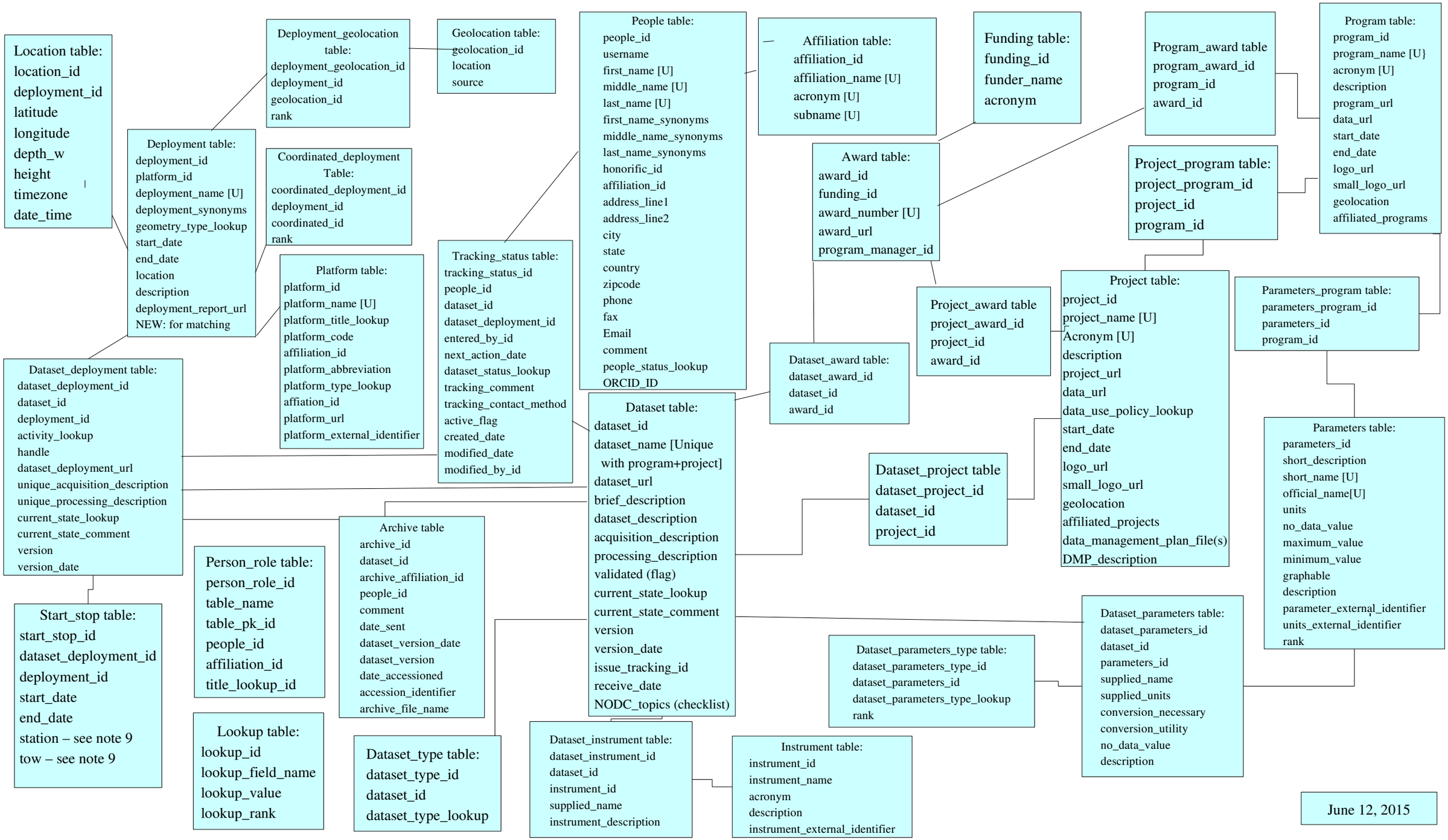
Location table:
location_id
deployment_id
latitude
longitude
depth_w
height
timezone
date_time
Deployment_geolocation table:
deployment_geolocation_id
deployment_id
geolocation_id
rank
Geolocation table:
geolocation_id
location
source
People table:
people_id
username
first_name [U]
middle_name [U]
last_name [U]
first_name_synonyms
middle_name_synonyms
last_name_synonyms
honorific_id
affiliation_id
address_line1
address_line2
city
state
country
zipcode
phone
fax
Email
comment
people_status_lookup
ORCID_ID
Affiliation table:
affiliation_id
affiliation_name [U]
acronym [U]
subname [U]
Funding table:
funding_id
funder_name
acronym
Program_award table
program_award_id
program_id
award_id
Program table:
program_id
program_name [U}
acronym [U]
description
program_url
data_url
start_date
end_date
logo_url
small_logo_url
geolocation
affiliated_programs
Deployment table:
deployment_id
platform_id
deployment_name [U]
deployment_synonyms
geometry_type_lookup
start_date
end_date
location
description
deployment_report_url
NEW: for matching
Coordinated_deployment Table:
coordinated_deployment_id
deployment_id
coordinated_id
rank
Award table:
award_id
funding_id
award_number [U]
award_url
program_manager_id
Project_program table:
project_program_id
project_id
program_id
Platform table:
platform_id
platform_name [U]
platform_title_lookup
platform_code
affiliation_id
platform_abbreviation
platform_type_lookup
affiation_id
platform_url
platform_external_identifier
Tracking_status table:
tracking_status_id
people_id
dataset_id
dataset_deployment_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_award table
project_award_id
project_id
award_id
Project table:
project_id
project_name [U]
Acronym [U]
description
project_url
data_url
data_use_policy_lookup
start_date
end_date
logo_url
small_logo_url
geolocation
affiliated_projects
data_management_plan_file(s)
DMP_description
Parameters_program table:
parameters_program_id
parameters_id
program_id
Dataset_award table:
dataset_award_id
dataset_id
award_id
Dataset_deployment table:
dataset_deployment_id
dataset_id
deployment_id
activity_lookup
handle
dataset_deployment_url
unique_acquisition_description
unique_processing_description
current_state_lookup
current_state_comment
version
version_date
Dataset table:
dataset_id
dataset_name [Unique with program+project]
dataset_url
brief_description
dataset_description
acquisition_description
processing_description
validated (flag)
current_state_lookup
current_state_comment
version
version_date
issue_tracking_id
receive_date
NODC_topics (checklist)
Dataset_project table
dataset_project_id
dataset_id
project_id
Parameters table:
parameters_id
short_description
short_name [U]
official_name[U]
units
no_data_value
maximum_value
minimum_value
graphable
description
parameter_external_identifier
units_external_identifier
rank
Archive table
archive_id
dataset_id
archive_affiliation_id
people_id
comment
date_sent
dataset_version_date
dataset_version
date_accessioned
accession_identifier
archive_file_name
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_deployment_id
deployment_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
instrument_external_identifier
June 12, 2015

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. Add geometry_type_lookup column to the Platform table.
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.

35. April 28, 2010. Fix spelling of “geolocation” in both the Project and Program tables in the schema picture. There is no “s”, i.e. the word is not plural. Update the February 12, 2009 entry to reflect the addition of the geometry_type_lookup column to the Platform table. Add username to the People table schema diagram. It had been left out. Change the database so that the first, middle and last name synonyms columns display after their corresponding first, middle and last name columns. Move the affiliated_programs column in the Program table to appear after the geolocation column.

36. July 16, 2010. Change project_id to program_id in the schema diagram in the Program table. Add new Archive table; add creation_date to all tables and automate creation_date and modified_date using triggers; add trigger to define geometry_type; add parameter_uri and units_uri to Parameters table; add instrument_uri to Instrument table; implement honorific_id in People table to replace title column in People table; change order of columns in Tracking_status table. For a time, the title column will remain in the People table, but it will be removed once all code is updated.

37. July 26, 2010. Add dataset_id to the Archive table picture. It was in the table but left out of the picture. Add dataset_platform_id to the Archive table to deal with the situation where there is no dataset_url specified and we have to rely on the dataset_plotform_url's. This will also allow us to track when new data are added to existing, archived Datasets.

38. August 16, 2010. Add handle to Dataset table. Add comment to People table.

39. August 30, 2010. Fix spelling of graphable in the Parameters table. It had been spelled as graphical. Rename parameter_uri to parameter_external_identifier and units_uri to units_external_identifier in the Parameters table and instrument_uri to instrument_external_identifier in the Instrument table.

40. February 17, 2011. Add archive_file_name to Archive table. We are making a major change to the metadata database schema by dividing the Platform table into two, Platforms and Deployments. (Yes, we know, it should have been done this way in the first place.) At the same time, we are separating out synonym into a separate table; geolocation into an intersection table and a Geolocation table; and coordinated_deployments into a Coordinated_deployments table. All have ranks associated with the entries to specify ordering of the connections. Also, several other tables are affected since their foreign keys and platform_id have to be replaced by deployment_id. This includes the Dataset_platform table and the Location table. Also, there is a change in name of dataset_platform_id to dataset_deployment_id in the Tracking table. The Dataset_platform table is renamed Dataset_deployment table. The

change to the Archive table will take effect immediately. However, all other changes will take time to test and implement. The schema diagram was modified to reflect the connections between the Tracking-status table and the Dataset_deployment, Dataset, and People tables.

41. March 7, 2011. After review, we have made the following additional changes to those made in item 40: added in deployment column in the Deployments table (left out in error); renamed Deployment_geolocation to Deployments_geolocation; renamed Coordinated_deployments to Coordinated_deployment; added affiliated_id and platform_url to the Platform table; renamed Synonyms table to Deployments_synonym table; added Authority table. Note that the platform_url points to the specific ship if available, or to the operator's site if not available. While it is possible for the platform_code to be NULL in the Platforms table, we will try to force it to be unique. The deployment column in the Deployments column should be constrained to be unique (as in the past). Change the Deployments table name to Deployment. References to dataset_platform_id are changed to dataset_deployment in tables Dataset_deployment, Start_stop,Tracking_status, and Archive. The entries in the Lookup table for dataset_platform_current_state_lookup and platform_activity_ lookup are changed to dataset_deployment_current_state_lookup and deployment_activity_lookup, respectively. In the renamed Dataset_deployment table, platform_activity_lookup is renamed activity_lookup and dataset_platform_url is renamed dataset_deployment_url.

42. April 4, 2011. These are changes to the schema picture: The spelling of the primary key in the Coordinated_deployment table was changed to coordinated_deployment_id; removed the dataset_deployment_id column in the Archive table; and the authority_url name needed to be left justified in the Authority table. In the Deployment table, change the deployment column name to deployment_name.

43. April 26, 2011. Add the missing line between the Dataset and Dataset_funding tables. The People_role table will have the data changed in the table_name column to reflect the table name change from Dataset_platform to Dataset_deployment.

44. May 6, 2011. Add the location column back into the Deployment table for free field input of location Information.

45. May 11, 2011. Fix the diagram, changing the Platforms table name to Platform. Fix the diagram primary key in the Deployment_synonym table to be deployment_synonym_id. Fix the database Start_stop table column name from dataset_platform_id to dataset_deployment_id. Add in the authority_id, affiliation_id and platform_url columns in the Platform table. They were left out by accident. Add foreign key constraints to the Platform table and the Start_stop table.

46. May 26, 2011. Add platform_title_lookup to the Platform table since we are splitting out the vessel/platform title from the platform_name, for example, R/V Knorr is stored as Knorr with the R/V title accesses via the Lookup table.

47. December 14, 2011. Reorganize the way we represent and store aware information by creating an award table. The three tables which contained award information, Dataset_funding, Program_funding and Project_funding are renamed Dataset_award, Program_award and Project_award and their corresponding keys are renamed Dataset_award_id, Program_award_id and project_award_id respectively. The award information columns, project_number, award_number, and award_number respectively are replaced by a foreign key, award_id, to the Award table. Also, the award_url columns are removed and kept in the Award table, along with new information, program_manager_id, a foreign key referencing the program managers name in the People table. Other changes made at this time include: removing the title column in the people table; making the username column in the people table unique; changing the zero values of the honorific_id column to be NULL; add foreign key constraints to the dataset_deployment_id and deployment_id columns in the Start_stop table.

48. February 16, 2012: Fix the notes from December 14, 2011 correcting the references from Dataset_award (etc) to Dataset_funding. Add platform_external_identifier to the Platform table. Also, remove the modification_history table as it is no longer used. Add in the triggers for the created_date and modified_date in the Award table.

49. June 13, 2012: The award_url column was not shown in the schema diagram as part of the Award table. This was added. The connection between the Dataset and Dataset_deployment tables was not shown in the schema diagram. This was added. There were no changes to the schema design, only the diagram was corrected.

50. On October 18, 2013 the longitude column in the Location table was changed from a FLOAT variable to a FLOAT(7,4) variable since the precision on this computer is only six (6) significant digits and we want to preserve numbers to four (4) significant digits after the decimal point. The diagram did not change so the date on the diagram was not changed.

51. December 6, 2013: On November 14, 2013 we completed the migration from the stand-alone metadata database to a Drupal (version 7) implementation. There were several issues needing immediate attention but nothing serious enough that necessitated a regression. In the Drupal implementation, the handle column in the Dataset content page was not retained. This information is included in the Archive content page as this content type can record several different kinds of archiving information, including NODC accession numbers and MBL DOIs. The Deployment_synonym table has been removed and the deployment_synonyms column added to the Deployment table (content type). This was done because Drupal, by default, supports the possibility of multiple values for a field.

52. March 25, 2014: Add [U] or [Unique] to contents that are to be unique. In some case, like the People table, the Dataset table, and the Affiliation table, the entries are unique but only when connected to other contents. For People, the name is unique when combining first, middle and last names. For Dataset, the dataset_name is unique when tied to a project acronym and program acronym. For Affiliation, the entry is unique when name, acronym, and subname are combined. In the Platform table, authroity_id is replaced by affiliation_id and the Authority table is removed. The data_management_plan_url is added to the Project table. Of course, all primary keys are unique in each table.

53. November 20, 2014: Add data_management-plan_file(s) and DMP_description to the Project table. Add issue_tracking_id, receive_date and NODC_topics to the Dataset table. Add a placeholder for match up link to the Deployment table. Add ORCID_ID to people table.