



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 to the 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 therefore 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 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 an InnoDB type table. Delete project_id from the Dataset table and program_id from the Project table. Add small_logo_url to the Project and Program tables. (This change was done some time ago.) Add affiliation_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 Dataset_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.