ICAT 3.3 Data Dictionary

Shoaib Sufi
Devigo,
Daresbury Science and Innovation Centre,
`shoaib.sufi@devigo.com`

July 31, 2008

Abstract

This document is the data dictionary for ICAT 3.3.x; explaining the purpose of the tables and columns used in the schema.

Chapter 1

Introduction

This document is created from the comments written during the development of the ICAT 3.3.x schema. The ICAT 3.3.x schema was developed using Oracle JDeveloper; the JDeveloper project files contain data definition, contraint definition, index definition, comments and a schema diagram. The JDeveloper project files for ICAT 3.3.x are available at time of writing from https://esc-cvs.dl.ac.uk/svn/dl/metadata/icat/trunk/jdeveloper/icat, please contact STFC database services for access to these files by e-mailing databaseservices@stfc.ac.uk.

Chapter 2

Tables

2.1 APPLICATIONS

Stores descriptions of the Oracle Application Express applications which use this schema. The table USER_ROLES is linked to APPLICATIONS. In future version of ICAT these tables will be removed as applications interact with ICAT via the ICAT API.

Column Name	Comments
APP_CODE	Unique identifier for this application in the system.
APP_NAME	Name of the application.
APP_DESCRIPTION	Description of the application.

2.2 DATAFILE

Information about the datafiles associated with this facility. This table stores information about logical names and physical locations. Some parameters (e.g. create time and fixity information) are also stored however the majority of this information is now in the DATAFILE_PARAMETER table e.g. the uA hours value for datafiles which gave the ISIS scientists an idea of how much data was collected has been moved to associated entries in the DATAFILE_PARAMETER table.

Column Name	Comments
ID	Key
DATASET_ID	Key of parent dataset.
NAME	Logical name of datafile.
DESCRIPTION	Description of contents, mechanism of capture or other relevant details pertaining to the datafile. It is expected that the topics covered by the description will be consistent across datafiles collected at a particular facility.
DATAFILE_VERSION	There are many situations in which a collected file may need to be replaced by a newer version. For example if the software creating the datafiles is incorrect or being fed incorrect data (incorrect algorithms or stale sample information) then erroneous data could be present in the final data file. Correction of the datafile due to these type of errors could then be recorded and a new version registered.
DATAFILE_VERSION_COMMENT	Describes why there was a new version of the datafile and what has changed. If the change which caused a new version of the file be produced is relevant for user and their data analysis they can then choose to update their files to the newer versions.
LOCATION	The physical location of the file. This will be facility specific e.g. in the case of Diamond Light Source this would be an SRB schemed URL. The locations are expected to be fully qualified but they could be relative if the parent dataset location is specified, again this should be consistent across all the datafiles associated with a facility.
DATAFILE_FORMAT	Format of the file. Permitted values are specified in DATAFILE_FORMAT.
DATAFILE_FORMAT_VERSION	This can be removed in future versions of ICAT as the data already exists in DATAFILE_FORMAT.
DATAFILE_CREATE_TIME	Useful parameter in the datafile which can be used for various things. It allows one to search for all files which were modified after they were created to solve consistency issues. Can be used to calculate the actual extent in of any experiment.
DATAFILE_MODIFY_TIME	Used with DATAFILE_CREATE_TIME.
FILE_SIZE	This is the actual size in bytes of the file (as oppose to the size on a storage device which maybe greater due to block size issues). This column should be a parameter but it is needed often and thus is present at this level.
COMMAND	This stores the command line used to create this file. This is primarily seen as related to datafiles which have come out of data analysis work. Also it is present here and not in the RELATED_DATAFILES table as the source datafile may not be held in the catalog, but be externally referenced. This could be an area where registration of dependencies in data analysis is seen as best practice.
CHECKSUM	In later releases of ICAT this piece of fixity information should be moved to a DATAFILE_PARAMETER and the checksum_standard should also be captured.
SIGNATURE	In later releases of ICAT this piece of fixity information should be moved to a DATAFILE_PARAMETER and the signature_standard should also be captured.

2.3   DATAFILE_FORMAT

Holds the valid datafile formats that are permitted in DATAFILE.

Column Name	Comments
NAME	Name of the format.
VERSION	Version of the format.
FORMAT_TYPE	The type of the format e.g. binary, HDF5, XML.
DESCRIPTION	A description of the format relating to the type of data that it holds, this could be a link to more information about the format also.

2.4   DATAFILE_PARAMETER

Stores name-value pairs of metadata associated with datafiles. The type of names are constrained by their entry in the PARAMETER table. This feature aids building up controlled vocabularies for a particular facility.

Column Name	Comments
DATAFILE_ID	Key of the related datafile.
NAME	Name of the parameter.
UNITS	Units of the parameter.
STRING_VALUE	If the DATAFILE_PARAMETER.NAME field has a value expressed as a string then the data will be present here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
NUMERIC_VALUE	If the DATAFILE_PARAMETER.NAME field has units expressed as a number then the value will be set here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
RANGE_TOP	If the value is a range this holds the maximum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
RANGE_BOTTOM	If the value is a range this holds the minimum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
ERROR	Holds the error range for the STRING_VALUE, NUMERIC_VALUE or RANGE_TOP and RANGE_BOTTOM combination.
DESCRIPTION	Where and how the parameter was recorded or extracted as oppose to a definition of the parameter; the latter being defined in the PARAMETER table.

2.5   DATASET

Groups a set of files. Illustrative groupings include pre-experimental data including electronic copies of the proposal or simulation data, experimental data, analysed data and final data.

Column Name	Comments
ID	None Available
SAMPLE_ID	Key to the related sample from the SAMPLE table. This can be unset as calibration datasets will not have a sample. As an aside it should be noted that data in the SAMPLE table is often sourced from the Proposal and that what people propose and what they bring maybe different so some process of manual reconciliation may be needed as this is not always possible to automate.
INVESTIGATION_ID	Key of the related INVESTIGATION.
NAME	The name of the dataset. This naming convention depends upon facility specific policies and is a function of how data is registered with the ICAT.
DATASET_TYPE	The rules governing what dataset type is given for a particular investigation type are facility specific. This may be an example of a business rule which would be of use in a rule engine. An Example of the rule is the context of the facilities at STFC is: If INVESTIGATION.INV_TYPE == EXPERIMENT or INVESTIGATION.INV_TYPE == CALIBRATION then dataset type could be any of the following: 1. PRE_EXPERIMENT_DATA 2. DETECTOR_CALIBRATION (one detector) 3. EXPERIMENT_CALIBRATION (with a calibration sample) 4. EXPERIMENT_RAW 5. LASER_SHOT (CLF specific) 6. LASER_DIAGNOSTICS (CLF specific) 7. TARGET_DATA 8. SIMULATION 9. ANALYSIS
DATASET_STATUS	This show the status of the data collection. Examples of values are empty, ongoing (e.g. CLF waiting for the glass plate to be analysed) or complete.
LOCATION	This is the location or root designation of the dataset. This is useful for specifying all the files in a directory, or for giving a physical root designation to nested files locations specified in the flat datafile table to allow applications to display and download the nested structures.
DESCRIPTION	This give a description of the dataset. This could hold a variety of information or could be empty. An example of use is storage of a description of the data collected at a facility filled in by data acquisition systems e.g. important parameters (e.g. temperature might change a few degrees, but the angle of the sample might be very important so this is the place such information should be stored) or this could be used to describe how the data was 'cut'; in the case of CLF they cut by shot - i.e. one dataset per shot which is important information for understanding the data organisation. Also when DATASET_TYPE is PRE_EXPERIMENT_DATE this field could be used to describe why certain files have been added and what their purpose is (e.g. simulation files).

2.6 DATASET_PARAMETER

Used to store parameters which are relevant to the entire dataset. For example, it could be used to hold instrument configuration parameters relevant to this dataset (more relevant for CLF but as an aside in ISIS instrument configuration usually only happens once per cycle, so this might be attached to datasets associated with a calibration investigation).

Column Name	Comments
DATASET_ID	Key of related dataset.
NAME	Name of the parameter.
UNITS	Units of the parameter, N/A when no Unit applies.
STRING_VALUE	If this name has a value expressed as a string then the data will be present here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
NUMERIC_VALUE	If this name has units expressed as a number then the value will be set here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
RANGE_TOP	If the value is a range this holds the maximum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
RANGE_BOTTOM	If the value is a range this holds the minimum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
ERROR	Holds the error range for the STRING_VALUE, NUMERIC_VALUE or RANGE_TOP and RANGE_BOTTOM combination.
DESCRIPTION	Where and how the parameter was recorded or extracted as oppose to a definition of the parameter; the latter being defined in the PARAMETER table.

2.7 DATASET_STATUS

Holds the status of datafiles; for example are they empty and therefore placeholders, being added to or complete.

Column Name	Comments
NAME	A meaningful status name. In future versions of ICAT this could be changed to STATUS.
DESCRIPTION	A description of what the status term means, for example if this is empty does it mean that this is a placeholder for data or that no data will ever be added and this is a tag of some kind.

2.8 DATASET_TYPE

Holds the valid set of dataset types.

Column Name	Comments
NAME	Designates a particular dataset type e.g. experiment_raw.
DESCRIPTION	A description of the dataset type e.g. raw data collected at the facility during an experiment.

2.9 FACILITY_CYCLE

Designates the time between any two shutdown periods for a facility i.e. active time when calibrations and experiments are being done.

Column Name	Comments
NAME	Facility specific designator of a cycle. Facilities have systematic naming formats for this e.g. ISIS use the start and end dates for creating cycle names. In terms of data organisation, facilities have found cycles a useful way of delineating their data.
START_DATE	Start date of the cycle.
FINISH_DATE	Finish date of the cycle.
DESCRIPTION	A description of the cycle including any noteworthy events relevant for understanding the data archive state.

2.10 FACILITY_INSTRUMENT_SCIENTIST

Holds information pertaining to facility designated authorities with regards to particular user scheduled experimental equipment. For example, for ISIS this would be instruments, for Diamond this would be beamlines and for CLF this would be target areas. These designated authorities have privilege in the ICAT system to access to all the data collected at their instruments.

Column Name	Comments
INSTRUMENT_NAME	The facility name of the device from which data is being collected e.g. Instrument name, Beamline name or Target area name for the STFC facilities.
FEDERAL_ID	The organisational identifier associated with the designated authority. For example in the case of STFC this is the Active Directory user name that the individual uses to login to corporate services.

2.11 FACILITY_USER

Holds information pertaining to people associated with investigations whether proposers or experimenters. It also is associates facility issued user identifiers with corporate identifiers of the organisational context within which the facilities operate. The former is used to register proposals and associated data and the latter used to access services offered by the organisation to access that data for example.

Column Name	Comments
FEDERAL_ID	FEDERAL_ID should be self consistent across the database and usually refers to a user numbering system which is valid across a (virtual or real) organisation e.g. the CICT issued Federal Identifiers at STFC. Thus it is generalised as a string as this can accommodate a numbering system also.
TITLE	The TITLE of the individual (for example Mr, Mrs, Ms, Miss, Professor, Dr.) This information is often taken from facility user database or business systems.
INITIALS	Initials of the user often taken from the facility business system.
FIRST_NAME	The first name of the facility user.
MIDDLE_NAME	Middle name if any of the facility user. In future versions of ICAT this should probably be changed to MIDDLE_NAMES i.e. plural.
LAST_NAME	Last name of the facility user

2.12   ICAT_AUTHORISATION

Contains authorisation information pertaining to user and their roles.

Column Name	Comments
ID	Key of the ICAT_AUTHORISATION entry.
USER_ID	This is the user identifier used for authentication with the ICAT system - e.g. federal id or 'ANY' for public or an context designation (e.g. <facility>_GUARDIAN for the process which adds data on behalf of the facility or SUPER for the super user).
ROLE	ROLE from the ICAT_ROLE table.
USER_CHILD_RECORD	This is an optimisation. This connects a child record where the ELEMENT_ID is null and ELEMENT_TYPE is DATASET indicating that the user in that case can create datasets in the investigation with the actual parent investigation - this makes it easier for the ICAT API to quickly check if a user has create privileges on a particular investigation.
ELEMENT_TYPE	Can be INVESTIGATION or DATASET.
ELEMENT_ID	If null then this has special meaning please see the authorisation specification for ICAT 3. Otherwise this is the key of the ELEMENT_TYPE inside ICAT.
PARENT_ELEMENT_TYPE	Needed if element_type is a dataset to give context to the parent investigation.
PARENT_ELEMENT_ID	Key of parent element for this record.

2.13   ICAT_ROLE

This information specifies the individual actions that a ROLE can perform. A ROLE represents a set of permitted actions. These actions should be reflected in the applications which interact with ICAT, e.g. the ICAT API. For definitive information about the authorisation system please refer to the ICAT authorisation guide in the documentation.

Column Name	Comments
ROLE	The Role name.
ROLE_WEIGHT	Calculated from the combined weights given for the actions a role can perform. This is meant to allow an easy way of working out a hierarchy of authority.
ACTION_INSERT	Specifies the ability to create metadata in the ICAT.
ACTION_INSERT_WEIGHT	Weight associated with this action.
ACTION_SELECT	Specifies the ability to search metadata in the ICAT.
ACTION_SELECT_WEIGHT	Weight associated with this action.
ACTION_DOWNLOAD	Specifies the ability to download data linked to in the metadata in the ICAT.
ACTION_DOWNLOAD_WEIGHT	Weight associated with this action.
ACTION_UPDATE	Specifies the ability to update metadata in the ICAT.
ACTION_UPDATE_WEIGHT	Weight associated with this action.
ACTION_DELETE	Specifies the ability to mark as deleted metadata in the ICAT.
ACTION_DELETE_WEIGHT	Weight associated with this action.
ACTION_REMOVE	Specifies the ability to remove metadata in the ICAT.
ACTION_REMOVE_WEIGHT	Weight associated with this action.
ACTION_ROOT_INSERT	Specifies the ability to create roots of metadata hierarchies in the ICAT, i.e. investigations in this version of ICAT.
ACTION_ROOT_INSERT_WEIGHT	Weight associated with this action.
ACTION_ROOT_REMOVE	Specifies the ability to remove roots of metadata hierarchies in the ICAT, i.e. investigations in this version of ICAT.
ACTION_ROOT_REMOVE_WEIGHT	Weight associated with this action.
ACTION_SET_FA	Specifies the ability to set the facility acquired flag in ICAT; this flag has special features and can override other actions and disallow them. More information is available in the ICAT authorisation specification.
ACTION_SET_FA_WEIGHT	Weight associated with this action.
ACTION_MANAGE_USERS	Specifies the ability to manage users in ICAT, i.e. adding and deleting them from investigations and changing their privileges.
ACTION_MANAGE_USERS_WEIGHT	Weight associated with this action.
ACTION_SUPER	Specifies the ability to do anything you might do with direct access to the schema. However the set of abilities one can do with this user could make the database inconsistent so caution in use of this user is advised. For example, an issue was found with registering files in ISIS which caused an error which was the correct behaviour as otherwise files would have been mis-cataloged if the user had ACTION_SUPER. This example supports the reason for having the <facility>_GUARDIAN user for data registration.
ACTION_SUPER_WEIGHT	Weight associated with this action.

2.14   INSTRUMENT

The name of this table should be changed in future version of ICAT to something more generic. As this table lists the name for the facility designated device which is normally associated with data collection.

Column Name	Comments
NAME	The name of the instrument.
SHORT_NAME	Needed as often (e.g. ISIS) instrument short names cannot be generated automatically from long names e.g. CRISP (CSP), HRPD (HRP), PRISMA (PRS).
TYPE	The type or classification of the device used for data collection such that similar facilities could understand what the instrument does from the type. E.g. at ISIS HRPD is a Powder Diffractometer, other Neutron sources also have Powder Diffractometers, but the specific name of their device would be different.
DESCRIPTION	A longer description or link to longer description of the device used for data collection.

2.15   INVESTIGATION

Holds information about investigations (i.e. experiments, calibrations etc). How investigations (more specifically experiments) are derived from proposals is facility specific. For example, in ISIS there is mainly a one to one correspondence, whereas in Diamond a proposal often maps onto many investigations.

Column Name	Comments
ID	Key for the table referred to in other tables as INVESTIGATION_ID.
INV_NUMBER	This is the experiment number e.g. the RB number from ISIS. In the case of ISIS, this is usually derived from the proposal number, but proposals can be split into separate experiments each with their own number, in which case it actually maps to the approved proposal number.
VISIT_ID	Sometimes (e.g. in the case of DLS) investigations are consortium based, i.e. carried out by a range of people from different institutions who manage their own time slot on the instrument. So in effect the investigation is multi-faceted where different groups should not have access to the data from other groups. Thus a visit identifier is required to differentiate the different groups in the consortium based proposal and the experiments that they perform. In the case of ISIS and CLF this can be set to null or a constant value.
FACILITY	Derived from THIS_ICAT.FACILITY_SHORT_NAME often needed in front end applications and for differentiating results in multiple ICAT queries and so is here as an optimisation.
INSTRUMENT	Multiple instruments per approved proposal are different investigations with different instrument inside ICAT. Feature requested by DLS and ISIS, the other solutions was to attach the instrument at the dataset level but this was seen as problematic due in part to common searches being in terms of instrument and experiment number as oppose to title and drilling down further.

TITLE	The proposal is usually the source of the INVESTIGATION.TITLE however this could be modified by the facility to reflect more accurately the real experiment being performed as oppose to the one specified in the proposal.
INV_TYPE	Valid value will be from amongst the ones available from INVESTIGATION_TYPE.NAME_VALUES e.g. common ones will be experiment or calibration. Note - as calibrations can be ad hoc not just at the beginning of a cycle but also when an instrument is fixed then these are modelled as separate investigations. The linkage to these by experimental investigations is proposed to done via the range of timestamps on the collected datafiles in that particular experiment.
INV_ABSTRACT	Description of the experiment, e.g. based on the proposal.
PREV_INV_NUMBER	Experiment number of a preceding and related experiment, e.g. in a chain of such experiments. This has not be used in practise and may be a candidate for removal in future versions of ICAT as aggregation of studies can be done in the STUDY_INVESTIGATION table.
BCAT_INV_STR	Short hand for representing a best guess at the Principal Investigator - this is used at ISIS when mining metadata from the back catalog (where no matching proposal information was available) and the column is short for back catalog investigator string. This could be changed to Principal_Investigator_Institution in future versions of ICAT to make it more widely of use.
GRANT_ID	This was meant to hold information about who has funded the experiment. However in practise this was never used by any of the facilities, so should in future versions of ICAT be removed.
INV_PARAM_NAME	Holds the defining parameter name for investigation based on the facility, e.g. in ISIS this may be run number range for the experiment. This may later be removed and replaces with a INVESTIGATION_PARAMETER table in versions of ICAT after 3.3. These investigation parameters could also be used to replace keyword, but this needs some discussion. Also if nested parameters were supported then the topic table could be removed also. This would require significant rework to dependent software system but would be an excellent generalisation mechanism for ICAT.
INV_PARAM_VALUE	The facility specific value of the inv_param_name. E.g. in the case of ISIS the actual range of numbers in some specified format for the range of values designating the run numbers.
INV_START_DATE	The official start date-time of the experiment. E.g. in ISIS as of the time of writing this could be derived from the date-time that the first raw datafile was read from the instrument.
INV_END_DATE	The official end date-time of the experiment. E.g. in ISIS as of the time of writing this could be derived from the date-time that the latest raw datafile was created on the instrument.

RELEASE_DATE	This is the date in the future that the raw data will be made available to other users (or publicly available) - this is informed by the data policy of the facility (e.g. 3 years for ISIS).
SRC_HASH	This stores a hash key to identify the records in a business system (e.g DUO Desk) which is the source of approved proposals in ICAT. From the Diamond DUO Desk primary key values are hashed together; this is needed as one ICAT record is often sourced from multiple records in the business system.

2.16 INVESTIGATION_TYPE

Holds the designation of valid investigation types.

Column Name	Comments
NAME	The name given to the INVESTIGATION_TYPE, e.g. experiment.
DESCRIPTION	Description of what the type denotes e.g. for commercial_experiment this might be: "A scientific experiment performed by a commercial company".

2.17 INVESTIGATOR

A facility user can be involved with more than one investigation and iinvestigation can have more than one facility user. This is a many-to-many mapping table which models this fact.

Column Name	Comments
INVESTIGATION_ID	The key of the investigation associated with the investigator.
ROLE	The role of the facility user in this investigation, for example Principal Investigator, Co-Investigator, etc.

2.18 KEYWORD

Holds the keywords associated with the investigations. Keywords can comve from a variety of sources, for example using words from the investigation title and/or abstract with the stop words removed, from specified key information in the proposal or user supplied with the proposal. Keywords maybe populated in other ways also. In future versions of ICAT implementing INVESTIGATION_PARAMETERS would do away with the need for a separate keyword table.

Column Name	Comments
INVESTIGATION_ID	The key of the investigation associated with this keyword.
NAME	The actual keyword. Note these are stored in a case sensitive way as they could alter the meaning of the keyword. A pertinent example would be chemical formulas.

2.19 PARAMETER

This table contains information about the valid parameters that can be used to describe samples, datasets and datafiles. It is recommended that a single parameter uses a single unit type so that data pertaining to that parameter type is using a uniform unit system through out the catalog aiding selection of data based

on values of a particular parameter. The PARAMETER table should hold a description of the types of information that the facility collects. However when data is registered into ICAT parameters that don't already exist are added but marked as unverified, so that they can be checked later and either accepted as a new valif facility parameter type or reconciled with another type or other relevant changes made.

Column Name	Comments
NAME	The name of the parameter; following the SI units is recommended where applicable in which case the PARAMETER.NAME maps onto SI base quantity or SI derived quantity e.g. Celsius temperature. If following SI is not possible then following the SI methodology of naming is recommended where applicable. Note SI quantities may not always be applicable as often you are recording the temperature of something not just the temperature in which case the Units should be as close to SI as possible.
UNITS	The unit (preferably SI symbol if applicable) for this parameter. Note any given parameter name can be held at multiple unit levels, this is needed to support data from different sources e.g. user office systems, values collected at proposal time and values collected at data registration. Use N/A when no Unit applies. An example of a value would be ?C.
UNITS_LONG_VERSION	The long version of units as short hand unit is normally used in practice. This should, if applicable be the SI Name, for example degree Celsius.
SEARCHABLE	Y or y - for allowing searches using this parameter and anything else for not searcheable including null.
NUMERIC_VALUE	Y or y denote that the value of the parameter is a number - anything else denotes that it is a string.
NON_NUMERIC_VALUE_FORMAT	Where the value is a string, this allows that value to be documented according to the rules or a regular expression.
IS_SAMPLE_PARAMETER	Y or y denote that the parameter is relevant for association with samples.
IS_DATASET_PARAMETER	Y or y denote that the parameter is relevant for association with datasets.
IS_DATAFILE_PARAMETER	Y or y denote that the parameter is relevant for association with datafiles.
DESCRIPTION	This describes the PARAMETER.NAME and is not a definition of the units.
VERIFIED	If Y this means that the parameter was loaded from the facility spreadsheet, list of approved values or that the parameter was unverified but has been checked and is now verified. N means that the parameter is not verified, i.e. it has not been approved by the facility; often this will be the case when datafiles or datasets are registered and the parameter values they have associated with them are new or have not been recognised as valid values before. Rather then not allow registration, it is allowed, but the values are flagged for later checking and verification to keep the mechanism of generality in the ICAT under some control such that random values are minimised.

2.20   PUBLICATION

This supports the linking of investigations to publications. This can be both publications that the current investigation is based on and publications created from the investigation.

Column Name	Comments
ID	Key for records in the PUBLICATION table.
INVESTIGATION_ID	The key of the related investigation.
FULL_REFERENCE	A citable reference to the work.
URL	A link to where the publication is available.
REPOSITORY_ID	Short hand identifier for the repository where the publication is available.
REPOSITORY	The name of the repository which contains a copy of the publication, e.g. STFC ePubs.

2.21   RELATED_DATAFILES

This table show how datafiles are related. It works off the axiom that there can only be one relationship between any two specified files. It would, for example, support the capture of information showing how newer updated versions of captured datafiles were generated. It is to be understood that the source is related to the destination and the RELATION column holds information pertaining to this relationship. The relationship might hold information about a transformation or might be another type of relationship e.g. newer version.

Column Name	Comments
SOURCE_DATAFILE_ID	The key of the source datafile in the DATAFILE table.
DEST_DATAFILE_ID	The key of the destination datafile in the DATAFILE table.
RELATION	Specifies the nature of the relationship - i.e. <dest_file> a <relation> of <source_file>. where relation could be for example: subset, newer version, reduced, used the configuration of - e.g. sample environment (configuration, temperature, pressure).

2.22   SAMPLE

This table stores the sample information for both the experiment (abstract) and the datasets (instances) e.g. in the case of the information from the proposal system the abstract sample information should have the instance column set to NULL. It should be noted that Samples are often substituted if the stated samples instances do not produce good results. Scientists usually have reserve samples or other samples which they want to try out (this usually happens half of the time as an estimate on ISIS) this can happen due to a variety of reasons:

1. The sample instances is not producing good results

2. The experiment is progressing better than expected and the experimental team have time left to look at a reserve sample

This is not usually a problem if the sample safety characteristics are the same; usually the will just substitute the sample. However if the substituted samples:

1. Have very different sample safety characteristics

2. Are totally different samples from the original and they person is attempting to queue jump by analysing a particularly "hot" (from a research perspective) compound.

Then a station scientist must be asked/involved as

1. Sample can can explode and the people going in to clear up need to know if these are for example radioactive

2. This is queue jumping and is attempted often but is considered to be unfair play and is actively stopped

Column Name	Comments
ID	The key for the sample information in ICAT.
INVESTIGATION_ID	The key of the related investigation.
NAME	Descriptive name of the sample.
INSTANCE	This is a designator representing difference instances of the sample e.g. (1, 2, 3) brought by the experimenters. This will be NULL in the case where the abstract sample is being described.
CHEMICAL_FORMULA	This is the chemical formula of the sample. Note this can be NULL in cases when there is a complex layered target, where the chemical formula is unknown, or only partially known or when trying to study the interface between two solids; in these cases more support in describing these situations in ICAT may be needed.
SAFETY_INFORMATION	This field holds sample safety or sample hazard information. Where the safety information has numerous aspects, then these can be stored as sample parameters and a note stored here to that effect.
PROPOSAL_SAMPLE_ID	A copy of the sample_id in the database from which this record was imported. This lets us propagate changes made to a sample's child tables, such as sample_parameter.

2.23 SAMPLE_PARAMETER

Stores name-value pairs of metadata associated with the sample. The type of names are constrained by their entry in the PARAMETER table, allowing for building up controlled vocabularies for a particular facility around pertinent sample information.

Column Name	Comments
SAMPLE_ID	The key of the related sample in ICAT.
NAME	The name of the sample parameter, e.g. disposal method.
UNITS	The units of the parameter, SI if applicable.
STRING_VALUE	If the field NAME has a value expressed as a string then the data will be present here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
NUMERIC_VALUE	If the field NAME has units expressed as a number then the value will be set here. Whether a number or value is stored is determined by what is set is the PARAMETER.NUMERIC_VALUE column.
ERROR	Holds the error range for the STRING_VALUE, NUMERIC_VALUE or RANGE_TOP and RANGE_BOTTOM combination.
RANGE_TOP	If the value is a range this holds the maximum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
RANGE_BOTTOM	If the value is a range this holds the minimum value. This has not been used in practise and might be a candidate for removal in the next version of ICAT.
DESCRIPTION	Where and how the parameter was recorded or extracted as oppose to a definition of the parameter; the latter being defined in the PARAMETER table.

2.24 SHIFT

This stores information pertaining to when experiments are actually performed. This information is needed in scenarios where programs at the data collection device check with ICAT or its proxy IKitten to determine whether experimenters at the device and what experiment they are performing can be accurately ascertained. While this is more of an operational than archive feature, the metadata associated with this situation can be used to help data registration and filling out fields of metadata in the data files themselves.

Column Name	Comments
INVESTIGATION_ID	The key of the related investigation.
START_DATE	The start date and time of the shift.
END_DATE	The end date and time of the shift.
SHIFT_COMMENT	A comment on the shift. This could, for example explain why the data and time of the shift had been changed.

2.25 SOFTWARE_VERSION

This hold information about versions of the software used to do data reduction and analysis. This has not been used in practice in ICAT 3.1 to 3.3. It maybe used in the future, but will need re-visiting to see whether or not it meets the need of storing such information.

Column Name	Comments
ID	The key of the software version record in ICAT.
NAME	The name of the program.
SW_VERSION	The version string of the program.
FEATURES	The features of the program.
DESCRIPTION	Description of the program. This could for example include a URL for further information about the software.
AUTHORS	The authors of the program.

2.26   STUDY

STUDY is used to aggregate investigations. This is not used in ICAT 3.3 but may be a basis to allow facility users to group their experiments or sets of related experiments in the future.

Column Name	Comments
ID	The key of the study in ICAT.
NAME	Unique name given to the study.
PURPOSE	The reason for aggregating this particular set of investigations, i.e. the aggregation criteria used.
STATUS	Ongoing or complete, as there could be additional investigations planned in the future which could be applicable to this study.
RELATED_MATERIAL	This field holds information related to this study. For example this could be related studies in different facility ICATs or similar sample investigated at a different facility (e.g. DLS or CLF if work done at ISIS). To allow the connection of different sources of relevant information at the moment this is unstructured text but this may become more structured in future versions of ICAT.
STUDY_CREATION_DATE	When the study was created. This could be removed in future version of ICAT as the audit column CREATE_TIME now holds this information.
STUDY_MANAGER	This should map to the FACILITY_USER.FEDERAL_ID, as the user who creates the study need not be an investigator but should be known to be a registered user of the facility. Authorisation rules would need to be added to ICAT_AUTHORISATION.

2.27   STUDY_INVESTIGATION

Mapping table holding the link between studies and investigation.

Column Name	Comments
STUDY_ID	The key of the study in ICAT.
INVESTIGATION_ID	The key of the investigation in ICAT.
INVESTIGATION_VISIT_ID	Should be removed in a future version of ICAT.

2.28   STUDY_STATUS

A lookup table holding information about the valid status values for studies.

Column Name	Comments
NAME	The name give to the status.
DESCRIPTION	A description of what the study status actually means.

2.29 THIS_ICAT

Reflective information about the facility served by this instance of ICAT.

Column Name	Comments
DAYS_UNTIL_PUBLIC_RELEASE	The number of days until the raw data is made publicaly available for a given experiment. A value of 0 means that the data should be made immediately available.

2.30 TOPIC

Used to build a taxonomy of terms relevent to investigations.

Column Name	Comments
ID	The key of the record in ICAT.
NAME	The name of the classification term.
PARENT_ID	The key of the parent record in the TOPIC table. This will be some special value where the topic is at the root of the hierarchy.
TOPIC_LEVEL	The level in the hierarchy the term is at.

2.31 TOPIC_LIST

Stores the link between investigations and their topics.

Column Name	Comments
INVESTIGATION_ID	The key of the linked investigation.
TOPIC_ID	The key of the linked topic (refers to the leaf node).

2.32 USER_ROLES

Stores special roles, e.g. Administrators, assigned to the users of the Oracle Application Express applications which interact with ICAT. The table APPLICATIONS is linked to USER_ROLES. These users specified will not necessarily be facility users. In future version of ICAT these tables will be removed as applications interact with ICAT via the ICAT API.

Column Name	Comments
APP_CODE	The application code refers to values in the APPLICATIONS table.
USERNAME	The user's corporate username. Must be upper case. This will be there federal Identifier at STFC.
ROLE	The role of the user with regards to the application.