Data Management for the CRESP Amchitka Project

Vikram M. Vyas, Yuri Mun and Lisa Bliss

This Appendix describes the procedures followed by the project personnel for inventorying and tracking data and information associated with the project.

Several intermediate drafts of documents, as well files containing raw data or processed results were generated and transmitted during the course of the project. The objective of the Data Management component was to serve as a centralized repository for information, and as the coordinating node for organizing laboratory analysis results into usable databases. Not all personnel were involved with all parts of the project. Furthermore, for some classes of information, project personnel were also grouped in different categories in order to preserve integrity of information. A master document was prepared and distributed to project personnel in order to clearly identify the lines of communication, the people who were responsible for generating the information, and the people entitled to receive the distributed information.

During all phases of the project, drafts of documents as well as raw data files were emailed to the Data Management component for centralized archiving. These documents were backed up every week on CDs and kept in a physically separate building to protect against data loss.

I. Input of Information to Database[1]

A. Transfer of information from logbooks to Database

The biological specimens collected were listed in two logbooks – the Invertebrate (Specimens) Logbook, and the Vertebrate (Specimens) Logbook. Information from these logbooks was entered into electronic spreadsheets for subsequent use. The data entry process followed these steps:

1. Personnel at DM enter information from carbon copies log books into the excel spreadsheets. Any queries or clarifications are flagged.
2. All queries regarding entry of information are resolved with personnel responsible for collecting the specimens and maintaining expedition log books. If any changes had to be made, they are entered into the original and carbon copies of the log book and signed by authorized personnel.
3. Any necessary changes were also made to the excel spreadsheets.
4. The excel spreadsheets were independently verified against the entries in the log book.

[1] See appendix 1 for detailed information on the database structure.

B. Modifications and Additions to Database

1. When currently unprocessed specimens were processed into composites, the information in the database was updated to keep track of the new samples.
2. Recoded specimen ID information and shipment information from shipment notifications were entered into the database.
3. Information from additional sources, such as GPS locators, was synthesized with the log book information in the database.

II. Inventorying and Relocation of Specimens:

A. Initial Expedition Inventory

All specimens collected during the Amchitka expedition were inventoried, as follows:

1. Requested chests of biological specimens were shipped from the cold storage in Newark. Personnel responsible for inventorying were notified of the chest numbers and quantity of chests that had been requested for shipping, with the date and expected time of arrival.
2. Inventorying personnel counted chests delivered from cold sotrage, and verified that all requested chest numbers (and total quantity of chests) were present. The verification information was entered into the inventorying logbook which was distinct from the previously mentioned expedition logbooks. The chain of custody (COC) form for each opened chest will be retained and scanned electronically for records. The COC form for each opened chest was countersigned by two inventorying personnel.
3. As each bag of specimens was brought out of the chest, its specimen ID and any other identifying information (such as species name and collection location, if available) was entered into the inventorying log book. Photocopies of the log book pages were stored in two different buildings.
4. Every logged specimen was be re-sorted as being fish, bird, or other, and put into chests with other specimens of the same classification. Chests containing re-sorted specimens were given new numbers, and chest numbers corresponding to each class (bird, fish or other) will be noted. The re-sorted and filled chests will be resealed in presence of at least two members, and the chests were shipped back to cold storage unless otherwise required by specimen batching and processing.
5. Information from inventorying log book was entered into an excel spreadsheet, and double checked independently. Any mismatches were corrected.
6. Reviewed and approved inventories of opened chests were matched against log-book entries to tally all specimens with database entries. Water and Sediment samples were inventoried similarly, with the exception that chests were already shipped directly to the laboratory from the expedition, so the inventorying of chests arriving from cold storage did not have to take place.

III. Shipment of Processed Specimens

Processed specimens were shipped to INEEL or Vanderbilt University (see figures 1 and 2). The shipping process involved the following steps for documentation of shipped samples and notification:

1. The email notification of shipments included: name of sender, name of primary receiver, date of shipment, estimated time of arrival of shipment, list of specimens contained in package, and type of analysis requested.
2. A telephone call was made to a designated person at the receiving laboratory to inform them of the shipment and expected date of arrival.
3. On receipt of shipment, the receiving laboratory sent an electronic confirmation of arrival. The confirmation included a formal acknowledgement that the receiver was in custody of the specimens listed in the COC forms. Upon receipt of samples, the Lab Manager
    a. Opened the container and obtained the chain-of-custody form.
    b. Verified the contents against the chain-of-custody form.
    c. Checked and noted the condition of the samples before storage. Date of receipt and sample ID were noted in laboratory log book.
    d. Stored samples in the sub-zero freezer.
    e. Photocopied chain-of-custody form and placed it in log notebook. Filed the original in the Lab Manager office.
4. Additional individual laboratory chain-of-custody forms for drying/ashing, radiochemistry, alpha spec, and ICP-MS contained the following information:
    a. ID of original sample – e.g. S-F-1
    b. New ID of sample –included original ID first then e.g.
        i. S-F-1-DA (dried and ashed)
        ii. S-F-1-R (radiochemistry)
        iii. S-F-1-AS-Am (Alpha Spec Americium) or AS-U or AS-Sr
        iv. S-F-1-MS-Am (ICP-Mass Spec Americium)
    c. Date the sample is processed
    d. Brief description of analysis performed on the sample[2]
5. All electronic confirmations were saved and organized in an electronic tracking system developed for this project.
6. Specimens sent to Vanderbilt University for gamma analysis were not destroyed on analysis. These re-analyzable specimens were returned to Rutgers University with a new COC form. The laboratory (VU laboratory at this point) followed the notification process as described in steps 1-4.

---

[2] Notes

- All laboratories in the Vanderbilt CEE program were accessible only by cleared lab personnel with electronic or mechanical key access. Samples left for periods, for gamma counting, drying, ashing, chemical procedures, etc. remained in locked rooms.
- All analyses for Alpha Spec and Mass Spec were destructive, and no sample remained after analyses.
- Additional samples analyzed only for gamma emitters were held in long term freezer storage, with original sample identifiers intact.

IV. Distribution of Analytical Results

Laboratory analysis results were subjected to QA/QC procedures. After that, data were released to Data Management component for incorporation into database and distribution to assigned CRESP members for analysis.

Appendix 1

STRUCTURE OF RELATIONAL DATABASE

by Vikram Vyas and Yuri Mun

OBJECTIVE: To organize electronic data on specimens and their lab analysis results in an integrated database.

BACKGROUND: The biological specimens collected during the Amchitka expedition were assigned individual as well as composite IDs. The specimens were re-coded to mask location and species specific information when specimens were processed and sent to laboratories for analysis. An electronic database was used to organize all information pertaining to biological specimens and the corresponding laboratory analysis results.

METHODS: An interactive electronic database integrated information for every step of the project, from specimen collection to laboratory analysis. The principal technical features were the development of a relational database query engine based on specimen IDs to keep track of individual and composite specimens, use of geographic coordinates collected by GPS locators to identify collection locations, and the use of GIS software to place the data in a geo-spatial context. The information was evaluated through quality assurance steps prior to its assimilation in the database, and the quality evaluation status of laboratory analyzed information was tracked through the use of pre-designated flags. All information was entered and stored in Microsoft Excel spreadsheets, and the excel spreadsheets were also entered into a Microsoft Access database to permit querying.

RESULTS: The database was used initially for reconciling logged inventories and chest inventories of specimens. Thereafter, the organized information was used in identifying and mapping locations of specimens by species, and in identifying individual specimens that went into composites sent for laboratory analysis. In case of laboratory analysis results of particular interest, the database was used to identify other specimens collected in the same location. The database permitted listing of number of composites sent for laboratory analysis by species, type of analysis requested, and sample size. Finally, and most importantly, the database enabled the integration of original and coded specimen ID information, so that all information pertaining to original specimen IDs could be related to laboratory analysis results that referred to coded IDs. This, in turn, enabled production of tables that listed all laboratory analysis results by original as well as coded IDs. Such tables were used for Quality Assurance as well as analysis of final results.

Figure 1 below shows the conceptual structure of the database.

Figure 1. Conceptual Structure of Integrated Amchitka Database