**The duplicate VMs cleanup process**

involves insert/modification records in existing 4 tables:

SBR.VALUE\_MEANINGS

SBR.DESIGNATIONS

SBR.DEFINITIONS

SBREXT.AC\_ATT\_CSCSI\_EXT

For the duplicate VMs cleanup process  were created

5 tables in SBREXT schema

SBREXT.MDSR\_VM\_DUP\_REF --to keep truck of final and dup records

SBREXT.MDSR\_VM\_DUP\_REF\_TMP to upload synonyms which are determinated by client

SBREXT.MDSR\_DUP\_VM\_ERR  --to keep truck of errors

SBREXT.MDSR\_SYNONYMS\_XML --to load webservices XML respond

SBREXT.MDSR\_CONCEPTS\_SYNONYMS  --to load concept names and synonyms

Package MDSR\_CLEAN\_VM\_DUPLICATES of 8 Stored Procedures and 1 function in SBREXT schema

FUNCTION MDSR\_GET\_CONCEPT\_SYNONYM(p\_code IN VARCHAR2 ,p\_NAME IN VARCHAR2);

    PROCEDURE MDSR\_CALL\_NCI\_WEBSERVICE;

    PROCEDURE MDSR\_UPDATE\_SYNONYMS\_XML;

    PROCEDURE MDSR\_INSERT\_CONCEPT\_SYN;

    PROCEDURE MDSR\_INSERT\_VM\_FINAL\_DUP\_REF;

    PROCEDURE MDSR\_CREATE\_DUP\_VM\_DES\_DEF;

    PROCEDURE MDSR\_CREATE\_DUP\_VM\_CSI;

    PROCEDURE MDSR\_CHECK\_FVM\_DES\_DEF\_CSI;

    PROCEDURE MDSR\_VM\_DUP\_ROLLBACK;

CHANGES made by the process IN SBR.VALUE\_MEANINGS:

UPDATE to "RELEASED" workflow status for each Final VM.

UPDATE to "RETIRED ARCHIVED" workflow status duplicate VM record and add to CHANGE\_NOTE 'Use VM public ID: FIN\_VM Version 1.0 instead. Modified by caDSR script.

CHANGES made by the process INSBR.DEFINITIONS

Add preferred definition record with definition type= “Prior preferred definition” – if not in definitions type list.

DEFINITIONS are created  for FINAL VM from existing DEFINITIONS of retired VM if do not exist yet.

CHANGES made by the process INSBR.DESIGNATIONS

DESIGNATION is created for each Retired VM.

DESIGNATIONS are created for FINAL VM from existing DESIGNATIONS of retired VM if do not exist yet.

CHANGES made by the process IN SBREXT.AC\_ATT\_CSCSI\_EXT

CLASSIFICATIONS of DESIGNATIONS are created for FINAL VM  from retired VM CLASSIFICATIONS of DESIGNATIONS

CLASSIFICATIONS of DEFINITIONS where created for FINAL VM  from retired VM CLASSIFICATIONS of DEFINITIONS

Table SBREXT.MDSR\_VM\_DUP\_REF is created as  across reference table between final and retired SBR.VALUE\_MEANINGS duplicate records

marked by the duplicate VMs cleanup process.

Table SBREXT.MDSR\_DUP\_VM\_ERR is created to store reasons of failures of the process.

**Following procedures** were developed load concept’s synonyms NCI thesaurus to DB table   
 SBREXT.MDSR\_CONCEPTS\_SYNONYMS:

(1) PROCEDURE MDSR\_CALL\_NCI\_WEBSERVICE;

    (2)PROCEDURE MDSR\_UPDATE\_SYNONYMS\_XML;

    (3)PROCEDURE MDSR\_INSERT\_CONCEPT\_SYN;

**(4)MDSR\_INSERT\_VM\_FINAL\_DUP\_REF Procedures DESCRIPTION**

**1st essential group of VM duplicate records and Current Algorithm of identifying Final and duplicate VMs:**

*duplicate VMs with Concepts(CONDR\_IDSEQ is NOT NULL) but Concepts Name not like "integer".*

There are 3 group of duplicate sets of not RETIRED VMs:  
1. When find VM with long name =Concept name in duplicate set.  
2 When not find VM with long name =Concept name in duplicate set, but find VM long name = Synonym name   
3.When not find VM with long name =Concept name or VM long name = Synonym name in duplicate set.

Final VM was identified from the set of VM with the same Concept Derivation Rule ID (CDRI):  
1.When Final VM long name =Concept name and is most current VM .  
Duplicate VMs for this CDRI can have Long name =Concept name or Synonym  
2.When Final VM long name = Synonym name and is most current VM (No VM long name =Concept name in find duplicate set of VMs).  
Duplicate VMs for this CDRI can have Long name = Final VM long name or other Synonym  
3.When No VM long name =Concept or other Synonym name in find duplicate set of VMs( per Concept Derivation Rule ID)  
Final VM is most current VM per Concept Derivation Rule ID.  
Duplicate VMs for this CDRI have same Long name as Final VM.

***SPEP 1A***

First insert into MDSR\_VM\_DUP\_REF  FINAL VM Records with CONDR\_IDSEQ is not null

and Concepts Name not like "integer" and Concepts Name=VM.LONG\_NAME:

INSERT INTO SBREXT.MDSR\_VM\_DUP\_REF

(FIN\_VM,FIN\_IDSEQ ,VM\_ID,VM\_IDSEQ,CONCEPTS\_CODE,CONCEPTS\_NAME,LONG\_NAME,CONDR\_IDSEQ, PROCESSED,CONCEPT\_SYNONYM ,DATE\_INSERTED,PREFERRED\_DEFINITION)

select  distinct FIN\_VM ,VM\_IDSEQ FIN\_IDSEQ,VM\_ID,VM\_IDSEQ,name CONCEPTS\_CODE,CONCEPT\_NAME,VM\_NAME,CONDR\_IDSEQ,'FINAL','CONCEPT',SYSDATE,PREFERRED\_DEFINITION

  from

(

select  max(VM\_ID) over  (partition by CONDR\_IDSEQ order by CONDR\_IDSEQ ) as FIN\_VM

,VM\_ID,VM\_NAME,CONCEPT\_NAME,CONDR\_IDSEQ,VM\_IDSEQ ,name,PREFERRED\_DEFINITION from

(

select VM\_ID,VM\_IDSEQ,VM.LONG\_NAME VM\_NAME,CN.LONG\_NAME CONCEPT\_NAME,VM.CONDR\_IDSEQ,name,

UPPER(trim(VM.PREFERRED\_DEFINITION))PREFERRED\_DEFINITION

from SBR.VALUE\_MEANINGS VM,

SBREXT.CON\_DERIVATION\_RULES\_EXT DER,

SBREXT.CONCEPTS\_EXT CN,

(

select COUNT(\*),CONDR\_IDSEQ from SBR.VALUE\_MEANINGS VM

where   UPPER(ASL\_NAME) not like '%RETIRED%' AND CONDR\_IDSEQ is not null

GROUP BY CONDR\_IDSEQ HAVING COUNT(\*)>1

)DUP

where   VM.CONDR\_IDSEQ=DER.CONDR\_IDSEQ

AND VM.CONDR\_IDSEQ=DUP.CONDR\_IDSEQ

and replace(DER.NAME,'Rh Positive Blood Group','C76251')=CN.PREFERRED\_NAME

AND UPPER(VM.ASL\_NAME) not like '%RETIRED%'

AND TRIM(UPPER(VM.LONG\_NAME)) =TRIM(UPPER(CN.LONG\_NAME))

AND instr(DER.NAME,'C45255')=0

))

where FIN\_VM=VM\_ID

UNION

  select   FIN\_VM ,VM\_IDSEQ FIN\_IDSEQ,VM\_ID,VM\_IDSEQ,name CONCEPTS\_CODE,CONCEPT\_NAME,VM\_NAME,CONDR\_IDSEQ,'FINAL','CONCEPT',SYSDATE,PREFERRED\_DEFINITION from

(

select  max(VM\_ID) over  (partition by CONDR\_IDSEQ order by CONDR\_IDSEQ ) as FIN\_VM

,VM\_ID,VM\_NAME,CONCEPT\_NAME,name,CONDR\_IDSEQ,VM\_IDSEQ ,PREFERRED\_DEFINITION from

(

SELECT VM\_ID,VM\_IDSEQ,name,trim(UPPER(CONCEPT\_NAME))CONCEPT\_NAME,trim(UPPER(LONG\_NAME)) VM\_NAME,VM.CONDR\_IDSEQ,UPPER(trim(VM.PREFERRED\_DEFINITION))PREFERRED\_DEFINITION

FROM  SBR.VALUE\_MEANINGS VM,

(SELECT M.CONDR\_IDSEQ,name, LISTAGG(M.LONG\_NAME,' ') WITHIN GROUP (ORDER BY M.ELM\_ORDER) as CONCEPT\_NAME

FROM  (select CONDR\_IDSEQ,name,spl.preferred\_name,ELM\_ORDER,LONG\_NAME

from

(select distinct

CONDR\_IDSEQ,name,

  trim(regexp\_substr(replace(replace(name,'Rh Negative Blood Group','C76252'),'Rh Positive Blood Group','C76251'), '[^:]+', 1, levels.column\_value)) as preferred\_name,levels.column\_value ELM\_ORDER

from

(select \*from SBREXT.CON\_DERIVATION\_RULES\_EXT

where  instr(name,':')>0) t,

table(cast(multiset(select level from dual connect by level <= length (regexp\_replace(t.name, '[^:]+')) + 1) as sys.OdciNumberList))

  levels) spl,

  sbrext.concepts\_ext  c

  where spl.preferred\_name=c.preferred\_name) M

GROUP BY name,M.CONDR\_IDSEQ)VW,

(select count(\*),CONDR\_IDSEQ from SBR.VALUE\_MEANINGS VM

where  UPPER(ASL\_NAME) not like '%RETIRED%'

having count(\*)>1GROUP BY CONDR\_IDSEQ )DR

where    VW.CONDR\_IDSEQ=VM.CONDR\_IDSEQ

AND  VW.CONDR\_IDSEQ=DR.CONDR\_IDSEQ

and UPPER(ASL\_NAME) not like '%RETIRED%'

and  trim(UPPER(VM.LONG\_NAME))=trim(UPPER(CONCEPT\_NAME))

and instr(NAME,'C45255')=0 ))

where FIN\_VM=VM\_ID

ORDER BY 4,2 desc ;

**Statement description from business rules.**

This is union of 2 selects:

1.select of single VM concepts

select  distinct FIN\_VM ,VM\_IDSEQ FIN\_IDSEQ,VM\_ID,VM\_IDSEQ,name CONCEPTS\_CODE,CONCEPT\_NAME,VM\_NAME,CONDR\_IDSEQ,'FINAL','CONCEPT',SYSDATE,PREFERRED\_DEFINITION

  from

(

select  max(VM\_ID) over  (partition by CONDR\_IDSEQ order by CONDR\_IDSEQ ) as FIN\_VM

,VM\_ID,VM\_NAME,CONCEPT\_NAME,CONDR\_IDSEQ,VM\_IDSEQ ,name,PREFERRED\_DEFINITION from

(

select VM\_ID,VM\_IDSEQ,VM.LONG\_NAME VM\_NAME,CN.LONG\_NAME CONCEPT\_NAME,VM.CONDR\_IDSEQ,name,

UPPER(trim(VM.PREFERRED\_DEFINITION))PREFERRED\_DEFINITION

from SBR.VALUE\_MEANINGS VM,

SBREXT.CON\_DERIVATION\_RULES\_EXT DER,

SBREXT.CONCEPTS\_EXT CN,

(

select COUNT(\*),CONDR\_IDSEQ from SBR.VALUE\_MEANINGS VM

where   UPPER(ASL\_NAME) not like '%RETIRED%' AND CONDR\_IDSEQ is not null

GROUP BY CONDR\_IDSEQ HAVING COUNT(\*)>1

)DUP

where   VM.CONDR\_IDSEQ=DER.CONDR\_IDSEQ

AND VM.CONDR\_IDSEQ=DUP.CONDR\_IDSEQ

and replace(DER.NAME,'Rh Positive Blood Group','C76251')=CN.PREFERRED\_NAME

AND UPPER(VM.ASL\_NAME) not like '%RETIRED%'

AND TRIM(UPPER(VM.LONG\_NAME)) =TRIM(UPPER(CN.LONG\_NAME))

AND instr(DER.NAME,'C45255')=0

))

where FIN\_VM=VM\_ID

2.select of multiple VM concepts

select   FIN\_VM ,VM\_IDSEQ FIN\_IDSEQ,VM\_ID,VM\_IDSEQ,name CONCEPTS\_CODE,CONCEPT\_NAME,VM\_NAME,CONDR\_IDSEQ,'FINAL','CONCEPT',SYSDATE,PREFERRED\_DEFINITION from

(

select  max(VM\_ID) over  (partition by CONDR\_IDSEQ order by CONDR\_IDSEQ ) as FIN\_VM

,VM\_ID,VM\_NAME,CONCEPT\_NAME,name,CONDR\_IDSEQ,VM\_IDSEQ ,PREFERRED\_DEFINITION from

(

SELECT VM\_ID,VM\_IDSEQ,name,trim(UPPER(CONCEPT\_NAME))CONCEPT\_NAME,trim(UPPER(LONG\_NAME)) VM\_NAME,VM.CONDR\_IDSEQ,UPPER(trim(VM.PREFERRED\_DEFINITION))PREFERRED\_DEFINITION

FROM  SBR.VALUE\_MEANINGS VM,

(SELECT M.CONDR\_IDSEQ,name, LISTAGG(M.LONG\_NAME,' ') WITHIN GROUP (ORDER BY M.ELM\_ORDER) as CONCEPT\_NAME

FROM  (select CONDR\_IDSEQ,name,spl.preferred\_name,ELM\_ORDER,LONG\_NAME

from

(select distinct

CONDR\_IDSEQ,name,

  trim(regexp\_substr(replace(replace(name,'Rh Negative Blood Group','C76252'),'Rh Positive Blood Group','C76251'), '[^:]+', 1, levels.column\_value)) as preferred\_name,levels.column\_value ELM\_ORDER

from

(select \*from SBREXT.CON\_DERIVATION\_RULES\_EXT

where  instr(name,':')>0) t,

table(cast(multiset(select level from dual connect by level <= length (regexp\_replace(t.name, '[^:]+')) + 1) as sys.OdciNumberList))

  levels) spl,

  sbrext.concepts\_ext  c

  where spl.preferred\_name=c.preferred\_name) M

GROUP BY name,M.CONDR\_IDSEQ)VW,

(select count(\*),CONDR\_IDSEQ from SBR.VALUE\_MEANINGS VM

where  UPPER(ASL\_NAME) not like '%RETIRED%'

having count(\*)>1GROUP BY CONDR\_IDSEQ )DR

where    VW.CONDR\_IDSEQ=VM.CONDR\_IDSEQ

AND  VW.CONDR\_IDSEQ=DR.CONDR\_IDSEQ

and UPPER(ASL\_NAME) not like '%RETIRED%'

and  trim(UPPER(VM.LONG\_NAME))=trim(UPPER(CONCEPT\_NAME))

and instr(NAME,'C45255')=0 ))

where FIN\_VM=VM\_ID

The yellow highlighted part of the select takes apart name from SBREXT.CON\_DERIVATION\_RULES\_EXT connected by ‘:’ and replaces them with Concept Names from SBREXT.CONCEPTS\_EXT.

Green part concatenates Concept Names in to string for given CONDR\_IDSEQ.

*STEP 1B*

insert into MDSR\_VM\_DUP\_REF DUPLICATE Retired VM Records for each FINAL VM found in step 1A

when long name of Retired VM can be = CONCEPT\_NAME  or SYNONYM (MDSR\_CLEAN\_VM\_DUPLICATES.MDSR\_GET\_CONCEPT\_SYNONYM(CONCEPTS\_CODE,vm.LONG\_NAME)=1)

*STEP 2A*

insert Final VM records with matching synonyms for single concepts

 2.Provide the client with the List of FINAL and DUP VM with multiple concepts where duplicate VM LONG\_NAME<>CONCEPT\_NAME.

   then load them to SBREXT.MDSR\_VM\_DUP\_REF as duplicates.

3.FIND set of DUP VM where no one VM name =CONCEPT\_NAME  (per VW.CONDR\_IDSEQ)

   After the most current VM (Final VM) found ,they are inserted  to  in the group, the older records inserted in to SBREXT.MDSR\_VM\_DUP\_REF

   with Final VM Public ID and VM\_IDSSEQ.

2.duplicate VMs without Concepts (CONDR\_IDSEQ is NULL).

  The records in these sets are grouped by VM Long\_name and PREFERRED\_DEFINITION.

  After the most current VM (Final VM) found  in the group, the older records inserted in to SBREXT.MDSR\_VM\_DUP\_REF along

  with Final VM Public ID(FIN\_VM) and VM\_IDSSEQ(FIN\_IDSEQ).

  To avoid insert duplicate records into SBREXT.MDSR\_DUP\_VM\_ERR we delete from found duplicate set of records (A)

  already existed records in SBREXT.MDSR\_DUP\_VM\_ERR(B)

 Query A.

   retrieves only records which will be set to retired with final VM record ID.

   select  FIN\_VM,vm\_id,VM\_IDSEQ,CONDR\_IDSEQ,CONCEPTS\_NAME

   from (

   select max(VM\_ID) over  (partition by VM.CONDR\_IDSEQ order by VM.CONDR\_IDSEQ ) as FIN\_VM,VM\_ID,VM\_IDSEQ,VM.CONDR\_IDSEQ,CN.NAME CONCEPTS\_NAME

   from

   SBR.VALUE\_MEANINGS VM,

   CON\_DERIVATION\_RULES\_EXT CN

   where   VM.CONDR\_IDSEQ=CN.CONDR\_IDSEQ

   AND UPPER(ASL\_NAME) not like '%RETIRED%'

   AND instr(CN.NAME,'C45255')=0)

   where FIN\_VM<>VM\_ID

  Query B.

   select  FIN\_VM,vm\_id,VM\_IDSEQ,R.CONDR\_IDSEQ,CONCEPTS\_NAME

   from SBREXT.MDSR\_VM\_DUP\_REF R,

   SBREXT.CON\_DERIVATION\_RULES\_EXT CN

   where R.CONDR\_IDSEQ=CN.CONDR\_IDSEQ

   AND instr(CN.NAME,'C45255')=0

  (A MINUS B) will return duplicate VM records wich are not in SBREXT.MDSR\_VM\_DUP\_REF yet.

3.duplicate VMs with Concepts but Concepts Name like "integer".

  The records in these sets are grouped by Concepts Name and Concepts Value.

  Concepts Name is taken from SBREXT.CON\_DERIVATION\_RULES\_EXT table and Concepts Value from SBREXT.COMPONENT\_CONCEPTS\_EXT table,

  After the most current VM (Final VM) found  in the group, the older records inserted in to SBREXT.MDSR\_VM\_DUP\_REF

  with Final VM Public ID and VM\_IDSSEQ.

  Table SBREXT.COMPONENT\_CONCEPTS\_EXT might have many Concepts Value for one record in SBREXT.CON\_DERIVATION\_RULES\_EXT.

  In order to group records based on Concepts Name and Concepts Value we use aggregated value and instead of table SBREXT.COMPONENT\_CONCEPTS

  ,we are using following qoery.

  (SELECT CONDR\_IDSEQ, LISTAGG(CONCEPT\_VALUE,',') WITHIN GROUP (ORDER BY CONCEPT\_VALUE) as CONCEPT\_VALUE\_AG

  FROM  SBREXT.COMPONENT\_CONCEPTS\_EXT

  where CONCEPT\_VALUE is not null

  GROUP BY CONDR\_IDSEQ)CC

 The SP SBREXT.MDSR\_CREATE\_DUP\_VM\_DES\_DEF is created for:

1.(CURSOR  C1 -final VM and not processed records from duplicate sets).

    a. Set "RELEASED" workflow status for each Final VM.

    b. Add preferred definition with definition type= “Prior preferred definition” – if not in definitions type list.

 2.(CURSOR  C2- not final VM and not processed records from duplicate sets).

     1.Update SBR.VALUE\_MEANINGS records found matching in SBREXT.MDSR\_VM\_DUP\_REF table:

                a. Set "RETIRED ARCHIVED" workflow status.

                b. add to CHANGE\_NOTE 'Use VM public ID: FIN\_VM Version 1.0 instead. Modified by caDSR script.'

     2.Designation is created for each Retired VM for final VM as

    3.(CURSOR C\_DEF  all DESIGNATIONS for retired VM)

      DESIGNATIONS are created  for FINAL VM from existing DESIGNATIONS of retired VM if do not exist yet.

      select count(\*) into V\_cnt

      from SBR.DESIGNATIONS

      where AC\_IDSEQ=i.FIN\_IDSEQ and CONTE\_IDSEQ=n.conte\_idseq and trim(upper(NAME))=trim(upper(n.NAME))

      and trim(upper(DETL\_NAME))=trim(upper(n.DETL\_NAME)) and trim(upper(LAE\_NAME))=trim(upper(n.LAE\_NAME));

    4.(CURSOR C\_DES all DEFINITIONS for retired VM)

      DEFINITIONS are created  for FINAL VM from existing DEFINITIONS of retired VM if do not exist yet.

      select count(\*) into V\_cnt

      from SBR.DEFINITIONS

      where AC\_IDSEQ=i.FIN\_IDSEQ and CONTE\_IDSEQ=n.conte\_idseq

      and upper(trim(DEFINITION))=upper(trim(n.DEFINITION)) and

      trim(upper(DEFL\_NAME))=trim(upper(n.DEFL\_NAME)) and trim(upper(LAE\_NAME))=trim(upper(n.LAE\_NAME));

 The SP SBREXT.MDSR\_CREATE\_DUP\_VM\_CSI is created for:

 1.(CURSOR  C1 -retired VM CLASSIFICATIONS for DESIGNATIONS).

    Create CLASSIFICATIONS of DESIGNATIONS for FINAL VM from retired VM CLASSIFICATIONS of DESIGNATIONS

    avoiding duplicates in CLASSIFICATIONS set.

2.(CURSOR  C2 -retired VM CLASSIFICATIONS for DEFINITIONS).

   Create CLASSIFICATIONS of DEFINITIONS for FINAL VM from retired VM CLASSIFICATIONS of DEFINITIONS

    avoiding duplicates in CLASSIFICATIONS set

 The SP SBREXT.MDSR\_CHECK\_FVM\_DES\_DEF\_CSI is created for:

   1.To check if all DESIGNATIONS where created for FINAL VM  from retired VM DESIGNATIONS

   If any missed, the Procedure  set DES column to 'NC' in SBREXT.MDSR\_VM\_DUP\_REF

   2. To check if all DEFINITIONS where created for FINAL VM  from retired VM DEFINITIONS

   If any missed, the Procedure set DEFN column to 'NC' in SBREXT.MDSR\_VM\_DUP\_REF

   3. To check if all CLASSIFICATIONS of DESIGNATIONS where created for FINAL VM  from retired VM CLASSIFICATIONS of DESIGNATIONS

   If any missed, the Procedure set DES\_CL column to 'NC' in SBREXT.MDSR\_VM\_DUP\_REF

   4. To check if all CLASSIFICATIONS of DEFINITIONS where created for FINAL VM  from retired VM CLASSIFICATIONS of DEFINITIONS

    If any missed, the Procedure set DEFN\_CL column to 'NC' in SBREXT.MDSR\_VM\_DUP\_REF

   5.When all of 4 columns are NULL then column  PROC='P' .Which means the retired VM record and associated records in SBR.DESIGNATIONS,

   SBR.DEFINITIONS,SBREXT.AC\_ATT\_CSCSI\_EXT tables were added to FINAL VM if not existed.