ALEPH v18 New Authority Control Setup and Testing

Fang Peng
&
David Bolotine

Stony Brook University

ELUNA 2008 conference at Long Beach California
July 2008

Contact:
Fang.Peng@StonyBrook.edu, David.Bolotine@StonyBrook.edu

Introduction

This document is based on Stony Brook University Library's experience in implementing Authority Control in Aleph 18. It is a result of an intense development and testing effort that we undertook in the summer and fall of 2007.

The effort centered around three full conversions of Stony Brook University library data and extensive tests by a skilled and very attentive team. The team included authority catalogers/librarians, reference librarians and the Web OPAC committee. It was led by me and David Bolotine, who is head of cataloging and is an expert on Library of Congress Authority handling.

David developed many test scenarios to fully vet the logic and to figure out how to best implement it.

As a result of this effort, Stony Brook University was the first Aleph customer to fully implement full Authority Control in a production environment. We have been using this logic since December 2007. The professional librarians at Stony Brook University are very happy with the improved features of the system and continually notice where it returns better results. We have received positive comments from our users.

We developed this guide as a road map to implementing the logic on production. We have shared copies with over 30 universities now and have improved the document in response to their feedback.

Aleph 18 New Authority Control Setup:

To implement Authority Control, Aleph has to be in version 18.01 with Service Pack 1117 or higher. SLGA18 in this document refer to Aleph Version 18.01 System Librarian's Guide – Authorities.

XXX01: The files need to be set up in the Bibliographic library

tab_acc_category : (xxx01)
This is a new file and can be copied from a_tree. Change the parameters of second column to 1 for the fields .
1 - The system will take into account the second and third
!   positions of the tags for heading uniqueness and authority
!   matching algorithms.
!!!!!-!
100## 1
110## 1
111## 1
130## 0
240## 0
245## 1
246## 1
247## 1
430## 0
440## 0
600## 1

610## 1
611## 1
630## 1
648## 1
650## 1
651## 1
655## 1
700## 1
710## 1
711## 1
730## 0
800## 1
810## 1
811## 1
830## 0

tab_filing: (xxx01)
Check tab00 ACC (browse index) section, column 5 which is the filing procedure that is specified in tab_filing. The filing procedure for Author, Title and Subjects are 01, 11 and 12 in the following case.

```
!BROWSE Index              tab00
!   2       3     4 5  6   7                          11
!-!!!!!-!!!!!-! !!-!!-!-!---!!-!!!!!-!!!!!!!!!!!!!!!!!!!!!!!
H TTL    ACC     11 00      00        Titles
H TTLC   ACC     61 00      00        Titles Chinese
H TTLJ   ACC     62 00      00        Titles Japanese
H TTLK   ACC     63 00      00        Titles Korean
H TRT    ACC     11 00      00        Title/Related Titles
H ATH    ACC     01 00      00        Authors
H AUT    ACC     01 00      00        Authors/Names
H AUTC   ACC     51 00      00        Authors Chinese
H AUTJ   ACC     52 00      00        Authors Japanese
H AUTK   ACC     53 00      00        Authors Korean
H AWT    ACC     01 00      00        Authors+Titles
H SRS    ACC     11 00      00        Series
H SUB    ACC     12 00      00        Subjects (all)
H SUL    ACC     11 00      00        LC Subjects
H SUM    ACC     11 00      00        MeSH Subjects
H SUE    ACC     11 00      00        NLC English Subjects
H ACO    ACC     11 00      00        A.Corp
H TUT    ACC     11 00      00        Uniform title
H SPE    ACC     11 00      00        Person
H SCO    ACC     11 00      00        Corp.
H SME    ACC     11 00      00        Meeting
H SUT    ACC     11 00      00        U.Title
H SGE    ACC     11 00      00        Geographic
H PST    ACC     22 00      00        LC Call Number
H GNO    ACC     20 00      00        Gov.Doc.Number
H IPS    ACC     20 00      00        Item Process Status
H IPP    ACC     20 00      00        Sub Library Code+Ite
H XYZ    ACC     12 00      00        Subjects -vxyz
H SULX   ACC     11 00      00        LC Subjects -vxyz
```

The following **lines in Blue** need to be added into tab_filing (xxx01) in author, title and subject filing procedures.
(F- filing routine for z01_rec_key in oracle table z01
N- Normalization filing routine for z01_normalized_text in oracle table z01.
D- display text filing routine for z01_display_text in oracle table z01)

<p style="color:red">**tab_filing**</p>

```
01 D end_punctuation     :,=;/.
01 N end_sub_punctuation  :,=;/.
```
<p style="color:blue">**01 N to_blank   #**</p>
<p style="color:blue">**01 N add_prefix_hash   7**</p>
<p style="color:blue">**01 F to_blank   #**</p>
<p style="color:blue">**01 F add_prefix_hash   7**</p>

```
01 F del_subfield
01 F to_lower
01 F abbreviation
01 F suppress
01 F compress           '
01 F to_blank          /-:;
01 F to_blank          !@$%^&*()_+-={}[]:";<>?,./~`
01 F mc_to_mac
01 F pack_spaces
01 F char_conv         FILING-KEY-01
!*rem titles subjects
11 D end_punctuation     :,=;/.
11 N end_sub_punctuation  :,=;/.
```
<p style="color:blue">**11 N to_blank   #**</p>
<p style="color:blue">**11 N add_prefix_hash   7**</p>
<p style="color:blue">**11 F to_blank   #**</p>
<p style="color:blue">**11 F add_prefix_hash   7**</p>

```
11 F del_subfield
11 F to_lower
11 F suppress
11 F abbreviation
11 F numbers
11 F non_filing
11 F compress           '
11 F to_blank          !@$%^&*()_+-={}[]:";<>?,./~`
11 F expand_num
11 F pack_spaces
11 F char_conv         FILING-KEY-01
!* for subject (all)
12 D end_punctuation     :,=;/.
12 N end_sub_punctuation  :,=;/.
```
<p style="color:blue">**12 N to_blank   #**</p>
<p style="color:blue">**12 N add_prefix_hash   7**</p>

12 F to_blank   #
12 F add_prefix_hash   7
12 F to_lower
12 F del_subfield
12 F abbreviation
12 F suppress
12 F numbers
12 F compress   ""`<>%
12 F to_blank   ./-:;,{}[]()?!_
12 F pack_spaces
12 F char_conv   FILING-KEY-01
12 F expand_num
12 F non_filing
12 F to_blank   ^

Caution: Be sure, you need to have **to_blank   #** filing routine before **add_prefix_hash 7.** The "#" shouldn't appear in any filing routine after **add_prefix_hash** routine.

XXX10: The files need to set up in Authority library.

tab_acc_category: (xxx10)
This is a new file and can be copied from a_tree **.** Change the value of second column from 0 to 1.

```
! 1   2
!!!!!-!
100## 1
110## 1
111## 1
130## 0
150## 1
151## 1
155## 1
400## 1
410## 1
411## 1
430## 0
450## 1
451## 1
455## 1
COR## 0
500## 1
510## 1
511## 1
530## 0
550## 1
551## 1
```

Note: 5XX is optional in the tab_acc_category. If you use tab07 for defining links between authority headings for the creation of z103 link records, the 5XX has to be in the tab_acc_category. This is not in the SLGA18 doc. (Thanks to Allison from Harvard and Randy from Exlibris for discovering it).

tab11_acc: (xxx10)

The following two lines need to be added into tab11_acc of authority library (SLGA18 p13-14)

```
4####     7 *       DUP  -7
5####               GXX  -wi056
```

tab_filing: (xxx10)
Add two lines to the ACC routine number 11. You may have a different number of filing procedures for ACC index routine. You can find it in tab00. It is the value of column 5 in ACC section.   Make sure the "#" doesn't appear in any routine (to_blank, compress,etc) after add_prefix_hash.

```
!ACC browse index        stb10/tab00        --------------------
H GEN   ACC    11 00                       General
H TOP   ACC    11 00                       LCSH
H LCN   ACC    11 00        00             Class numbers
H PER   ACC    11 00        00             Personal names
H COR   ACC    11 00        00             Corporate names
H MET   ACC    11 00        00             Meeting names
H TIT   ACC    11 00        00             Uniform titles
H GEO   ACC    11 00        00             Geographic names
H GNR   ACC    11 00        00             Genre
H SBD   ACC    11 00        00             GenSub-division
H CBD   ACC    11 00        00             ChronSub-division
H GBD   ACC    11 00        00             GeogSub-division
!--GXX for 5XX, DUP for Ambiguous heading 07/19/2007
H Z0101 ACC    11 00        00             General (Brief rec.)
H GXX   ACC    11 00        00             General (Secondary)
H DUP   ACC    11 00        00             Duplicate
```

tab_filing
11 F to_blank   #
11 N add_prefix_hash   7

tab00 : (xxx10)
Added GXX for 5XX, DUP for Ambiguous headings in the ACC index section.
H Z0101 ACC   11 00   00   General (Brief rec.)
H GXX   ACC   11 00   00   General (Secondary)
H DUP   ACC   11 00   00   Duplicate

tab_fix: (xxx10)
Add fix_doc_aut_duplicate in INS, INS2 and   UE-01. SLGA18 p.39 states: "add fix_doc_aut_duplicate in OCLC." This is wrong and will create $$7 for every cross

reference even though it is not a conflict heading. Be sure, no fix_doc_aut_duplicate is added to the OCLC fix routine. The tab_fix for stb10 should look like:

UE-01 fix_doc_aut_duplicate
INS   fix_doc_preferred
INS   fix_doc_aut_duplicate
OCLC fix_doc_perferred
OCLC fix_doc_merge
INS2 fix_doc_aut_duplicate
UE-01 fix_doc_aut_duplicate

edit_field.eng (stb10)
Add "-7" to column 6 to strip out subfield $$7. Should look like:

```
! 2    3    4 5        6         7 8        9              10
!-!-!!!!!-!-!-!!!!!!!!!!!-!-!-!!!!!!!!!!!-!!!!!!!!!!
1 # GEN## C    -7
2                          a  A
2                          b  A  ^
2                          c  A  ^
2                          d  A  ^
2                          q  A  ^
2                          t  A  ^                    ^
2                          z  A  ^--^
2                          x  A  ^--^
2                          y  A  ^--^
2                          v  A  ^--^
2                          p  A  ^--^
1 # 4#### S    -w297
2                          a  A  ^
2                          a  8  ^--^
2                          b  A  ^
2                          c  A  ^
2                          d  A  ^
2                          v  A  ^--^
2                          x  A  ^--^
1 # 4#### D    -w2897
2                          a  A  ^
2                          b  A  ^
2                          c  A  ^
2                          d  A  ^
2                          q  A  ^
2                          t  A  ^
2                          v  A  ^--^
2                          x  A  ^--^
2                          #  A
```

Minimal Indexes necessary to run and activate New Authority Control:

Important:

Before running index jobs, you need have all of the above files ready and in place, then using p_manage_21 globally change UPD from "N" to "Y" in authority library. Then you can run the index as following sequence. The jobs can be run from Service Utility in the GUI or Command line in the Server.

STB01:

p_manage_102,
Aleph18>csh   -f   p_manage_102   stb01,stb10,1,000000000,999999999,1xxt,9,>& $alephe_scratch/p_manage_102.log &

p_manage_02
Aleph18>csh   -f   p_manage_02   stb01,0,000000000,999999999,,9,Y,N,>& $alephe_scratch/p_manage_02_stb01.log &

STB10:

 p_manage_02,
Aleph18>csh   -f   p_manage_02   stb10,1,000000000,999999999,,7,Y,N,>& $alephe_scratch/p_manage_02_stb10.log &

p_manage_12 :
to create link table z103 if you use tab07 for link between authority headings.

Aleph18>csh   -f   p_manage_12   stb10,Y,N,000000000,999999999,06,>& $alephe_scratch/p_manage_12_stb10.log &

Testing

The list below is a sample of subject headings and cross references which were identified in version 17 as ambiguous or conflicting headings. After the new authority control implementation they were searched and all passed the test. What follows are illustrated examples of the types of problems encountered and their ultimate solution.

Portraits --- Used in Subject Heading (150) and Series title cross reference (430)

PAS --- Used as Corporate bodies cross reference (410) in 8 different records

China --- China used as geographic subject heading (151)

Sin --- Used as subject heading (150) and as cross reference (451) of 'China'

Phonology --- Used as cross reference(450) in two separate subject headings

Mythology --- Used as Uniform tile 130 and topic heading 150

Environmental biotechnology --- X-ref 450 and 130

Developmental disabilities --- 150 and 130

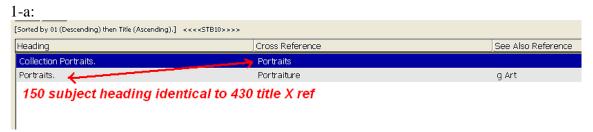*Abstinence* --- Two identical cross references (450) in separate headings

Arts, Useful --- as above

Let us see the difference before and after new authority implementation.

Example 1: Portraits

When a term in an index in the authority library matches the spelling of a term in another index the system cannot function correctly.

Screenshot 1-a confirms the duplication by identifying their different index locations on the brief list display: a subject heading and a matching cross reference for a series uniform title.

1-a:

[Sorted by 01 (Descending) then Title (Ascending).] <<<<STB10>>>>		
Heading	Cross Reference	See Also Reference
Collection Portraits.	Portraits	
Portraits.	Portraiture	g Art

150 subject heading identical to 430 title X ref

Screenshot 1-b indicates a problem when a subject browse search reveals existing documents and an adjacent x-ref symbol. This occurs when authority records in different indexes have identical texts.
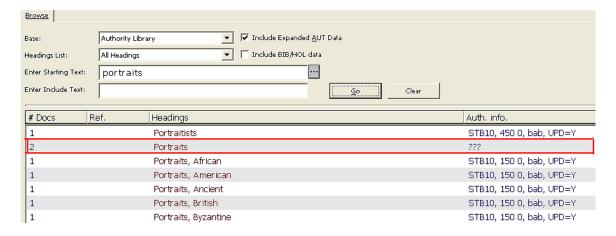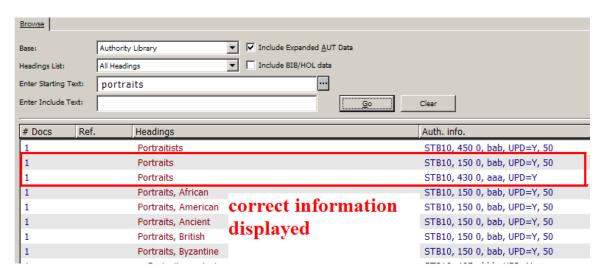
1-b:

Before implementation:

After implementation:

Screenshot 1-c confirms the problem in an authority library search when the conflict symbols of three question marks appear in the right column.

1-c:

Before implementation:

After implementation:

Screenshot 1-d shows the results in the OPAC browse subject search. The term refers to the cross reference of the uniform title which is wrong and what is the result of clicking on heading information.

1-d:
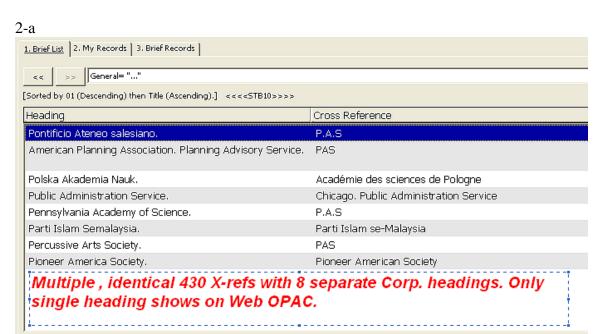Before Implementation:

Browse List: Subjects (all)

No. of Recs	Entry
	Portraitists - [Heading information]
	See: Portrait painters
3	Portraits - [Heading information]

Click on it and you get title heading instead of subject heading which is wrong

Sys. no.	000001478
Heading	Collection Portraits

After implementation:

Browse List: Subjects (all)

No. of Recs	Entry
	Portraitists - [Heading information]
	See: Portrait painters
3	Portraits - [Heading information]
	Portraits, African - [Heading i

Click on it and you will see correct subject heading information

Sys. no.	000568101
Heading	Portraits
Search also under	subdivision Biography--Portraitsunder names of countries, cities subdivision Artunder names of individual persons who lived befor individual persons who lived after 1400 and individual families, ar wars, e.g. United States--History--Civil War, 1861-1865--Portraits
Broader topic	Art
Broader topic	Biography
Broader topic	Pictures

Example 2: PAS

      This problem occurs when the identical cross references are used for different author headings. PAS is a cross reference for eight of corporate bodies.

Screenshot 2-a displays the eight corporate bodies the cross reference generated.

2-a

Screenshot 2-b: The multiple hits by the same term in the Authority file are signaled by question marks.

2-b:

Screenshot 2-c displays the result in the OPAC. Most corporate bodies do not display.

2-c:

Before implementation:

Browse List: Authors

No. of Recs	Entry
1	Parziale, Jim
	PAS - [Heading information] See: Percussive Arts Society
	P.A.S - [Heading information] See: Percussive Arts Society
	Pas - [Heading information] See: Public Administration Service
1	Pas, Alice van der, 1934- - [Heading information]

After implementation: All headings display.

Browse List: Authors

No. of Recs	Entry
1	Parziale, Jim
	Pas - [Heading information] See: Public Administration Service
	PAS American Planning Association. Planning Advisory Service - [Heading information] See: American Planning Association. Planning Advisory Service
	P.A.S American Planning Association. Planning Advisory Service - [Heading information] See: American Planning Association. Planning Advisory Service
	P.A.S Pennsylvania Academy of Science - [Heading information] See: Pennsylvania Academy of Science
	PAS Pennsylvania Academy of Science - [Heading information] See: Pennsylvania Academy of Science
	PAS Percussive Arts Society - [Heading information] See: Percussive Arts Society
	P.A.S Percussive Arts Society - [Heading information] See: Percussive Arts Society
	PAS Pioneer America Society - [Heading information] See: Pioneer America Society
	PAS Polska Akademia Nauk - [Heading information] See: Polska Akademia Nauk

Example 3: China and Sin

Subject headings share the same word. One is a Topical heading 150 (Sin) and the other a 451 (Sin) See cross reference for the Geographical heading 'China'. If UPD is set to Y this can be the most serious and dramatic scenario because the cross reference blocks the correct heading in the bibliographic record and converts it to a corrupted form resulting in the 650 "Sin" to be displayed as 650 "China".

Screenshot 3-a OPAC displays 'Sin' as subject. No number of bib records found since all replaced by 'China'. Authority record on right confirms the 150.

3-a:

Screenshot 3-b displays the presence of the Subject Authority for China. Just below is the 451 now a conflicting term. There are over 50+ bib records with 'China' as a subject.

3-b:

Let us check the each record closely:

The dominant heading is the 651 (China) its x-ref reference (Sin) overlaid the 650 matching text (Sin) as following screenshot 3-c:

3-c

24500	|6 01 |a Hai wai xue zhe lun Zhongguo / |c Zhang Jingfu zhu bian.
24500	|6 01 |a 海外学者论中国 / |c 张劲夫主编.
250	|6 02 |a Beijing di 1 ban.
250	|6 02 |a 北京第1版.
260	|6 03 |a Beijing : |b Hua xia chu ban she : |b Xin hua shu dian jing xiao, |c 1994.
260	|6 03 |a 北京 : |b 華夏出版社 : |b 新华书店经销, |c 1994.
300	|a 6, 5, 506 p. : |b ill. ; |c 21 cm.
504	|a Includes bibliographical references.
651 0	|a China. ⟶ **This is correct geographical heading**
7001	|a Zhang, Qingfu |6 04
7001	|6 04 |a 张劲夫.

10010	|a Pittenger, W. Norman |q (William Norman), |d 1905-
24510	|a Cosmic love and human wrong : |b the reconception of meaning of sin, in the light
2600	|a New York : |b Paulist Press, |c c1978.
300	|a v, 102 p. ; |c 20 cm.
500	|a "The chapters ... were originally delivered as lectures at Texas Christian University
650 0	|a China. ⟶ **Wrong, it was 'Sin' subject heading which**
650 0	|a Process theology **replaced by 'China'**

After implementation:

The new program has separated the topical (650) from the geographic fields (651) even though they have the same text. In the screenshot below the "China" in 650 is separated from "China" in 651. To correct "China" in 650 to "Sin", a manual effort has to be made.

Browse List: Subjects (all)

No. of Recs	Entry
	Chin, Yung, 1924- - [Heading information] See: Jin, Yong, 1924-
50+	China - [Heading information]
18	China
1	China -- 1949- -- Eco
1	China -- 1949- -- Eco
4	China -- Administrative and political divisions

651 field is correct geo heading

650 text 'Sin' replaced by 'China' which must be corrected manually

After manually changing "China" to "Sin" in each record in the 650 bib records, search under this as a subject and the Heading information will display correctly.

Browse List: Subjects (all)

No. of Recs	Entry
	Simyang-si (China) - [Heading information] See: Shenyang (Liaoning Sheng, China)
18	Sin - [Heading information]
	Sin - [Heading information] See: China
1	Sin (Assyro-Babylonian deity)
2	Sin -- Biblical teaching

Stony Brook University   18
Library Information System – Fang Peng, David Bolotine
July 23, 2008

The Problem and Solution:

Problem 1: OCLC export authority record didn't overlay older records.

Solve: Created a new overlay routine for authority library.

tab_match
tab_merge
tab_merge_overlay

add following line in OCLC routine of tab_fix
!--added for overlay authority records
!--files need to change on tab_match, tab_merge, tab_merge_overlay
OCLC fix_doc_merge

Problem 2: New/updated authority record didn't change bib record headings.

This heading was not a conflict heading. After exporting new authority record by overlaying the older one, there was $$7 generated and appended to cross reference which causing the heading not match.

New export overlaid original record. But created $$7 for every cross reference.

```
AU System No. 560909 (Mukden Incident, China, 1931) Year:

Leader          LDR  __ _  ^^^^^cz^^a2200181n^^45^0
Control No. ID  003  __ _  OCoLC
Date and Time   005  __ _  20070730024911.0
Fixed Data      008  __ _  050301||^anannbabn^^^^^^^^^^|a^ana^^^^^^
LC Control No.  010  __ a  sh 85088295
Catalog. Source 040  __ a  DLC
                     b  eng
                     c  DLC
                     d  ICU
                     d  DLC
LC Class. No.   053  0 a  DS783.8
Topical Term    150  __ a  Mukden Incident, China, 1931
See).Trac.Topic 450  __ a  Manchurian Incident, China, 1931
                     7  Mukden Incident, China, 1931
See).Trac.Topic 450  __ a  Mukden Incident, 1931
                     w  nne
                     7  Mukden Incident, China, 1931
See).Trac.Geo.N 451  __ a  Manchuria (China)
                     x  History
                     y  Incident, 1931
                     7  Mukden Incident, China, 1931
SAF.Trac.Topic  550  __ a  Sino-Japanese Conflict, 1931-1933
                     w  g
Sourc.Dat.Found 670  __ a  Britannica Micro.
                     b  (Mukden Incident, also called Manchurian Incident (1931), seizure of the Manchurian city of Mukden
                        by Japanese troops, followed by the Japanese invasion of all of Manchuria and the establishment of
                        the Japanese-dominated state of Manchukuo in the area)
```

Solve:

Command out fix_dic_aut_duplicate from OCLC routine.

!OCLC fix_doc_aut_duplicate

!--remove fix_doc_aut_duplicate from OCLC routine added UE-01 for overlay not generate $$7

UE-01 fix_doc_aut_duplicate

Problem 3: New/updated authority records did not generate bib links.
(The Subjects (all) does not display new Auth Info in OPAC and GUI. It displays in Subjects (LC))

The subject heading of authority was updated. The heading was flipped to bib records. But only LC subject index(SUL) linked to the authority record. All subject headings (SUB) were not linked to authority records.

stb01.tab_aut : add line SUB   2 STB10

stb01.tabl20 : add 2 lines

```
    1 SUB  SUB  150##                0
    2           450##        -wi5    0 SEEF

    1 SUB  SUB  151##                0
    2           451##        -wi5    0 SEEF
```

** Can't add '!' with comments between the group. Otherwise, after '!... ' won't work.

Problem 4: ACC sorting is fine in stb10, but the sorting in stb01 is still not right.

This is in stb10
DES
D.E.S.
Des Abbayes, H. (Henry)
Des Abbayes, Henry

This is in stb01
Des Granges, Charles Marc, 1861-
DES   ($$7Great Britain. Dept. of Education and Science)
D.E.S.
Des Hons, Gabriel
Des Houliéres, Antoinette, ca. 1634-1694
Des Houliéres, Madame (Antoinette), ca. 1634-1694
DES   ($$7International Committee of the Red Cross)
D.E.S.
Des Issarts, Jean Henri Bancal, 1750-1826

Solve: Add following lines and rerun indexing.
01 N to_blank   #
01 N add_prefix_hash   7
01 F to_blank   #
01 F add_prefix_hash   7
 (F- filing z01_rec_key, N- Normalization – z01_normalized_text, D- display text – z01_display_text in z01)

Problem 5: 'See Also' (5XX) disappeared from Web OPAC "heading information" after using new authority mechanism.

Solve: add 5XX in the stb10/tab_acc_category and run p_manage_12 to recreate z103 link table.