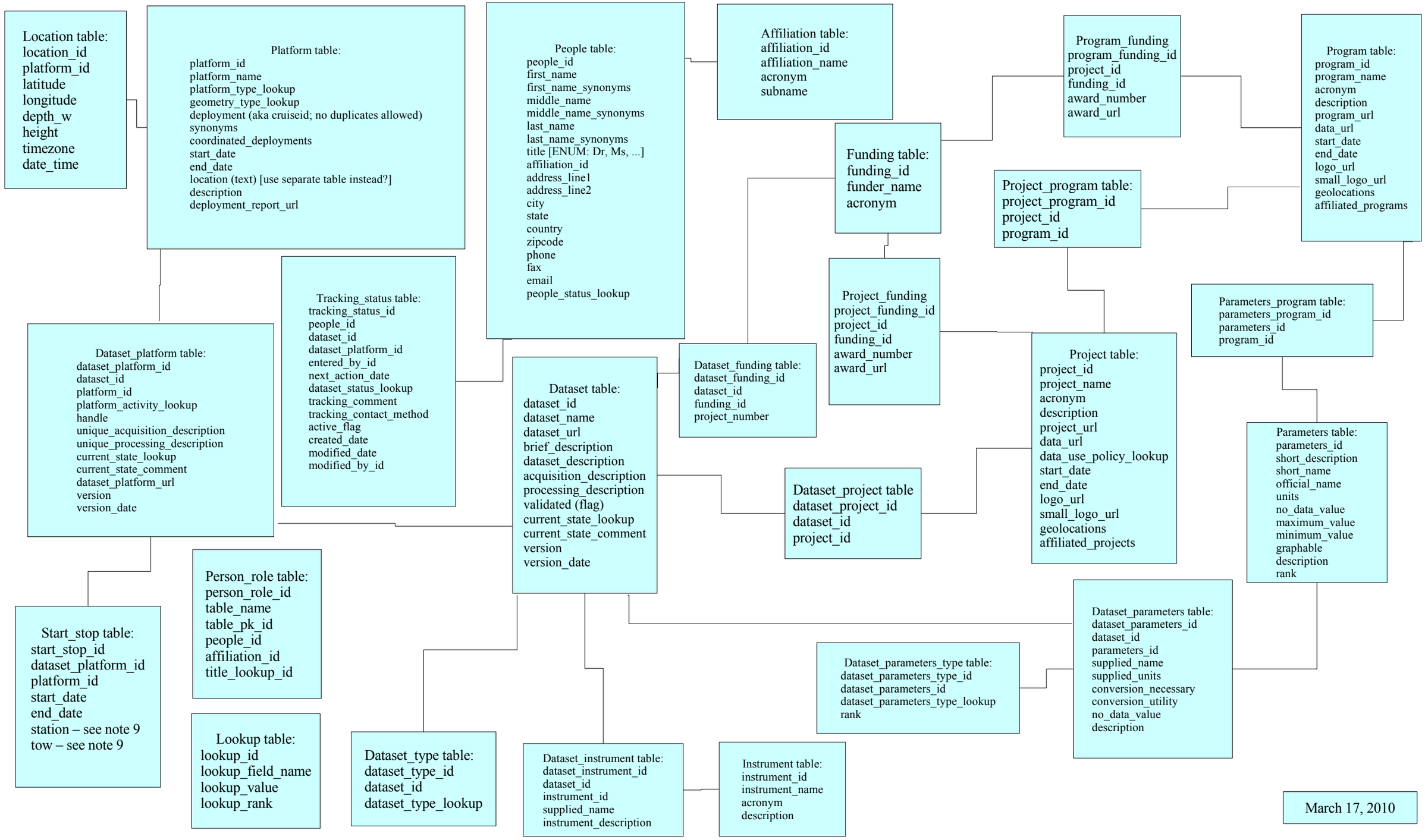
Location table:
location_id
platform_id
latitude
longitude
depth_w
height
timezone
date_time
Platform table:
platform_id
platform_name
platform_type_lookup
geometry_type_lookup
deployment (aka cruiseid; no duplicates allowed)
synonyms
coordinated_deployments
start_date
end_date
location (text) [use separate table instead?]
description
deployment_report_url
People table:
people_id
first_name
first_name_synonyms
middle_name
middle_name_synonyms
last_name
last_name_synonyms
title [ENUM: Dr, Ms, ...]
affiliation_id
address_line1
address_line2
city
state
country
zipcode
phone
fax
email
people_status_lookup
Affiliation table:
affiliation_id
affiliation_name
acronym
subname
Program_funding
program_funding_id
project_id
funding_id
award_number
award_url
Program table:
program_id
program_name
acronym
description
program_url
data_url
start_date
end_date
logo_url
small_logo_url
geolocations
affiliated_programs
Funding table:
funding_id
funder_name
acronym
Project_program table:
project_program_id
project_id
program_id
Tracking_status table:
tracking_status_id
people_id
dataset_id
dataset_platform_id
entered_by_id
next_action_date
dataset_status_lookup
tracking_comment
tracking_contact_method
active_flag
created_date
modified_date
modified_by_id
Project_funding
project_funding_id
project_id
funding_id
award_number
award_url
Parameters_program table:
parameters_program_id
parameters_id
program_id
Dataset_platform table:
dataset_platform_id
dataset_id
platform_id
platform_activity_lookup
handle
unique_acquisition_description
unique_processing_description
current_state_lookup
current_state_comment
dataset_platform_url
version
version_date
Dataset table:
dataset_id
dataset_name
dataset_url
brief_description
dataset_description
acquisition_description
processing_description
validated (flag)
current_state_lookup
current_state_comment
version
version_date
Dataset_funding table:
dataset_funding_id
dataset_id
funding_id
project_number
Project table:
project_id
project_name
acronym
description
project_url
data_url
data_use_policy_lookup
start_date
end_date
logo_url
small_logo_url
geolocations
affiliated_projects
Parameters table:
parameters_id
short_description
short_name
official_name
units
no_data_value
maximum_value
minimum_value
graphable
description
rank
Dataset_project table
dataset_project_id
dataset_id
project_id
Person_role table:
person_role_id
table_name
table_pk_id
people_id
affiliation_id
title_lookup_id
Start_stop table:
start_stop_id
dataset_platform_id
platform_id
start_date
end_date
station – see note 9
tow – see note 9
Dataset_parameters table:
dataset_parameters_id
dataset_id
parameters_id
supplied_name
supplied_units
conversion_necessary
conversion_utility
no_data_value
description
Dataset_parameters_type table:
dataset_parameters_type_id
dataset_parameters_id
dataset_parameters_type_lookup
rank
Lookup table:
lookup_id
lookup_field_name
lookup_value
lookup_rank
Dataset_type table:
dataset_type_id
dataset_id
dataset_type_lookup
Dataset_instrument table:
dataset_instrument_id
dataset_id
instrument_id
supplied_name
instrument_description
Instrument table:
instrument_id
instrument_name
acronym
description
March 17, 2010

The BCO-DMO Metadata Database Schema

Notes:

1. All tables have a created_date field.
2. The platform table replaces the cruiseid table to be more general. It could also have been called the deployment table.
3. I decided not to change name specific id's, like people_id to be just id even though it works well elsewhere. The problem that I see is that it will make reading linking tables harder to understand.
4. The location table contains one or more positions/times for this dataset, getting the data from the event log perhaps or from the meteorological data sets.
5. Do we need the tables for instrument and dataset_instrument? - Yes.
6. Do we need to store max and min values of positions, and other parameters? STILL TO BE DECIDED
7. There will be more than one location table entries for a dataset. Hence the need for a linking table, dataset_platform table.
8. June 21, 2007. Replaced depth with depth_w in location table.
9. June 26, 2007. The location table is now connected tot he Platform table, representing a summary of the "best" navigation for the cruise. We need to encode information about the starting and stopping of measurements and that is done using the new Start_stop table, mainly of dates/times. This table does include fields for latitude and longitude, but if these are not present, the system should look up the position information in the location table, dead reckoning between positions if necessary, assuming a straight line between fixes. To reconstruct the cruise track, it will be necessary to sort the retrieved information from the location table by time (or maybe location_id if the data are added in order). It would be best if the times were stored in the location table as GMT, with timezone information included with each record. This information may be used to convert input data provided in local time to be stored in GMT. It was noted that we do not have a "Scientific_personnel" table to record who participated during the cruise. This could be added, but at this time it is felt that names connected to each of the data sets, via the Dataset_platform_people table, is sufficient. Another discussion to have is whether tow, cast, and/or station number information should also be saved within the start_stop table. These are currently included, but ultimately the overhead in storing and maintaining this information may be too high. Note that the Start_stop table contains the platform_id, which is redundant information, given that it could be obtained via the dataset_platform_id value, but platform_id is provided as a way to simplify subsequent retrieval.
10. September 13, 2007. Added dataset_platform_url to the dataset_platform table. Add the dataset_status and dataset_platform_status tables. The "status" field name in both tables should probably be an enumerated list while the comment list will enable more free form information.
11. October 10, 2007. Add Project and Program tables. Replace project with project_id in the Dataset table. Clarify that the status information in the Dataset_status table is an ENUM list and therefor a controlled vocabulary. Add acronym field names to the affiliation, funding and instrument tables. Possible ENUM values for the title in the Dataset_platform_people table are originator, contact, analyst, technician, curator, manager, student, principal investigator, scientific investigator, and publisher.

12. October 11, 2007. Remove dataset_id from the Location table and change the dataset_platform_id to platform_id in the Location table. In the Platform table, the deployment field is a non duplicates field. Add synonyms to this table. It will be a comma separated list of synonyms for deployment. Add a version field to the Dataest_platform and Dataset tables. In the Dataset_status table, add entered_by_id and next_action_date fields. In this table, the comment field will contain potentially lots of text so it should be able to accommodate many characters. Add program_id and contact_id fields to the Project table. Add geolocations to the Program table. Remove the conversion_utility field from the Parameters table. The geolocations field in the Project and Program tables is a comma separate list of geographical areas (e.g. Pacific Ocean, Georges Bank, etc.) but could also contain Marsden Squares and/or C-squares numbers.
13. January 23, 2008. Remove the latitude and longitude fields from the Start_stop table. These can be determined by using the start_date and end_date values and doing a lookup and possibly interpolation to the data in the Location table. Add entered_by_id and next_action_date fields to the Dataset_platform_status table. Fix spelling of this table name. The status field in this table is an enumerated list, similar to the status field in the Dataset_status table.
14. January 24, 2008. Remove the dataset_id from the Modification_history table as this value may not always be ready. Separate out the table name and start and end local ID value from within the description field from this table and put them in their own columns, table_changed, start_local_id, and end_local_id, respectively.
15. April 22, 2008. Remove unnecessary dataset_id field from the Project table. Add co_pi2.id to both the Project and Program tables. Fix the spelling of dataset_parameters_id in the Dataset_parameters table. Add related_projects column to the Project table and related_programs column to the Program table. Add the new table People_status, comparable to the Dataset_status table, but for people. Yet to be done is the implementation of the new columns to store the information about how to plot the different datasets. It should also be noted that while the status field in the three status tables are listed as of type ENUM, this has yet to be implemented in the actual tables. In order to proceed with this, we need to define what values this field can take on.
16. May 6, 2008. Change related_program and related_project to affiliated_programs and affiliated_projects. Add coordinated_platforms to Platform table. Fix spelling of end_local_ed to end_local_id in Modification_history table. Add first_name_synonyms, middle_name_synonyms, and last_name_synonyms to the People table to handle the situation that people change their names. Fix the headings of each box to be consistent, i.e. ending in “table:”. Fix the formatting of the Modification_history table: the entries needed to be left justified.

17. June 14, 2008. Change the MySQL database engine to InnoDB from MyISAM in order to support foreign key declarations. Add in the foreign key declarations for all foreign keys except for the people table. Add a small_logo_url column to the Project and Program tables. Add the Lookup table and replace several ENUM columns with lookup ids into the Lookup table. This applies to the People_status, Dataset_platform_people, Dataset_platform_status, Dataset_status, Dataset_platform, and Platform tables. Note that until the new lookup values are added the original ENUM fields will be retained in the database. However, once the lookup values are added, the original ENUM field columns will be deleted. Change the entered_by_id field in the Modification_ history table to be an INT type and a foreign key as well. The "cast" field name from the Start_stop table has been removed. It was removed some time earlier from the database, but the schema did not reflect this until now.
18. July 8, 2008. Add a many to many linking table between the Dataset table and the Project table (called Dataset_project) and between the Project table and the Program table (called Project_program). Remove the affiliated_project and affiliated_program columns once these data have been transformed into the new linking tables. Add a the handle column to the Dataset_platform table to record the DOI or handle assigned by the data archiving agency.
19. August 25, 2008. Remove Modification_history table from scheme figure to make room for two new tables. Add the Dataset_type table and the Dataset_parameters_type table. These tables could not have dataset_type_lookup and dataset_parameters_lookup entries declared as foreign keys since the lookup table must first be recreated as a InnoDB type table. Delete project_id from the Dataset table and program_id from the the Project table. Add small_logo_url to the Project and Program tables. (This change was done some time ago.) Add affliation_id to the Dataset_platform_people table, but do not declare as required or as a foreign key until application code is updated. Add conversion_necessary to the Dataset_parameters table. It can take the values of either 'yes' or 'no', with a default of 'no'. The lookup table was updated to include the new entries needed by the Dataset_type table and the Dataset_parameters_type table.
20. September 5, 2008. Remove project_id from the Dataset table and program_id from the Project table as these were replaced by the Dataset_project and Project_program tables. [These columns still exist in the database but will be removed once the new tables are fully implemented in the software.] Add dataset_id to the Datset_type table schema picture and add dataset_parameters_type_id to the Dataset_parameters_type table schema picture as these were inadvertently left out of the schema picture when these tables were added August 25, 2008. Fix the line joining the People_status table to the People table.

21. September 15, 2008. Remove project_id from the Dataset table and program_id from the Project table. They are replaced by the Dataset_project and Project_program tables, respectively. Rename fill_value to no_data_value in the Parameters table and the Dataset_parameters table. Rename standard_name to short_description in the Parameters table. Remove common_name and equivalent_name from the Parameters table.
22. September 18, 2008. Correct the spelling of the “no_data_value” entry in the the Dataset_parameters table on page one.

23. October 16, 2008. Add the many to many table Parameters_program. Move the contents of Dataset_platform table's acquisition_description and processing_description to the Dataset table and rename the former columns to be unique_acquisition_description and unique_processing_description. Add description and deployment_report_url to the Platform table.
24. November 4, 2008. Apply changes mentioned in item 23 to the live database, in particular create the Parameters_program table, add acquisition_description and processing_description to the Dataset table, rename these columns to unique_acquisition_description and unique_processing_description in the Dataset_platform table, add description and deployment_report_url to the Platform table,move the coordinated_platforms column to the correct place in 'the Platform table, and change the rank column to INTEGER (10) instead of TINIINT in the Dataset_parameters_type table.
25. November 14, 2008. Change coordinated_platforms to coordinated_deployments in the Platform table.

26. February 11, 2009. The following changes have been made to the test database with the expectation that they will be made to the live database shortly. Add contact_id and co_pi2_id to the Program table. Add rank (as a decimal number) to the parameters table. Brief_description has been added to the Dataset table. Because of additions to be added to the Dataset_platform table (described below) it is possible that the validated flag will be removed from the Dataset table. Create a new table called Contact_status replacing the identical People_status table but adding keyword_string to this table. Add current_state and current_state_comment to the Dataset_platform table.
27. February 12, 2009. Add current_state and current_state_comment to the Dataset table. Make the changes to the live database mentioned in the February 11, 2009 entry.
28. February 26, 2009. The following changes were made to the test database but are anticipated to be made to the live database shortly. Remove the Contact_status, Dataset_status, and Dataset_platform_status tables. Add the Tracking_status table. Add people_status_lookup to the People table. Update the diagram to show the presence of the geometry_type_lookup column in the Platform table.
29. March 3, 2009. Implement the changes mentioned in item #28 and rename tracking_status_lookup to dataset_status_lookup in the Tracking_status table.

30. March 11, 2009. Add data_url to both the Project and Program tables. Replace the missing line between the Dataset and Dataset_platform tables.
31. April 28, 2009. Add the graphable column to the Parameters table, with values of Y or N. Change the columns current_state in the Dataset and Dataset_platform tables to current_state_lookup and make them integer values rather than text entries. It was noted that the Dataset_type table probably should have been called Mapserver_type table.
32. July 28, 2009. Replaced the Dataset_platform_people table with a new table called Person_role which will serve this function for the Dataset_platform, Platform, Project and Program tables. That is, it will be used by several tables using the column table_name to specify which table it applies to and table_pk_id for the primary key id from this specified table. Remove references to specific roles from several tables including chief_scientist_name_id and co_chief_scientist_name_id from the Platform table, and lead_pi_id, co_pi_id, co_pi2_id, and contact_id from the Project and Program tables.
33. October 26, 2009. These changes were made to the live database August 4, 2009, but due to problems in running this graphics program the diagram was not updated until today. We added two new intersection tables, Project_funding and Program_funding, to support the many to many relationship the Dataset table and the Program and Project tables. It was done to support the needs of OCB Project Office.
34. March 17, 2010. The following changes are being made. Add supplied_name to the Dataset_instrument table, similar in concept to the supplied_name in the Dataset_parameters table, to capture what the contributor calls their instrument. Add version_date to the Dataset table to insure we have a properly formated date for the version information. We will initially keep the version column in the Dataset table since it is possible that people will have their own text name (none date entry) for the version information. Add version_date to the Dataset_platform table. We will think about a way of filling in this field, if empty, initially using the version/version_date for Dataset. Add data_use_policy_lookup column to the Project table to keep the id number of the entry in the Lookup table containing the statement of the contributor's use policy. Add minimum_value and maximum_value to the Parameters table. It is hoped that this will suffice and we will not need comparable columns in the Dataset_parameters table. It should be noted that we plan to add “archived” and “restricted” as possible values in the lookup_table for the current_state_lookup value in the Dataset table. We considering how to initialize the current_state_lookup and current_state_comment columns in the Dataset_platform table. They have never been used so far (probably because OSPREY does not display these as input options). The program_name and acronym in the Program table are unique entries. No duplicates are allowed nor can they be NULL. Also, the instrument_name in the Instrument table should be unique and not NULL. These changes will also be added to the database definition. The affiliated_projects column was added to the Project table some time ago, but was not reflected in the schema diagram. That has been corrected. An initial study of using triggers suggests that we can add a trigger to the Dataset version_date column so that it stays current based on changes to the Dataset_platform version_date column data. We will investigate this further. Update the program table to reflect the addition of the new column, affiliated_programs. It too was added some time ago.