

PD/A CRSP Central Database

Interim Work Plan, Database Management

Doug Ernst and John Bolte
Department of Bioresource Engineering
Oregon State University
Corvallis, USA

(Printed as Submitted)

Introduction

This document discusses the Central Database status at time of transfer to Oregon State University and the current Database status as of December 1996. Publication of the Database on the Internet and through additional collaborative efforts is also discussed. Data table formats, definitions, and guidelines for submitting new data are discussed under Central Database Organization and Guidelines for Data Submission (1996). For a comprehensive review of the mission, mechanisms, and use of the Database, see the Final Report by the Special Committee on Database Management (Batterson et al., December 1991).

Prior Status of Database

The PD/A CRSP Central Database was maintained by the Management Entity from 1985 until Spring 1993, when it was transferred to Kevin Hopkins at the University of Hawaii at Hilo. The Database was received by John Bolte (DAST PI) and Doug Ernst (Database Manager) at Oregon State University from Kevin Hopkins in May of 1996. Over the period 1985-1996, the Database migrated through a series of computer operating systems and database software. When received at OSU, the Database contained data from 82 experiments, which were completed under Work Plans 1 through 7 from 1983 to 1995. The data were organized by experiment (site and time) and by data type (e.g. weather, water, fish) into 720 separate files, maintained in dbf format using FoxPro database software.

Considerable re-organization and editing of the Database was required to implement relational data structures and provide efficient mechanisms for data access and publication. Most of these problems had been recognized by the prior Database Manager and were brought to the attention of the new Database Manager when the Database was

transferred. Prior problems with the Database included:

- Lack of descriptions of experimental treatments, and thus lack of a corresponding fish production methodology context from which a Database user could identify and extract specific data;
- Lack of relational data organization, i.e. linking of associated data through common primary data;
- Existence of partially or fully duplicate records, i.e., data redundancy and conflict;
- Some lack of consistency in pond names, both within data sets of single experiments and among multiple experiments at a given site;
- Redundant and undefined names for some fish stocks and pond application materials;
- Poor organization of water quality data regarding time and depth of water samples;
- Inadequate pond area data required to interpret fish stocking numbers; and
- Incomplete data checking for reasonable range values.

In addition, the Database required better linkage to the greater PD/A CRSP literature and information base, through cross referencing and shared data files. Specific areas for improvement included:

- Information regarding principal investigators responsible for specific experimental data sets, to be used for data citations and referrals
- Site and facility descriptions and physical data, to be used in conjunction with experiment treatment specifications
- References to PD/A CRSP literature (Work Plans, Technical Reports, Annual Reports, etc.) and other external publications, to be used to augment experiment treatment descriptions, provide research objectives and context, and provide experimental results and discussions.

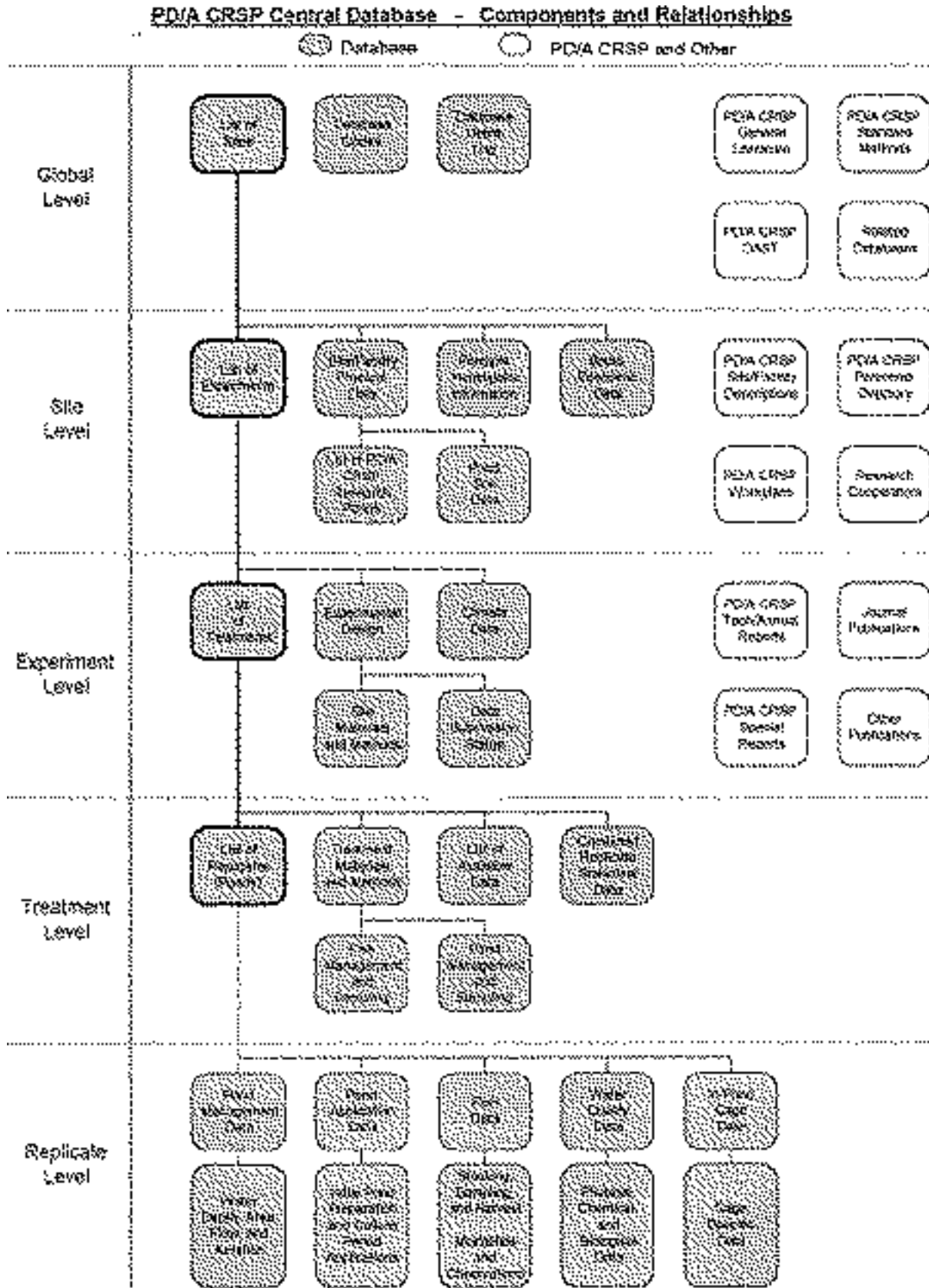


Figure 1. PD/A CRSP Central Database—Components and Relationships.

Current Status of Central Database

The Database is now managed under Microsoft Access and consists of one computer file (crsp.mdb, approx. 60 MB) containing multiple data tables (Figure 1; also see Central Database Organization and Guidelines for Data Submission). Data tables are made up of *records* (rows) and *fields* (columns) of data cells, analogous to a computerized spreadsheet. These data tables are related through *relational rules* and *primary keys* that, together with data field rules, prevent entry of incompletely defined, duplicate, or conflicting data. A list of all experiments currently in the Database is given in Table 1.

While efforts to solve problems with the Database continue, major problems have been eliminated. Through the use of data submission and data entry protocols, past problems will not be allowed to redevelop. New data submissions for the Database will not be accepted until experimental treatments are defined. Efforts that continue to better define experimental treatments include site and personnel information, and references to PD/A CRSP and other publications. Missing experiments from past Work Plans and the status of data submission from the current Work Plan remain to be determined. Investigators are encouraged to review the Database experiment list (Table 1) and see that all

Table 1. List of experiments in PD/A CRSP Central Database (total 82).

Site Name	Site Country	WP	Exp	Date Start	Date End
Abbassa	Egypt	07	1A	6/30/93	11/25/93
Aguadulce	Panama	01	01	2/6/84	5/18/84
Aguadulce	Panama	01	02	7/26/83	12/13/83
Aguadulce	Panama	02	01	1/30/85	5/3/85
Aguadulce	Panama	02	02	7/19/84	12/19/84
Aguadulce	Panama	03	01	3/18/86	5/30/86
Aguadulce	Panama	03	02	7/14/86	11/23/86
AIT	Thailand	04	02	3/8/88	9/20/88
AIT	Thailand	04	05	5/16/88	10/12/88
AIT	Thailand	04	06	12/20/88	5/22/89
AIT	Thailand	05	05	9/29/89	3/22/90
AIT	Thailand	05	09	12/4/91	5/6/92
AIT	Thailand	05	10	11/26/90	4/25/91
AIT	Thailand	06	01	9/10/92	1/8/93
AIT	Thailand	06	03	5/30/91	10/31/91
AIT	Thailand	06	04	12/6/91	7/2/92
AIT	Thailand	07	03	9/15/93	12/14/93
AIT	Thailand	07	05	12/1/94	5/4/95
AIT	Thailand	07	06	1/21/94	6/21/94
Ayutthaya	Thailand	01	02	8/1/84	1/4/85
Ayutthaya	Thailand	02	01	2/1/85	6/29/85
Ayutthaya	Thailand	02	02	8/3/85	12/27/85
Ayutthaya	Thailand	03	01	2/7/86	7/7/86
Ayutthaya	Thailand	03	02	8/6/86	12/29/86
Ayutthaya	Thailand	04	01	2/2/87	7/1/87
Ayutthaya	Thailand	04	03	2/5/88	6/7/88
Ayutthaya	Thailand	04	04	10/7/88	3/7/89
Ayutthaya	Thailand	05	02	10/12/89	3/8/90
Ayutthaya	Thailand	05	03	5/10/91	9/9/91
Ayutthaya	Thailand	05	04	10/25/90	3/21/91
Ayutthaya	Thailand	05	08	10/9/91	3/19/92
Ayutthaya	Thailand	06	05	6/11/92	11/13/92
Ayutthaya	Thailand	06	06	11/11/93	7/20/94
Ayutthaya	Thailand	06	07	1/21/93	12/9/93

data are submitted through Work Plan Seven (Sept. 1, 1993 to Aug. 31, 1995). Also remaining to be accomplished is determination of the completeness of the data that have been submitted and notification to investigators regarding conflicting data that have been deleted by the Database Manager.

It is important to realize that the pond-specific fish performance and water quality data contained in the Database is relatively meaningless without knowledge of how the ponds were managed (i.e., experimental treatments). It is also useful to know how ponds were organized regarding experiment-treatment-replicate hierarchies. There was some use of pond treatment codes in the Database, but these treatment codes were only partially used and defined. The PD/A CRSP Work Plans and Technical/Annual Reports were of limited use for defining treatment specifications in the Database, especially after Work Plan 3, given their superset relationship to the Database subset and lack of linking references.

To generate experiment treatment specifications to the best degree possible, a multistage procedure is being used: 1) compile all existing treatment data (Database and PD/A CRSP literature); 2) compile all fish stocking and pond application data in the Database and summarize as rate values; 3) develop a table of pond specific treatment specifications; and 4) have investigators review treatment specifications for completeness and accuracy. For development of fish stocking and pond application treatment specifications from the Database, the Database was first searched to find all ponds (sometimes cages) associated with each experiment and then searched to compile all fish stocking and material application data for each pond that was found. This treatment data (Database table ExpTreat) currently shows a grand total of 1311 replicates for the 82 experiments currently in the Database, with an approximate average of four treatments per experiment and four replicates per treatment. Of these 1311 replicates, it was found that a total of 226 had no fish stocked, and a total of 355 had

Table 1. Continued.

Site Name	Site Country	WP	Exp	Date Start	Date End
Ayutthaya	Thailand	06	6R	1/26/94	3/23/95
Ayutthaya	Thailand	07	6R	11/16/94	4/19/95
Bogor	Indonesia	01	01	6/19/84	11/12/84
Bogor	Indonesia	01	02	10/31/83	3/20/84
Bogor	Indonesia	02	02	1/10/85	6/10/85
Bogor	Indonesia	03	01	4/9/86	9/12/86
Bogor	Indonesia	03	02	10/14/86	3/12/87
Butare	Rwanda	01	01	2/26/86	7/22/86
Butare	Rwanda	01	02	10/1/86	2/26/87
Butare	Rwanda	03	01	7/9/86	12/4/86
Butare	Rwanda	03	02	12/18/85	5/15/86
Butare	Rwanda	04	01	1/1/99	1/1/70
Butare	Rwanda	04	02	6/23/88	12/8/88
Butare	Rwanda	04	03	5/5/89	9/28/89
Butare	Rwanda	04	04	12/29/88	3/30/89
Butare	Rwanda	05	01	10/30/89	3/20/90
Butare	Rwanda	05	03	7/31/90	1/31/91
Butare	Rwanda	05	07	8/24/91	1/3/92
Comayaga	Honduras	01	01	1/12/84	6/12/84
Comayaga	Honduras	01	02	7/11/84	12/10/84
Comayaga	Honduras	02	01	1/16/85	6/15/85
Comayaga	Honduras	02	02	7/26/85	12/23/85
Comayaga	Honduras	03	01	2/7/87	7/8/87
Comayaga	Honduras	03	02	6/5/86	11/5/86

no applications of feed or fertilizer. Whether these results represent actual pond management or data errors/omissions needs to be determined. It was also found that pond areas required to convert total fish stocking numbers to fish per unit area were inadequately reported (Database table PondSpecs). Investigators are asked to submit any past pond depth-area data not submitted to the Central Database and to bring all pond depth-area data up to date with new submissions to the Central Database.

The need to add data tables for sociological and economic data is recognized. Appropriate investigators will be contacted to determine what this socio-economic data would consist of and what is currently available. Collaboration will also be sought with the PD/A CRSP Management Entity to coordinate and standardize data management regarding aquaculture research sites and facilities, technical and annual report references, external publication references, and

investigator references. Hopefully, a Database user can easily determine all associated publications (CRSP and external) of a given data subset at the time the data are queried. Investigator reference data are required to properly cite data subsets and to provide contact information to interested Database users, analogous to the information given in journal articles.

Central Database Publication

A user and investigator interface to the Central Database is now provided at the designated Internet Web site of the Database ([http:// biosys.bre.orst.edu/crspDB/](http://biosys.bre.orst.edu/crspDB/)). This Database interface utilizes Cold Fusion software under a Windows NT operating environment. Tabular data presentation is available and graphical data presentation is under development. A log maintained at the Web site is used to document Database users, including optional contact

Table 1. Continued.

Site Name	Site Country	WP	Exp	Date Start	Date End
Comayaga	Honduras	04	01	8/11/88	12/20/88
Comayaga	Honduras	04	02	2/3/89	7/3/89
Comayaga	Honduras	05	01	7/27/89	12/21/89
Comayaga	Honduras	05	02	2/13/90	6/19/90
Comayaga	Honduras	05	05	8/8/90	1/3/91
Comayaga	Honduras	05	06	2/26/91	7/30/91
Comayaga	Honduras	06	01	9/5/91	2/4/92
FAC	Philippines	06	10	11/15/91	4/15/92
FAC	Philippines	06	11	7/1/92	11/4/92
FAC	Philippines	06	12	1/7/93	5/9/93
FAC	Philippines	07	09	5/31/94	10/15/94
FAC	Philippines	07	11	7/21/93	11/22/93
FAC	Philippines	07	13	11/22/94	4/24/95
Gualaca	Panama	01	01	2/15/85	5/21/85
Gualaca	Panama	01	02	7/8/85	12/2/85
Gualaca	Panama	03	01	1/27/86	6/25/86
Gualaca	Panama	03	02	7/21/86	12/9/86
Iloilo	Philippines	01	01	2/1/84	7/3/84
Iloilo	Philippines	01	02	7/11/83	12/12/83
Iloilo	Philippines	02	01	11/24/84	5/7/85
Iloilo	Philippines	02	02	8/16/85	2/28/86
Iloilo	Philippines	03	01	3/17/86	7/24/86
Iloilo	Philippines	03	02	9/3/86	11/13/86
Nong Sua	Thailand	01	01	1/1/99	1/1/70

information, objectives in data use, and comments on data use. The Database Web site is linked to the PD/A CRSP Web Site and to other aquaculture-related web sites. For intensive users of the Database, the entire Database could also be made available on electronic media, e.g., CD or 100 MB Zip Disk, given access to required hardware at the Database and user sites.

In the past, the Central Database has served mainly as a data repository with relatively few requests for data (about 30) as of December 1996. This lack of use by the aquaculture community was likely due to a combination of factors, including lack of awareness, difficulties in Database access, and lack of the necessary Database infrastructure to facilitate the search and extraction of specific data. Certainly, the Database problems itemized above would have been a major hindrance to anyone using the Database. If these assumptions are true, then publication of the Database at a Web site with a user-oriented interface should show clear improvement in the utilization of the Database. It may also be necessary, however, to better advertise the Database Web site beyond the PD/A CRSP community. A particularly under-utilized and promising audience is the aquaculture education community.

In addition to the Database Web site, the Database will be available at a world-wide environmental data Web site maintained by the Consortium of International Earth Science Information Network (CIESIN). Inclusion of some areas of the Database in FISHBASE, maintained by ICLARM/FAO, is being investigated. Potential collaboration with the aquaculture database maintained by the Network of Aquaculture Centers in Asia Pacific (NACA) will also be investigated. Other related databases, such as the Tropsoils CRSP database and the SADC Small Water Body database, will be contacted for the purpose of establishing a network of aquatic related databases and globally standardized data reporting.

The data search strategy supported by the Database Web site interface is based on a site, time, production-methods approach for defining and extracting data subsets. In this context, the "experimental treatment protocols" of investigators are analogous to the "fish production

methods" of fish culturists. The components of this methods-based database query include 1) experimental site; 2) inclusive dates; 3) fish/shrimp species; 4) fish/shrimp stocking density (fish/m²) and existence of polyculture; 5) initial fish/shrimp size (g); and 6) pond application fertilizers and feeds, including application frequency (days) and rates (kg/ha/wk).

These management components may be specified and applied in any order, whereby a single set of specifications, or through a series of iterative refinements, a user can define and extract a given subset of data. Investigators may go directly to a specific study by simply specifying site name and inclusive dates.

This approach to data queries provides the Database user with the ability to extract and compare any combination of fish production treatments. This ability is a unique characteristic of the PD/A CRSP research effort, with great potential for leveraging the usefulness of this data beyond report and journal publication. Further work is required to provide statistical tools for the user, including statistical summaries (e.g. minimum-maximum, mean, and standard deviation) and treatment comparisons (e.g. analysis of variance).

Principal investigator reference information is considered critical to properly acknowledge data subsets extracted from the Database. Data citations will appear automatically as users extract specific data sets, with the following format:

Pond Dynamics/Aquaculture Collaborative Research Support Program Central Database, (Diskette or Internet). (Inclusive dates of data used). Bioresource Engineering Dept., Oregon State University, Corvallis, OR USA. Available: Research country and site, principal investigator(s), inclusive dates of data used.

It is hoped that publication of data in the Database will be held in the same regard as its publication in reports and journals, and that equal incentives for investigators will apply. Publication of data in the Database should be viewed as an opportunity for additional research outreach, impact, and recognition.