Title: Specification of changes to the DLS Business System to ICAT 3.3.x link
Author: Shoaib Sufi
Company: Devigo
Version: 1.0
Related Task Id: 18
Date: 11 /07/2008

1 Introduction

This document covers the specification of changes to the DLS User Office System (DUO Desk) Link to ICAT.

Note this covers the data specification it does not cover the need for how error checking should work or how logging should be integrated. Please see Appendix A for notes on how implementation should proceed.

2 Rules for copying Data from DUO Desk to ICAT

The source columns are fully qualified table_name.column_name and they refer to tables and views inside DUO Desk. However if they are proceeded by ICAT then they refer to tables inside the ICAT schema instance.

The Destination columns are not fully qualified and refer to the table in the section heading which is inside the ICAT schema instance.

Conversion function specify &1, &2, &3 etc to match the first, second, third columns stated in the source columns column.

Database columns in the ICAT database that are not mentioned as targets for data in any destination can be ignored; they will either be set to null, a database default value or are updated using different mechanisms.

The process that does this mapping will be run multiple times so has to deal with inserts, update, optimisation and error processing the details of those are not specified here and looking at the existing system can aid in how the actual implementation should be structured please see Appendix A for further details.

Conditions for table joining are assumed and not stated.

Unless otherwise stated do not propagate any records across which pertain to failed investigation propagation attempts – i.e. if insertion into the INVESTIGATION table in ICAT failed then do not propagate associated keywords, shifts, publications, sample, sample_parameter and investigator. However there is an exception for facility users.

The batch migration job assumes an installed ICAT schema instance, data loaded from the spreadsheet and a link to the Diamond DUO Desk user office system.

The order of propagation should mimic the order of entities mentioned below i.e. following the numbering of 2.1 onwards below.

Records that are in ICAT that were sourced from DUO Desk but no longer existing in DUO Desk should be removed. If they cannot be removed this fact should be logged.

Records sourced from data in DUO Desk should only be inserted if they do not already exist in ICAT, and they should only be updated if any of the relevant columns specified below has actually changed.

2.1 INSTRUMENT

Condition 1: Instruments should only be copied across from DUO Desk into ICAT where they do not already exist in ICAT.

Source Columns	Conversion Function	Destination Column	Notes
INSTRUMENT.INSTR_NAME	Lower(&1)	NAME	
INSTRUMENT.INSTR_NAME	Lower(&1)	SHORT_NAME	
INSTRUMENT.INSTR_LIB		TYPE	
INSTRUMENT.INSTR_LIB		DESCRIPTION	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	If the records are updated then this column should not be modified as by definition it is set at create time.
		CREATE_ID set to name of process>	If the records are updated then this

			column should not be modified as by definition it is set at create time.
		FACILITY_ACQUIRED set to ‘Y’	

2.2 PARAMETER

2.2.1 From DUO Desk SAMPLE_PARAMETER

Condition 1: if the record already exists in ICAT set the is_sample_parameter to 'Y' if this is not already set and remember to update any of the audit columns that are relevant (i.e. mod_id, mod_time).

Condition 2: only attempt insert once for any instance of SAMPLE_PARAMETER.NAME in the source table

Source Columns	Conversion Function	Destination Column	Notes
SAMPLE_PARAMETER.NAME	Lower(&1)	NAME	
SAMPLE_PARAMETER.UNITS	NVL(&1,'N/A')	UNITS	
SAMPLE_PARAMETER.UNITS		UNITS_LONG_VERSION	
		SEARCHABLE set to 'Y'	
		NUMERIC_VALUE see Notes	If any of the SAMPLE_PARAMETER.VALUEs associated with this SAMPLE_PARAMETER.NAME do not match the Regular Expression '^-{0,1}\d*\.{0,1}\d+$' or the VALUE is NULL for all instances then this should be set to 'N' otherwise this should be set to 'Y'

		IS_SAMPLE_PARAMETER set to 'Y'	
		IS_DATASET_PARAMETER set to 'N'	On updates should not be modified
		IS_DATAFILE_PARAMETER set to 'N'	On updates should not be modified
SAMPLE_PARAMETER.COMMENTS		DESCRIPTION	
		VERIFIED set to 'Y'	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to 'Y'	

2.2.2 From Duo Desk SAMPLESHEET

Condition 1: if the record already exists in ICAT set the is_sample_parameter to 'Y' if this is not already set and remember to update any of the audit columns that are relevant (i.e. mod_id, mod_time).

Source Columns	Conversion Function	Destination Column	Notes
T.NAME	Lower(&1)	NAME	
	NVL(&1,'N/A')	UNITS set to 'text'	
		UNITS_LONG_VERSION set to 'text'	
		SEARCHABLE set to 'Y'	
T.VALUE	If any value of &1	NUMERIC_VALUE	

	does not match the Regular expression '^-{0,1}\d*\.{0,1}\d+$' or the &1 is NULL for all instances then this should be set to ‘N’ otherwise this can be set to ‘Y’		
		IS_SAMPLE_PARAMETER set to ‘Y’	
		IS_DATASET_PARAMETER set to ‘N’	On updates should not be modified
		IS_DATAFILE_PARAMETER set to ‘N’	On updates should not be modified
		VERIFIED set to ‘Y’	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to ‘Y’	

T is defined below:
T is a virtual table which needs defining as the names of columns in the SAMPLESHEET table are just not very good.

T.NAME	T.VALUE is value of
no	SAMPLESHEET.SMPS_NO
order_in_prop	SAMPLESHEET.SMPS_ORDER_IN_PROP
propsbm_no	SAMPLESHEET.SMPS_PROPSBM_NO
propos_no	SAMPLESHEET.SMPS_PROPOS_NO
description	SAMPLESHEET.SMPS_DESCRIPTION
acronym	SAMPLESHEET.SMPS_ACRONYM
is_crystal	SAMPLESHEET.SMPS_IS_CRYSTAL
is_powder	SAMPLESHEET.SMPS_IS_POWDER
is_solution	SAMPLESHEET.SMPS_IS_SOLUTION
concentration	SAMPLESHEET.SMPS_CONCENTRATION
source	SAMPLESHEET.SMPS_SOURCE
source_class	SAMPLESHEET.SMPS_SOURCE_CLASS
is_recombinant	SAMPLESHEET.SMPS_IS_RECOMBINANT
expr_host	SAMPLESHEET.SMPS_EXPR_HOST
exprhost_class	SAMPLESHEET.SMPS_EXPRHOST_CLASS
is_virus	SAMPLESHEET.SMPS_IS_VIRUS
is_toxin	SAMPLESHEET.SMPS_IS_TOXIN
is_prion	SAMPLESHEET.SMPS_IS_PRION
is_virfactor	SAMPLESHEET.SMPS_IS_VIRFACTOR
is_danger	SAMPLESHEET.SMPS_IS_DANGER
danger_txt	SAMPLESHEET.SMPS_DANGER_TXT
frozen	SAMPLESHEET.SMPS_FROZEN
capillary	SAMPLESHEET.SMPS_CAPILLARY
crystal_tray	SAMPLESHEET.SMPS_CRYSTAL_TRAY
other_holder	SAMPLESHEET.SMPS_OTHER_HOLDER
ligands	SAMPLESHEET.SMPS_LIGANDS

ligands_txt	SAMPLESHEET.SMPS_LIGANDS_TXT
laser	SAMPLESHEET.SMPS_LASER
laser_class	SAMPLESHEET.SMPS_LASER_CLASS
laser_wavelength	SAMPLESHEET.SMPS_LASER_WAVELENGTH
cooler	SAMPLESHEET.SMPS_COOLER
cryogenic_gas	SAMPLESHEET.SMPS_CRYOGENIC_GAS
pressurized_cell	SAMPLESHEET.SMPS_PRESSURIZED_CELL
propane	SAMPLESHEET.SMPS_PROPANE
danger_reception	SAMPLESHEET.SMPS_DANGER_RECEPTION
danger_reception_txt	SAMPLESHEET.SMPS_DANGER_RECEPTION_TXT
removed	SAMPLESHEET.SMPS_REMOVED
stored_esrf	SAMPLESHEET.SMPS_STORED_ESRF
user_name	SAMPLESHEET.SMPS_USER_NAME
user_email	SAMPLESHEET.SMPS_USER_EMAIL
user_phone	SAMPLESHEET.SMPS_USER_PHONE
opmode_no	SAMPLESHEET.SMPS_OPMODE_NO
opmode_date	SAMPLESHEET.SMPS_OPMODE_DATE
anom_scat_1	SAMPLESHEET.SMPS_ANOM_SCAT_1
anom_scat_2	SAMPLESHEET.SMPS_ANOM_SCAT_2
anom_scat_3	SAMPLESHEET.SMPS_ANOM_SCAT_3
anom_scat_4	SAMPLESHEET.SMPS_ANOM_SCAT_4
sci_justif	SAMPLESHEET.SMPS_SCI_JUSTIF
cell_a	SAMPLESHEET.SMPS_CELL_A
cell_b	SAMPLESHEET.SMPS_CELL_B
cell_c	SAMPLESHEET.SMPS_CELL_C
cell_alpha	SAMPLESHEET.SMPS_CELL_ALPHA
cell_beta	SAMPLESHEET.SMPS_CELL_BETA
cell_gamma	SAMPLESHEET.SMPS_CELL_GAMMA
space_group	SAMPLESHEET.SMPS_SPACE_GROUP
mad_ds	SAMPLESHEET.SMPS_MAD_DS

sad_ds	SAMPLESHEET.SMPS_SAD_DS
native_ds	SAMPLESHEET.SMPS_NATIVE_DS
comment	SAMPLESHEET.SMPS_COMMENT
bag_shift_type	SAMPLESHEET.SMPS_BAG_SHIFT_TYPE

2.3 INVESTIGATION

Condition 1: MEASURE.MES_UNI_ALL is greater then 0 indicating that this is an approved proposal with real resources allocated.

Condition 2: PLANNING.PL_DATE_DEB is not NULL and PLANNING.PL_DATE_FIN is not NULL

Condition 3: none of PROPOSAL.PROPOS_EFFACE, MEASURE.MES_EFFACE, INSTRUMENT.INSTR_EFFACE or PLANNING.PL_EFFACE are set to 'Y'

Note 1: Records that already exist in ICAT with matching INV_NUMBER, VISIT_ID, FACILITY_CYCLE, INSTRUMENT but SRC_HASH not set should have the SRC_HASH updated and this fact should be logged. (note this is different then what currently happens in 3.3.x – Appendix A – 2)

Note 2: Remove entries from ICAT which no longer exist in DUO Desk. However if it's not possible to remove the record in ICAT (e.g. as it has associated datasets) then this fact with the unique keys should be logged.

Source Columns	Conversion Function	Destination Column	Notes
		ID	On insert this is set to the ICAT sequence value of investigation_id_seq.nextval
PROPOSAL.PROPOS_NO	TO_CHAR(&1)	INV_NUMBER	
PROPOSAL.PROPOS_CATEG_CODE, PROPOSAL.PROPOS_CATEG_CPT, PLANNING.PL_VISIT_NO	&1||&2||'-'||&3	VISIT_ID	**Crucial column for DLS** – how they refer to experiments being performed.
		FACILITY set to the value of THIS_ICAT.FACILITY_SHORT_NAME in ICAT 3.3.x instance	

INSTRUMENT.INSTR_NOM	LOWER(&1)	INSTRUMENT	
PROPOSAL.PROPOS_TITLE, PROPOSAL.PROPOS_NO	NVL(SUBSTR(&1,1,255) , &2)	TITLE	
		INV_TYPE set to ‘experiment’	
DUO_PROPOSAL.EXP_ABSTRACT		INV_ABSTRACT	
		VERIFIED set to ‘Y’	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to ‘Y’	
PROPOSAL.PROPOS_NO, INSTRUMENT.INSTR_NO, PLANNING.PL_NO	GET_MD5(&1, &2, &3)	SRC_HASH	This column is used to store the primary keys of the 3 Duodesk tables which make up each ICAT Investigation record.

GET_MD5 is defined as:

```
FUNCTION get_md5 (
  p1  IN VARCHAR2,
  p2  IN VARCHAR2 DEFAULT NULL,
  p3  IN VARCHAR2 DEFAULT NULL,
  p4  IN VARCHAR2 DEFAULT NULL,
  p5  IN VARCHAR2 DEFAULT NULL,
  p6  IN VARCHAR2 DEFAULT NULL,
  p7  IN VARCHAR2 DEFAULT NULL,
```

```
  p8  IN VARCHAR2 DEFAULT NULL,
  p9  IN VARCHAR2 DEFAULT NULL,
  p10 IN VARCHAR2 DEFAULT NULL,
  p11 IN VARCHAR2 DEFAULT NULL,
  p12 IN VARCHAR2 DEFAULT NULL
  ) RETURN RAW DETERMINISTIC IS
BEGIN
  RETURN dbms_obfuscation_toolkit.md5(
    input =>
      utl_raw.cast_to_raw(
        p1||'-'||p2||'-'||p3||'-'||p4||'-'||p5||'-'||p6||'-'||
        p7||'-'||p8||'-'||p9||'-'||p10||'-'||p11||'-'||p12));
END get_md5;
```

2.4 SHIFTS

Condition 1: MEASURE.MES_UNI_ALL is greater then 0 indicating that this is an approved proposal with real resources allocated.

Condition 2: PLANNING.PL_DATE_DEB is not NULL and PLANNING.PL_DATE_FIN is not NULL

Condition 3: none of PROPOSAL.PROPOS_EFFACE, MEASURE.MES_EFFACE, INSTRUMENT.INSTR_EFFACE or PLANNING.PL_EFFACE are set to 'Y'

Note: when doing joins need to join icat.investigation with keys in duo (done via sources of unique keys) so that in one query we can find the correct investigation.id for populating the shift table.

Source Columns	Conversion Function	Destination Column	Notes
ICAT.INVESTIGATION.ID		INVESTIGATION_ID	
PLANNING.PL_DATE_DEB, PLANNING.PL_SHIFTS_DEB	SHIFT_TIME(&1,&2)	START_DATE	
PLANNING.PL_DATE_FIN, PLANNING.PL_SHIFTS_FIN	SHIFT_TIME(&1,&2)	END_DATE	
PLANNING.PL_COM		SHIFT_COMMENT	
		VERIFIED set to 'Y'	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified

		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to ‘Y’	

SHIFT_TIME notes:

Is used to generate the timestamp from the junction of the shift date and time slot parameters:
a = shift date
b = shift slot or time. there are three shift slot identified by the number 1,2,3
where 1 = 01:00; 2 = 09:00; 3 = 17:00.

SHIFT_TIME defined (in PL/SQL) as:

```
FUNCTION shift_time (a DATE, b NUMBER)
   RETURN DATE
AS
   shift   CHAR (5);
BEGIN
   CASE b
      WHEN 1
      THEN
         shift := '01:00';
      WHEN 2
      THEN
         shift := '09:00';
      WHEN 3
      THEN
         shift := '17:00';
      ELSE
         shift := '09:00';
   END CASE;
```

```
    RETURN TO_DATE (TO_CHAR (a, 'DD/MM/YYYY') || ' ' || shift,
                    'DD/MM/YYYY HH24:MI'
                   );
END;
```

2.5 FACILITY_USER

Note 1: Records are copied across even if there are failures in associated investigations.

Note 2: TBLPEOPLE has no unique key so we have to have to order the records for a person with federal id's having 'nulls' last to choose the ones where federal id is set. Also as the names for an individual might differ choosing only one TBLPEOPLE entry with the same federal id is required.

Source Columns	Conversion Function	Destination Column	Notes
TBLPEOPLE.USERNUMBER	TO_CHAR(&1)	FACILITY_USER_ID	
TBLPEOPLE.FEDID		FEDERAL_ID	
TBLPEOPLE.TITLE		TITLE	
TBLPEOPLE.INITIALS		INITIALS	
TBLPEOPLE.KNOWNAS		FIRST_NAME	
TBLPEOPLE.FAMILYNAME		LAST_NAME	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED	

		set to ‘Y’	

2.6 INVESTIGATOR

Condition 1: Find all the investigations in ICAT which match proposals in DUO Desk based on visit id and where the federal id of the investigator is set from. This will require joining ICAT.FACILITY_USER, DUODESK.TBLPEOPLE, DUODESK.PROPOSAL and DUODESK.PROPOSALSC to make linkages to extract all relevant data. Any links between INVESTIGATION and FACILITY_USER related to investigations which have not been able to be migrated in full need to be suppressed also.

Note 1: The source and destination of data is quite straightforward how these are derived however is complicated (please see Appendix A on how best to prceed note ROLE is specified differently in this document than in either 3.1.4 or 3.3.x).

Source Columns	Conversion Function	Destination Column	Notes
ICAT.INVESTIGATION.ID		INVESTIGATION_ID	
ICAT.FACILITY_USER.FACILITY_USER_ID		FACILITY_USER_ID	
		ROLE set to 'principal_experimenter'	The roles in Duo Desk are just not fine grained enough – e.g. normal_user, advanced_user, DLS_staff. They give the user a role at the facility but not a per investigation role so if this is not hard wired then the triggers won't work and if we don't set all the investigators to 'principal_experimentor' the actual principal will

			lose the ability to add users using the ICAT API.
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to ‘Y’	

2.7 KEYWORDS

Constraint 1: PROPOSAL.PROPOS_EFFACE must not be set to Y

Constraint 2: LOWER (T.COLUMN_VALUE) should have the following stopwords[1] filtered out:

I, a, about, an, are, as, at, be, by, etc, for, from, how, in, is, it, of, on, or, that, the, this, to, was, what, when, where, who, will, with, the

Constraint 3: for each T.COLUMN_VALUE added they must be unique in the same case – i.e. no duplicate keywords are permitted (case is significant).

Note 1: Keywords for investigations should be deleted if they no longer have a match in Duo Desk and they were created by the propagation process.

Source Columns	Conversion Function	Destination Column	Notes
ICAT.INVESTIGATION.ID		INVESTIGATION_ID	
T.COLUMN_VALUE		NAME	T is defined below.
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be

[1] http://en.wikipedia.org/wiki/Stop_words

			modified
		FACILITY_ACQUIRED set to ‘Y’	

Definition of T:

```
TABLE (CAST (split_at_whitespace (PROPOSAL.PROPOS_TITLE) AS vc_array) )
```

Definition of Function split_at_whitespace:

```
-- specialist function to convert a whitespace-delimited string into a
-- collection
FUNCTION split_at_whitespace(
  p_text IN VARCHAR2)
  RETURN vc_array IS
BEGIN
  -- reduce all consecutive whitespace characters to single spaces

  RETURN string_to_table(
    REGEXP_REPLACE(RTrim(p_text),'[[:space:]]+',' '), ' ');
END split_at_whitespace;
```

2.8 PUBLICATIONS

Constraint 1: T.COLUMN value is not null

Constraint 2: DUO_PROPOSAL.DESK_PROPOS_NO in DUO DESK matches INVESTIGATION.INV_NUMBER in ICAT.

Source Columns	Conversion Function	Destination Column	Notes
		ID	On insert this is set to the ICAT sequence value of publication_id_seq.nextval
ICAT.INVESTIGATION.ID		INVESTIGATION_ID	
T.COLUMN_VALUE	SUBSTR(&1,1,4000)	FULL_REFERENCE	T is Defined below
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to 'Y'	

Definition of T:

TABLE (CAST (string_to_table (RTRIM (duo_proposal.exp_publications), CHR (10)) AS vc_array))

Definition of function string to table:

```
-- converts a delimited string into a collection, using any delimiter
FUNCTION string_to_table(
  p_text IN VARCHAR2,
  p_delimiter IN VARCHAR2 DEFAULT ',')
  RETURN vc_array IS

  lv_text VARCHAR2(4000);
  lv_num PLS_INTEGER;
  retval vc_array := vc_array();
BEGIN
  lv_text := p_text || p_delimiter;

  LOOP
    lv_num := InStr(lv_text, p_delimiter);
    EXIT WHEN (Nvl(lv_num, 0) = 0);
    retval.extend;
    retval(retval.Count) := LTrim(RTrim(SubStr(lv_text, 1, lv_num - 1)));
    lv_text := SubStr(lv_text, lv_num + 1);
  END LOOP;

  RETURN retval;
END string_to_table;
```

2.9 SAMPLE

Constraint 1: SAMPLE.NAME is not null.

Source Columns	Conversion Function	Destination Column	Notes
		ID	On insert this is set to the ICAT sequence value of sample_id_seq.nextval
ICAT.INVESTIGATION.ID		INVESTIGATION_ID	
SAMPLE.NAME		NAME	
SAMPLE.CHEMICAL_FORMULA		CHEMICAL_FORMULA	
		SAFETY_INFORMATION set to 'See sample parameters'	
SAMPLE.ID		PROPOSAL_SAMPLE_ID	
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to 'Y'	

2.9.1 Samples Updated Locally in ICAT vs. Information from DUO Desk

Samples can be updated/inserted/deleted and could be updated directly in ICATDLS. This leads to a situation where there are not one but two sources of input to the table.However due to the presence of both the MOD_ID and CREATE_ID this makes determining allowable updates easier.

If a Sample entry was created by the process which feeds in from DUO Desk and any local changes are made to it then these local changes will be lost as the DUO Desk to ICAT link takes precedence.

If a Sample entry was created directly in ICAT and subsequently there is a conflict with data coming from the DUO Desk system then any updates or creation of data should fail and log what is happening so it can be manually resolved.

2.10 SAMPLE_PARAMETER

If there are any conflicts between entries added in 2.10.1 and 2.10.2 then the latter should take precedence.

2.10.1 General Sample Parameters

Constraint 1: SAMPLE_PARAMETER.NAME and SAMPLE_PARAMETER.VALUE are both not null.

Note 1: The values in SAMPLE_PARAMETER.NAME no longer match what is in 3.3.x use the specification below.

Source Columns	Conversion Function	Destination Column	Notes
ICAT.SAMPLE.ID		SAMPLE_ID	
SAMPLE_PARAMETER.NAME	LOWER(&1)	NAME	
SAMPLE_PARAMETER.UNITS	NVL(&1,'N/A')	UNITS	
		STRING_VALUE	If the associated (through name and units in the parameter table) ICAT.PARAMETER.NUMERIC_VALUE **is not** 'Y' then this is set to SAMPLE_PARAMETER.VALUE
		NUMERIC_VALUE	If the associated ICAT.PARAMETER.NUMERIC_VALUE **is** set to 'Y' then this is set to SAMPLE_PARAMETER.VALUE
SAMPLE_PARAMETER.ERROR		ERROR	
SAMPLE_PARAMETER.RANGE_TOP		RANGE_TOP	
SAMPLE_PARAMETER.RANGE_BOTTOM		RANGE_BOTTOM	

		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to 'Y'	

2.10.2Safety Related Sample Parameters

Constraint 1: T.NAME and T.VALUE are both not null.

Source Columns	Conversion Function	Destination Column	Notes
ICAT.SAMPLE.ID		SAMPLE_ID	
T.NAME		NAME	T specified below
		UNITS set to 'text'	T specified below
		STRING_VALUE	If the associated (through name and units in the parameter table) ICAT.PARAMETER.NUMERIC_VALUE **is not** 'Y' then this is set to value of

			T.VALUE
		NUMERIC_VALUE	If the associated ICAT.PARAMETER.NUMERIC_VALUE **is** set to 'Y' then this is set to value of T.VALUE
		MOD_TIME set to systimestamp	
		MOD_ID set to <name of process>	
		CREATE_TIME set to systimestamp	On updates should not be modified
		CREATE_ID set to name of process>	On updates should not be modified
		FACILITY_ACQUIRED set to 'Y'	

T specification:

T is a virtual table which needs defining as the names of columns in the SAMPLESHEET table are just not very good.

T.NAME is	**T.VALUE is value of**
no	SAMPLESHEET.SMPS_NO
order_in_prop	SAMPLESHEET.SMPS_ORDER_IN_PROP
propsbm_no	SAMPLESHEET.SMPS_PROPSBM_NO
propos_no	SAMPLESHEET.SMPS_PROPOS_NO
description	SAMPLESHEET.SMPS_DESCRIPTION
acronym	SAMPLESHEET.SMPS_ACRONYM
is_crystal	SAMPLESHEET.SMPS_IS_CRYSTAL
is_powder	SAMPLESHEET.SMPS_IS_POWDER
is_solution	SAMPLESHEET.SMPS_IS_SOLUTION
concentration	SAMPLESHEET.SMPS_CONCENTRATION
source	SAMPLESHEET.SMPS_SOURCE
source_class	SAMPLESHEET.SMPS_SOURCE_CLASS
is_recombinant	SAMPLESHEET.SMPS_IS_RECOMBINANT
expr_host	SAMPLESHEET.SMPS_EXPR_HOST

exprhost_class	SAMPLESHEET.SMPS_EXPRHOST_CLASS
is_virus	SAMPLESHEET.SMPS_IS_VIRUS
is_toxin	SAMPLESHEET.SMPS_IS_TOXIN
is_prion	SAMPLESHEET.SMPS_IS_PRION
is_virfactor	SAMPLESHEET.SMPS_IS_VIRFACTOR
is_danger	SAMPLESHEET.SMPS_IS_DANGER
danger_txt	SAMPLESHEET.SMPS_DANGER_TXT
frozen	SAMPLESHEET.SMPS_FROZEN
capillary	SAMPLESHEET.SMPS_CAPILLARY
crystal_tray	SAMPLESHEET.SMPS_CRYSTAL_TRAY
other_holder	SAMPLESHEET.SMPS_OTHER_HOLDER
ligands	SAMPLESHEET.SMPS_LIGANDS
ligands_txt	SAMPLESHEET.SMPS_LIGANDS_TXT
laser	SAMPLESHEET.SMPS_LASER
laser_class	SAMPLESHEET.SMPS_LASER_CLASS
laser_wavelength	SAMPLESHEET.SMPS_LASER_WAVELENGTH
cooler	SAMPLESHEET.SMPS_COOLER
cryogenic_gas	SAMPLESHEET.SMPS_CRYOGENIC_GAS
pressurized_cell	SAMPLESHEET.SMPS_PRESSURIZED_CELL
propane	SAMPLESHEET.SMPS_PROPANE
danger_reception	SAMPLESHEET.SMPS_DANGER_RECEPTION
danger_reception_txt	SAMPLESHEET.SMPS_DANGER_RECEPTION_TXT
removed	SAMPLESHEET.SMPS_REMOVED
stored_esrf	SAMPLESHEET.SMPS_STORED_ESRF
user_name	SAMPLESHEET.SMPS_USER_NAME
user_email	SAMPLESHEET.SMPS_USER_EMAIL
user_phone	SAMPLESHEET.SMPS_USER_PHONE
opmode_no	SAMPLESHEET.SMPS_OPMODE_NO
opmode_date	SAMPLESHEET.SMPS_OPMODE_DATE
anom_scat_1	SAMPLESHEET.SMPS_ANOM_SCAT_1

anom_scat_2	SAMPLESHEET.SMPS_ANOM_SCAT_2
anom_scat_3	SAMPLESHEET.SMPS_ANOM_SCAT_3
anom_scat_4	SAMPLESHEET.SMPS_ANOM_SCAT_4
sci_justif	SAMPLESHEET.SMPS_SCI_JUSTIF
cell_a	SAMPLESHEET.SMPS_CELL_A
cell_b	SAMPLESHEET.SMPS_CELL_B
cell_c	SAMPLESHEET.SMPS_CELL_C
cell_alpha	SAMPLESHEET.SMPS_CELL_ALPHA
cell_beta	SAMPLESHEET.SMPS_CELL_BETA
cell_gamma	SAMPLESHEET.SMPS_CELL_GAMMA
space_group	SAMPLESHEET.SMPS_SPACE_GROUP
mad_ds	SAMPLESHEET.SMPS_MAD_DS
sad_ds	SAMPLESHEET.SMPS_SAD_DS
native_ds	SAMPLESHEET.SMPS_NATIVE_DS
comment	SAMPLESHEET.SMPS_COMMENT
bag_shift_type	SAMPLESHEET.SMPS_BAG_SHIFT_TYPE

Note 1: Any of the above T.NAME entries that are not in the facilities spreadsheet as valid parameters for sample for DLS should be added.

Appendix A – How to proceed with implementation

1. The version of the batch migration package (BMP) originally rolled out for DLS was 3.1.0 however due to various issues this has been modified and currently the one in operation on the production ICATDLS 3.1.x instance is 3.1.4 and available here: https://esc-cvs.dl.ac.uk/svn/dl/metadata/icat/branches/3.1.4/propagation/install_scripts/schema_specific_scripts/icatdls/2_after_common_objects/020_icatdls_batch_migration_pkg.sql

2. However after the 3.1.0 release work started on the 3.2.x line of the BMP and this has significant differences in its implementation details. Primarily this supported separate create_id and mod_id, addition of src_hash to the investigation table to support more definite and efficient look ups of source data in DUO Desk. The latest version (which was modified less recently then the 3.1.4 version) is available at: https://esc-cvs.dl.ac.uk/svn/dl/metadata/icat/trunk/propagation/install_scripts/schema_specific_scripts/icatdls/2_after_common_objects/020_icatdls_batch_migration_pkg.sql

3. In addition to this there was a new view called SAMPLESHEET which we were advised by Bill Pulford (of DLS) of using as a source for populating sample parameter information inside ICAT and parameter information.

So how best to proceed:

Take 2. above, add any relevant changes made to 1 (Carmine will now what these are I think they are around working out accurate SHIFT times but there maybe others) above if still applicable and add the 3. SAMPLSHEET source of data as specified in 2.10.2 above. Then check that all the caveats and data propagation rules mentioned in this document are met. That would be a sensible way to proceed on implementation in a timely fashion. Note the populate_icat_authorisation function in 2. is no longer needed as this is taken care of by triggers. Also note that the way visit_id is sourced in migrate_investigations in 2. is wrong it should be derived the way it is in 1.