United States Department of the Interior

U.S. GEOLOGICAL SURVEY
Upper Midwest Environmental Sciences Center
2630 Fanta Reed Road
La Crosse, Wisconsin 54603

15 July 2003

To: LTRMP Fish Component Staff

From: Brian S. Ickes

Subject: Database reforms

We will shortly be initiating several database reforms within the fisheries component of the LTRMP that were discussed and agreed upon at our annual component meeting in La Crosse last spring. The overall goal of these reforms is to make our data more amenable to analytical efforts, data sharing efforts, and future scientific reviews of the component. Below, I outline the planned reforms that will directly affect you, provide my rationale, highlight actions that are going to be taken, and suggest protocol changes that will result in both stricter adherence to acceptable centralized database use, while also permitting de-centralized flexibility for soon-to-be unsupported items. Several recent changes in how we do business (e.g., shift from data entry contractor to electronic data capture in the field; an increased emphasis on analytical projects; development of alternative ways to serve our data), have provided the incentive and ability to begin implementing these reforms. If you have any questions or concerns, please do not hesitate to contact me directly.

Brian S. Ickes

Issue 1: Bycatch in core monitoring collections

Invertebrates

Issue: Our sampling methods are not standard for invertebrates and in no way even approach methods optimized for their detection and enumeration. Thus, while recorded as bycatch, certainly huge and unknown sampling error renders these observations nearly useless. Their existence in our central database requires a very real cost that cannot be justified in my opinion. Moreover, their existence creates opportunities for analytical errors, requires developing and maintaining superfluous programming code, and creates numerous problems as we move forward with efforts to serve our data to program partners and the public using novel and new technologies.

Action: I have decided to remove all invertebrate data from our central database and to no longer support the inclusion of invertebrate observations. Table 1 presents all invertebrate codes presently found in our level 1 database. Each station may choose to download invertebrate observations from their station, or all invertebrate observations for the entire program for local archiving. Invertebrate codes are not supported in the electronic data entry application for either core monitoring or special projects. I recognize that some taxa are endangered and that there is some value to documenting the existence of particular taxa, regardless of the sampling methods used. Thus, if you decide to use our protocols and data applications to continue recording invertebrate observations, you still may.

Protocol: Protocol for doing this is as follows. Record observations on a paper datasheet at the sample site - you cannot enter these electronically. Return to the office. Enter these data into the desktop version of the data entry application we provided last fall. Maintain and archive these data under state-specific protocols. I will allow until August 31 for everyone to download what he or she wants. On September 2, 2003 these data will be permanently removed from our central database.

Turtles

Issue: Turtles also exist in our collections as recorded bycatch. However, I would argue that some of our collection methods are in line with accepted methods for their detection and enumeration. Thus, there is inherently greater value in these data than the invertebrate data. On the other hand, their inclusion in the level 1 database creates problems similar to those for invertebrates.

Action: I have decided to remove all turtle observations from the level 1 browser and place them in a new special project (TURT). All future turtle observations will be entered into this special project. The electronic data entry application permits turtle codes in the special projects portion of the application but not in the core monitoring portion, so you will be able to enter these observations at the point of collection, though you will need to "back out" of the core monitoring data forms and launch the "special projects" portion of the application. This will permit us to separate turtle from fish within the level 1 database while also providing central archiving of turtle bycatch. The fish and turtle collections, while physically separated within the database, can be relationally linked by barcode at the time of analysis if necessary. Public access to these data in the future will need to be mediated by a component specialist as they will no longer reside in our publicly served database (level 1). Unfortunately, this is not likely to change unless a resounding need is heard (would require the development of a separate browser interface and maintenance of a separate Oracle database for public access).

Protocol: For all project M- collections, turtles are to be entered separately from the fish observations and placed into the special project TURT. Barcodes will be cross linked, so you may either re-enter all site data in the special project fields or enter just the barcode and cross link to the fish data at the time of analysis to obtain site data. Specific step-by-step protocols will be drafted once the database conversion is completed. Fish-oriented special projects should record turtle data in the special project code for that project.

Issue 2: Database coding standards

Project codes

Issue: We have a highly advanced database and utilities for enhancing, supplementing, and serving data from this database. I'd would like to see it used whenever possible when it is beneficial to a special project. However, it must also be used responsibly. Irresponsible use can quickly degrade the usefulness and high quality aspects of this database. The fish component presently has more special projects than all of the other components combined and I've learned that 1) metadata are rare on these projects, and 2) some project codes are being used by more than 1 project. We need to straighten this out ASAP, if for no other reason than we will get slammed in scientific reviews of our procedures, data, and products.

Action: The transition to electronic data capture permits me to impose some central control on this issue. Andy has been compiling MetaData from past projects and fixing some of the problems alluded to above. All new projects must be assigned a new special project code by me. I do not wish to make judgments about the merits of any special project – I just need to know what is in the database so the problems identified above do not reoccur. Thus, I only ask that basic MetaData be submitted prior to initiating the project (who, what, where, when, why and how). A new special project code will added to the lookup tables in the electronic data entry application for your use.

Protocol: Inform me of your intent to initiate a special project. Include information on who is in charge of the project, who is funding the project, why the project is being implemented, where the sampling is to occur, what the project entails, the time frame for the project, and how it is being implemented – these are database fields on my end, so be brief. I will assign a new special project code and append it to our MetaData tracking database and have Dave Hansen add it to the electronic data entry application for you to use.

Fish species codes

Issue: It is apparent to me that many fish species codes have been added over the years, often for intermediate levels of identification (e.g., taxonomic genus). I am uncertain of the process in the past or the degree of control, but as we move forward in the future (e.g., analytically and in new ways to serve our data), some issues become apparent. First, a code is supposed to represent a species unit (with the exception of Unidentified Family codes, but these are easily filtered). Addition of intermediate levels of identification clouds this distinction. In the past, we have handled this through modifications to our SAS programs, constantly screening out these codes. However, opportunities for logic errors are ripe, as I have recently observed in many of the reports from this years analytical effort. Moreover, a naïve user of our data has a good chance of "getting it wrong". Lastly, new ways to serve our data that are currently under development would benefit greatly by a consistent notion of species within our database.

Action: Andy Bartels and I have already fixed codes that resulted from transcription errors over the years. These had been handled in past within the SAS code, but now are permanently fixed in the Level 1 database. Every taxonomic family has a code now as well (e.g., U-IC – Unidentified Ictaluridae). We collapsed all genus level identifications into the appropriate Unidentified Family code and we will no longer support new codes for intermediate levels of identification – it's species or family. We did retain the UNID code for entirely unidentified fish, and YOYF for unidentified young of the year fish and larval fish. Also, NFSH has been retained for "No Fish". If you wish to enter a genus level identification, you may do so in the comments field after recording the data under the appropriate Unidentified Family code.

Protocols: Fish codes are now "hard wired" in the electronic data entry application for M- collections. We recognize that a truly new species may be encountered that is not hard wired. We accounted for this by providing a new species code "NWSP". Record observations under this code and if a positive or probable identification is possible, record this in the Comments field. Be sure to preserve the specimen for vouchering. At the end of the sampling season, we will pull out all NWSP observations, review the comment fields, and generate and assign the observations to a new species code particular to the newly observed species. The new code will be added to the lookup tables for the electronic data entry application so that it is available for the next year. Thus, NWSP is a temporary dumping ground for specimens that are new species to our program. This will allow us to easily update

and refresh lookup tables for the data entry application and will permit us to maintain accurate MetaData records.

Table 1. Invertebrate codes presently residing in the level 1 LTRMP fish database. Observations on these invertebrates will be permanently removed from the Level 1 LTRMP fish database and no longer supported as of September 2, 2003.

Oracle Fish Codes	Common Names
ASIA	Asiatic clam
CSFY	Caddis fly
CYFH	Crayfish
DBFY	Dobson fly
DGFY	Dragon fly
DMFY	Damsel fly
GLSH	Glass shrimp
HHWM	Horsehair worm
HXFY	Hexagenia fly
HXMY	Hexagenia mayfly
LECH	Leech
LICE	Gill lice
MDGS	Midges
MQTL	Mosquito larvae
MYFY	Mayfly
OHSH	Ohio shrimp
PILL	Pillbug
PLAN	Planorbid snail
PSEU	Pseudiron centrallis--riverene mayfly
SCUD	Amphipod
SPIR	Spiroid snails
STFY	Stonefly