PPN?????????                               ← Pas de n° PPN en 035 dans cette notice
-----
-->Cas 2 : Mise à jour de la notice. (000051012)     ← Type de l'opération sur cette notice

====> Notice n° 000000002                             ← Lecture de la deuxième notice du fichier
|-> 001 : PPN000045616 -->n° Aleph : 000058130
|-> 035 : ALP?????????
|-> 035 : PPN007599323 PPN035854405                 ← Plusieurs n° PPN en 035 pour cette notice
-----
-->Cas 2 : Mise à jour de la notice. (000058130)     ← Type de l'opération sur cette notice
Traitement terminé le 16/12/2002 à 03:22:29.

```

<sup>1</sup> Le nombre de notices avec un numéro PPN en 001 est calculé d'après l'index IDN : si ce nombre est inférieur à 1 : le traitement s'arrêtera en considérant que l'index IDN n'est pas valide.

<sup>2</sup> La notice lue contient un numéro PPN en 001 qui correspond à la notice numéro 0000051012 dans Aleph500

- Un fichier TR\*\*\*R\*\*\*001\_debug\_fin.txt : qui détaille les erreurs rencontrés lors du dé doublonnage des notices entre le fichier Su-Doc et la base Aleph500.

```

====> Notice n° 000000109
|-> 001 : PPN002674394 -->n° Aleph : ??????????
|-> 035 : ALP?????????
|-> 035 : PPN006515568(SYS000003816) PPN055957188(SYS000236171) PPN020128614
-----
**>Cas 5.2 : Mise à jour de la notice. (000003816)
Attention plusieurs notices Aleph correspondent.

====> Notice n° 000000221
|-> 001 : PPN004681002 -->n° Aleph : ??????????
|-> 035 : ALP?????????
|-> 035 : PPN008087040 PPN005145287 PPN007304625(SYS000056865)
PPN012720984(SYS000066889) PPN020776047
-----
**>Cas 5.2 : Mise à jour de la notice. (000056865)
Attention plusieurs notices Aleph correspondent.

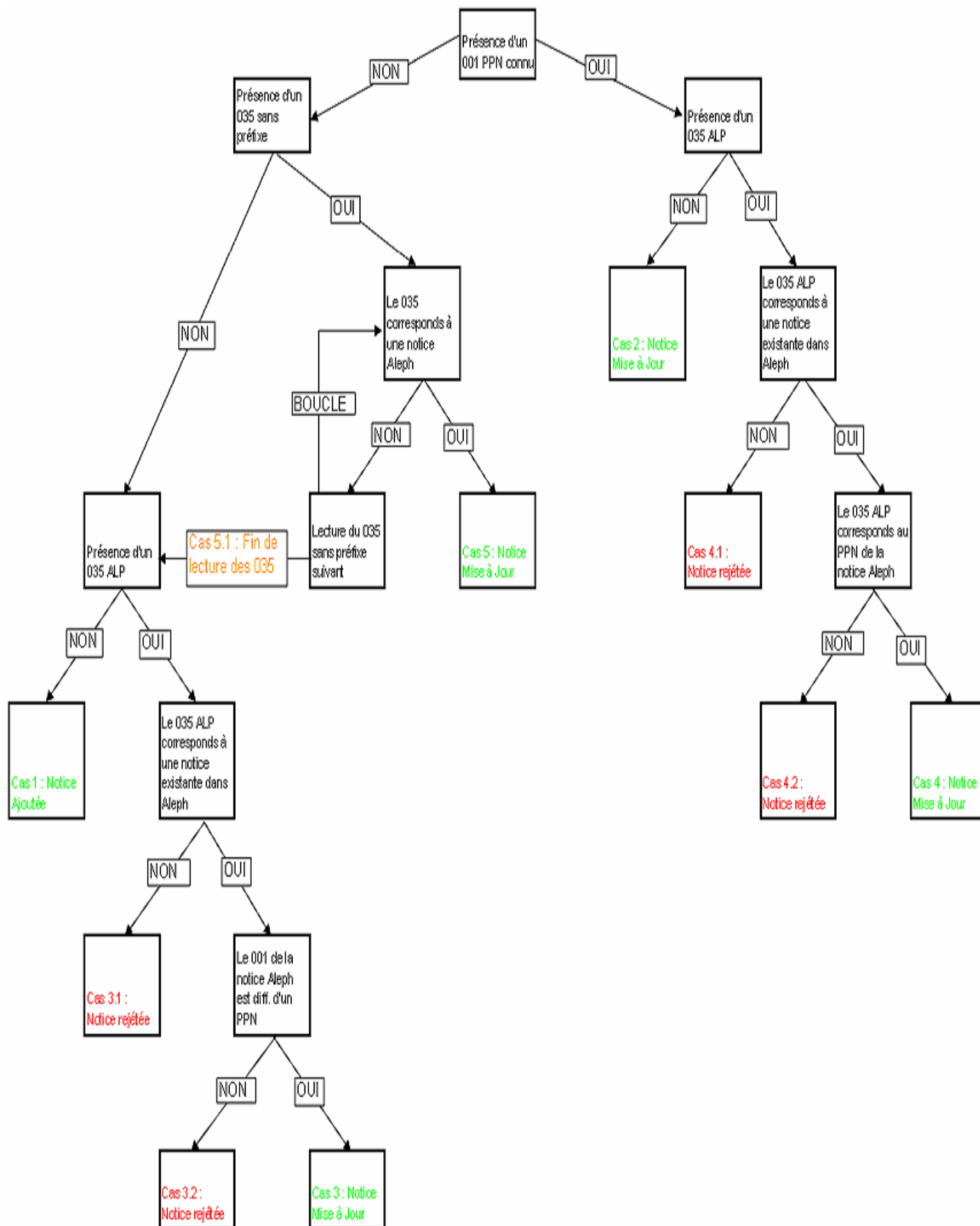
```



Les numéros SYS (voir entre les parenthèses) sont les numéros système Aleph500 correspondants aux numéros PPN.

**Les numéros de cas (type d'opérations) sont ceux présentés dans le graphe ci-dessous.**

### 3.2 Graphe de dédoublonnage des notices



### 3.3 Analyse en cas de dysfonctionnement

Emplacement des programmes : \$alephm\_source/sudyg/bin

Emplacement des données relatives aux traitements : \$alephm\_source/sudyg/data

Ce dernier répertoire contient :

- les originaux des fichiers envoyés par le Su-Doc,
- les différents fichiers intermédiaires,
- les fichiers résultant du traitement.

Emplacement de la log du traitement : \$alephe\_scratch/log\_su\_doc\_XXXXX

Emplacement des logs bibliothéconomiques : répertoire print de la base administrative.

#### Logs concernant les échanges de mail :

bin/full\_mail.box : liste de tous les mails reçus sur la boîte m505

bin/mail\_trt.log : idem que full\_mail.box + l'analyse des messages

En cas d'erreur dans les échanges de mails, cette log se termine par une série d'étoiles. Dans ce cas, il faut aller vérifier l'état du dernier mail transféré.

#### Exemple :

```
*****
Le sujet du message suivant doit être vérifié : le 20/12/02 à 13:00:08
*****
Return-Path: <abes_ftp@carmin.Su-Doc.abes.fr>
Received: from Su-Doc.abes.fr (carmin.Su-Doc.abes.fr [193.52.26.5])
        by lendit.univ-paris3.fr (8.9.3+Sun/jtpda-5.3.1) with ESMTTP id MAA27811
        for <m505@univ-paris3.fr>; Fri, 20 Dec 2002 12:19:35 +0100 (MET)
Received: (from abes_ftp@localhost)
        by Su-Doc.abes.fr (8.11.6+Sun/8.11.6) id gBKB1E412914;
        Fri, 20 Dec 2002 12:01:14 +0100 (MET)
Date: Fri, 20 Dec 2002 12:01:14 +0100 (MET)
Message-Id: <200212201101.gBKB1E412914@Su-Doc.abes.fr>
From: "GTD Service" <abes_ftp@carmin.Su-Doc.abes.fr>
To: m505@univ-paris3.fr
Subject: GTD status: 2 - Failure
MIME-Version: 1.0
Content-Type: text/plain; charset=iso-8859-1
Content-Transfer-Encoding: 8bit
X-UIDL: A~/!!~f$"!`~9"!DG#!
```

Dear customer,

Thank you for sending email to this address for requesting title data.  
The result of your request for this iln is shown in the subject of this  
email. For an explanation of the statuscodes, see below.

-----  
Request status (short):

Failure - 2 - Failure - unable to transfer the selected files

Transferred files information:

Filename	File size	Information
TR124R175A001.RAW	884267 bytes	Failure
425	Can't build data connection: Connection timed out.	
425	Can't build data connection: Host is unreachable.	
425	Can't build data connection: Connection timed out.	
125	- (warning) file transfer retry limit exceeded - file not transferred	
TR124R175B001.RAW	738512 bytes	Failure
425	Can't build data connection: Host is unreachable.	
425	Can't build data connection: Connection timed out.	
125	- (warning) file transfer retry limit exceeded - file not transferred	



```
Indexation en cours ... reste : 3111 notices
Indexation en cours ... reste : 3050 notices
Indexation en cours ... reste : 2982 notices
Indexation en cours ... reste : 2938 notices
Indexation en cours ... reste : 2824 notices
Indexation en cours ... reste : 2701 notices
Indexation en cours ... reste : 2576 notices
Indexation en cours ... reste : 2442 notices
Indexation en cours ... reste : 2314 notices
Indexation en cours ... reste : 2174 notices
Indexation en cours ... reste : 2032 notices
Indexation en cours ... reste : 1887 notices
Indexation en cours ... reste : 1747 notices
Indexation en cours ... reste : 1601 notices
Indexation en cours ... reste : 1456 notices
Indexation en cours ... reste : 1297 notices
Indexation en cours ... reste : 1144 notices
Indexation en cours ... reste : 1001 notices
Indexation en cours ... reste : 843 notices
Indexation en cours ... reste : 698 notices
Indexation en cours ... reste : 537 notices
Indexation en cours ... reste : 383 notices
Indexation en cours ... reste : 255 notices
Indexation en cours ... reste : 91 notices
Indexation en cours ... reste : 0 notices
Library SBN10 unlocked.
^[[2J^[[H^[[0m
ALEPH/ALPHA, Copyright Ex Libris.
version 505.14.1.5, 30-May-2001
/aleph/u50_5/sbn50/print/TR124R174C001_debug.txt
/aleph/u50_5/sbn50/print/TR124R174C001_debug_fin.txt
```

← Visuel sur le nombre de notices  
← restant à indexer <sup>5</sup>

← Déverrouillage de la base <sup>6</sup>

← Copie des logs

<sup>1</sup> Ce numéro provient de l'indicateur du répertoire « bin ». Par exemple, le fichier TR124R174B001.RAW dans le répertoire « bin » indique que le dernier fichier chargé est celui des notices « bibliographiques liées » numéro 174. Si plusieurs fichiers TR existent dans le répertoire « bin », le traitement s'arrête avec un message d'erreur. Il en va de même en cas d'absence de fichier. Si les numéros de traitement ne sont pas consécutifs, un message d'alerte sera affiché et le traitement stoppé.

<sup>2</sup> Les bases sont verrouillées par le traitement pour éviter des erreurs. Si la base est déjà verrouillée, le traitement attend que la base soit de nouveau accessible.

<sup>3</sup> La détection du type de notices se fait par le nom du fichier :

- A : bibliographique
- B : bibliographique lié
- C : autorité

<sup>4</sup> Si la connexion à la base est impossible, le traitement s'arrête.

<sup>5</sup> Si le nombre de notices ne diminue pas, il faut relancer manuellement la procédure d'indexation en tâche de fonds pour la base (util e 2 & util e 1).

<sup>6</sup> Tant que ce message ne figure pas dans le log, vous ne pouvez pas accéder à la base. Il ne faut surtout pas déverrouiller la base sans avoir préalablement remédié au problème ayant laissé la base dans cet état.

## 4. Annexe

### 4.1 Mail « Message de fin de traitement en central »

Return-Path: <pica@carmin.Su-Doc.abes.fr>  
Received: from Su-Doc.abes.fr (carmin.Su-Doc.abes.fr [193.52.26.5])  
by lendit.univ-paris3.fr (8.9.3+Sun/jtpda-5.3.1) with ESMTP id AAA21099  
for <m505@univ-paris3.fr>; Thu, 12 Dec 2002 00:38:31 +0100 (MET)  
Received: (from pica@localhost)  
by Su-Doc.abes.fr (8.11.6+Sun/8.11.6) id gBBNcTY18867  
for m505@univ-paris3.fr; Thu, 12 Dec 2002 00:38:30 +0100 (MET)  
Date: Thu, 12 Dec 2002 00:38:30 +0100 (MET)  
From: Pica Beheer <pica@carmin.Su-Doc.abes.fr>  
Message-Id: <200212112338.gBBNcTY18867@Su-Doc.abes.fr>  
To: m505@univ-paris3.fr  
Subject: For Jobld = 124 and Runld = 169, status is 9  
X-UIDL: ~g!!5&G"!H"p"!jNp!!

BibServ  
Jobld : 124  
Runld : 169  
Status : 9  
Comment : Job has been completed successfully

List of Selected titles  
DAT 20021212003615000  
RBF 670  
DSN /appl/tmp/iln74/TR124r169A001.raw  
ORS 1  
FOR M  
CSI iso5426

List of Linked titles  
DAT 20021212003617000  
RBF 688  
DSN /appl/tmp/iln74/TR124r169B001.raw  
ORS 1  
FOR M  
CSI iso5426

List of Linked Authorities  
DAT 20021212003622000  
RBF 2519  
DSN /appl/tmp/iln74/TR124r169C001.raw  
ORS 1  
FOR M  
CSI iso5426

### 4.2 Mail « GET TITLE DATA »

To : 'abes\_ftp@carmin.Su-Doc.abes.fr'  
From : 'm505@univ-paris3.fr'  
Subject : 'GET TITLE DATA'  
GTD\_ILN=74  
GTD\_YEAR=2001  
GTD\_FILE\_TO=194.206.252.70  
GTD\_ORDER=TR124\*  
GTD\_REMOTE\_DIR=/home/su

## 4.3 Mail « GTD Status »

Return-Path: <abes\_ftp@carmin.Su-Doc.abes.fr>  
Received: from Su-Doc.abes.fr (carmin.Su-Doc.abes.fr [193.52.26.5])  
by lendit.univ-paris3.fr (8.9.3+Sun/jtpda-5.3.1) with ESMTP id JAA23403  
for <m505@univ-paris3.fr>; Thu, 12 Dec 2002 09:01:02 +0100 (MET)  
Received: (from abes\_ftp@localhost)  
by Su-Doc.abes.fr (8.11.6+Sun/8.11.6) id gBC812706683;  
Thu, 12 Dec 2002 09:01:02 +0100 (MET)  
Date: Thu, 12 Dec 2002 09:01:02 +0100 (MET)  
Message-Id: <200212120801.gBC812706683@Su-Doc.abes.fr>  
From: "GTD Service" <abes\_ftp@carmin.Su-Doc.abes.fr>  
To: m505@univ-paris3.fr  
Subject: GTD status: 0 - Success  
MIME-Version: 1.0  
Content-Type: text/plain; charset=iso-8859-1  
Content-Transfer-Encoding: 8bit  
X-UIDL: C9!"!320!"!hpV"!eL^!!

Dear customer,

Thank you for sending email to this address for requesting title data.  
The result of your request for this iln is shown in the subject of this  
email. For an explanation of the statuscodes, see below.

-----  
Request status (short):  
Success - 0 - Normal successfull completion

Request information:  
Gtd\_Iln = 74  
Gtd\_Year = 2002  
Gtd\_File\_To = 195.221.79.2  
Gtd\_Remote\_Dir = /home/sud  
Gtd\_Order = TR124\*

Transferred files information:

Filename	File size	Information
TR124R169A001.RAW	1043002 bytes	Success
TR124R169B001.RAW	581347 bytes	Success
TR124R169C001.RAW	1794030 bytes	Success

-----  
Agence Bibliographique de l'Enseignement Supérieur  
Service Assistance du Système Universitaire de Documentation  
Email : svpsu@abes.fr  
T3511: 04 67 54 84 59  
-----

### CODE STATUS MEANING IN GENERAL

-----  
0 Success Normal successfull completion  
1 Partial Not all files were transfered successfully (see report)  
2 Failure Unable to transfer the selected files (see report)

Unexpected internal GTD Program errors (contact Abes about these errors)

101	GtdVar	GTD program error (variable not set)
102	GtdPath	GTD program error (path not found)
103	GtdFile	GTD program error (file not found)
104	GtdIni	GTD program error (INI-file invalid)
105	GtdMsg	GTD program error (MSG-file invalid)
106	GtdReqFile	GTD program error (no email request)
107	GtdMovErr	GTD program error (file move error)
108	GtdIniDef	Invalid definition for specified ILN in INI-file

Invalid GTD request errors

- 110 ReqInVHdr Email does not contain a valid header
- 111 ReqInvalid Email does not contain valid request keywords
- 112 ReqInVLN No or invalid ILN value specified - format is ###
- 113 ReqInVYr Invalid YEAR value specified - format is ## or #####
- 114 ReqNoILN No corresponding section found in GTD INI-file
- 115 ReqNotAuth The requestor is not authorized for the selected ILN

#### Connection & transfer errors

- 120 TfrNoHost No target-host defined or GTD\_FILE\_TO not specified
- 121 TfrNoOver Delete-after-transfer override requested (GTD\_DELETE) but not permitted (informational)
- 122 TfrCommErr Unable to setup a FTP session to the remote host;  
Possible reasons:
  - invalid host specified using GTD\_FILE\_TO
  - source and/or target hosts are down
  - user/password has been changed
  - FTP transfers are not allowed on source/target host
- 123 TfrListErr Unexpected error occurred while generating the list of files to transfer
- 124 TfrTransErr Unexpected error occurred during FTP file transfer
- 125 TfrLimExc File transfer retry limit exceeded (warning)

#### Information messages

- 130 NfoSrcSize Binary transfer file-size differ - retrying file transfer
- 131 NfoTfrSize Transferred file-sizes differ - retrying file transfer
- 132 NfoNoDel Unable to delete source-file from the source FTP-server
  
- 140 GtdMailOnly Option Mail\_Only was specified - no files transferred
- 141 NoFiles No files match the file selection mask on the source host