

2. DATA INVENTORY STRATEGY

2.1 Data types and data sources

The data resources for the Galveston Bay system can take many forms, which we have categorized as: point observations, time series, line series, areal delineations, anecdotal and regional statistics. Specific examples of these data forms are given in Table 1. Formats of data, i.e., the physical forms of the data set, are equally varied, as summarized, with examples, in Table 2. In addition to the form and format of data, we must also consider the range of sources of data and the motivation for data collection, presented in Table 3. Both the form and format of data sets are important in characterizing the data for inventory purposes, because they determine the potential utility of the data for a given scientific purpose, and the effort necessary to manipulate the data. The purpose and source of the data, as exemplified by the entries of Table 3, are more pertinent to the archival practices and accessibility of the data. As will be seen later, data from those programs that are implemented for the first purpose, routine monitoring, are by far the most accessible. In general, the entries of Table 3 are in decreasing order of accessibility and increasing probability of data loss.

This project focused on "raw" data, i.e., the original observations, imagery or measurements. This is in contrast to reduced or summary data, i.e., data which has been averaged, composited, or processed in some manner. Most literature references, especially in the formal scientific literature, employ only reduced data. Further, the increasing practice in the grey literature has been to present reduced data. This is due to a combination of space limitations, convenience and fashion. In the Galveston Bay NEP Data Inventory project, we have sought the original raw measurements whenever these still existed. The reason for this is simple. Any type of processing focuses upon one aspect of data interpretation at the expense of another. One may exhibit general trends in a variable by displaying the long-term time averages, or one may exhibit the horizontal structure in a variable by averaging measurements in the vertical. In the first case, information about time fluctuations is sacrificed, in the second case details of stratification are lost. Later researchers may be concerned about either of these, and would therefore require access to the raw measurements.

One of the major classes of data sources is the unpublished holdings of agencies and individual researchers. The approach to this class of data was stepwise, starting with inquiry letters and proceeding to direct contact; visits by the Principal Investigators (PI's) to inspect and assess holdings; completion of the inventory, assessment of data perishability and acquisition of copies where appropriate. For the key state and federal agencies (most of which are participants in the GBNEP), the strategy (proposed by the GBNEP) was to identify a point-of-contact in that agency who would facilitate the location of data holdings and make the necessary internal arrangements for the PI's to visit and inventory the data. Individual researchers posed a greater problem, in that there were many more of them, individually with smaller data sets, difficult to locate and contact, and frequently uncooperative.

Table 1: Forms Of Data

TYPE	EXAMPLE
POINT OBSERVATIONS	grab samples, soundings, temperature/salinity measurement, trawl catch
TIME SERIES	streamflow records, tide scrolls
LINE SERIES	cross-sectional profiles, scanner imagery
AREAL DELINEATIONS	maps, aerial photography
ANECDOTALS	event descriptions (fish kills, oil spills, hurricanes), strandings, historical references, oral recollections
REGIONAL STATISTICS	population profiles, bird rookeries, economic activity

Table 2: Data Formats

TYPE	EXAMPLE
OPEN LITERATURE	books, journals
GREY LITERATURE	technical reports, project studies, data reports
FILE DOCUMENTS	unpublished manuscripts, internal memoranda
TRANSIENT LITERATURE	newspapers, diaries, historical collections
DATA TABULATIONS	printouts, computer-encoded data bases, tabular summaries
ORGANIZED DATA ARCHIVES	indexed maps, aerial photos
RAW DATA	field sheets, strip charts, trip logs

Table 3: Data Objectives And Sources

PURPOSE	SOURCE
ROUTINE MONITORING	Federal agency (USGS, NWS/NCC)
OBJECTIVE-SPECIFIC MONITORING	Federal and state agencies
ENFORCEMENT	Federal, state, regional agencies
PROJECT-SPECIFIC INFORMATION	State, federal, private industry
SPONSORED RESEARCH	Universities, private industry, state agency (rarely)
PERSONAL RESEARCH	Universities, individuals (rarely)
DOCUMENTATION	Historians, journalists, individuals, private industry
INTEREST	Individuals

2.2 Project Approach

The Data Inventory project strategy consisted of several basic elements:

- (1) Simultaneous review of literature holdings, journals, and reports, and establishment of contacts with key agencies and researchers.
- (2) Early consultation with data-base specialists on selection of software and formulation of information system structure.
- (3) Subdivision and cross-referencing of information handling system (i.e., the Galveston Bay Data Inventory System) according to character of data: e.g., point observations, time series, line series, areal delineations, anecdotal.
- (4) Reliance upon point-of-contact for principal agencies, followed by systematic review of agency holdings, by discipline and geography.
- (5) Direct *personal* participation of PI's in contacts and data evaluation.
- (6) First-line reliance on letters, fax, photocopying and telephone.
- (7) Assessment of data perishability and initiation of appropriate action.
- (8) Matrix formulation linking STAC priority problems with subdiscipline data-types. Continuous re-appraisal of data coverage, quality, and interrelations vis-a-vis STAC priority problems.
- (9) Weighing of principle of diminishing returns versus criticality of data.
- (10) Documentation of sources, leads, and history as work progresses.

The significance of (1) is that the task of location proceeded simultaneously on several fronts: review of bibliographies and indexes; *direct* review of journals and reports (as opposed to computerized searches or published bibliographies); visits and contacts with likely sources. The purpose was to create, insofar as possible, a parallel activity rather than a serial, for maximal efficiency. At the same time, we assigned preliminary priority according to the age and anticipated inaccessibility of the information, so that the older, harder-to-find information was sought first. This led to the apparently paradoxical fact that the data sources nearest and most accessible by the PI's were contacted last. In fact, this represents a judgment of the requisite lag time in gaining access to the data. The greater the anticipated lag, the more lead time necessary.

The electronic product, the Galveston Bay Data Inventory System (GBDIS), began to be formulated at the outset of the project, hence (2) and (3) above. The obvious purpose was to distribute the effort of data entry throughout the project. An additional purpose was to identify and formulate data set features to be addressed before the actual appraisal process began. Thus, special-purpose appraisal forms

could be devised to facilitate the inventory process, speed up data entry, and minimize the possibility of error or omissions. Example data entry forms are shown in Figs. 1 and 2.

The point-of-contact approach (4) was very important to the project strategy. The GBNEP management committee recognized at the outset that considerable personnel time would have to be invested in tracking down data holdings in federal and state agencies. The labor time allocated for this project would be best concentrated on the actual data inventory process itself, rather than in the dissipative activities of identifying key personnel in the agencies, contending with archival procedures, submitting formal requests, and so forth. Considering that most of the key agencies with jurisdiction in the Galveston Bay area are participating in the GBNEP, it was proposed that each such agency designate an individual to serve as a point-of-contact. Ideally, this person should be fairly senior in the agency, so as to be familiar with agency procedures and personnel, and to be able to have the authority to encourage staff cooperation with the GBNEP. The PI's of this project would then work directly with the point-of-contact, as the interface to the agency.

It is important to emphasize (5), that all of this work was carried out by the PI's personally; no student help was employed. There were several reasons for this: to improve the responsiveness of the source agencies and individuals, to ensure accurate judgement of the quality and value of data sets, and to take advantage of the combined six decades of experience of the PI's in the Galveston Bay system. The principal activities of the PI's were to be: inquiry letters, direct contacts (apart from activity of the points-of-contact), on-site data appraisals, and preparation of data set reports. In addition, the PI's were personally responsible for the development of the GBDIS. Efficiency of prosecution was therefore an uppermost concern, hence the ordering of communication in (6): letters and facsimile are essentially parallel channels, while the telephone is serial, and much less efficient. Personal visits are the most inefficient of all, and were avoided until absolutely necessary.

An important property of a data set is its perishability, that is, whether it is reliably and permanently archived, exists elsewhere in duplicate, or could be subject to loss or discard. At the outset of the project, we anticipated that some data loss would have occurred and expected some data sets would be potentially susceptible to future loss (though we had no intimation of the scale of the problem). While the purpose of this project was the identification and appraisal of data, but not the acquisition of data *per se*, we felt an exception should be made for those data sets in imminent danger of destruction. Accordingly, (7) was identified as a specific strategy element, to allow the project the ability to actually obtain copies of data, or the actual data set itself, when considered endangered.

Figure 1. Example Data Inventory Form, General Information

GALVESTON BAY DATA INVENTORY PROJECT

GENERAL INFORMATION

1 GBNEP Data Ref. No:

2 Agency/Institution:
Name: _____
Address: _____
City: _____ State: _____ Zip: _____

3 Contact Person:
Name: _____
Address: _____
City: _____ State: _____ Zip: _____
Telephone No: () _____

4 Data Description:
Program Name: _____
Obj. of Progr: _____
Use of Data Coll: _____

5 Time Span of Data: From (MM/DD/YY): ___/___/___ To: ___/___/___ . Interval ___ Units ___

6 Data Coll Loc(s): Loc. Name: _____
WQ Seg. No(s): _____
Sta. Nos. _____

7 Type of Data:
Point Observ: _____
Analog Time Series: _____
Analog Line Series: _____
Areal Delineation: _____
Anecdotal: _____

GENERAL INFORMATION (Cont'd)

1 GBNEP Data Ref. No:

- 8 Source of Data:
- Open Literature: _____
 - Grey Literature: _____
 - File Document: _____
 - Transient Lit: _____
 - Data Tabulation: _____
 - Data Archive: _____
 - Raw Data: _____
 - Other: _____

9 Status of Data

a. Raw _____ Descrip: _____

b. Reprint - _____ Descrip: _____

c. Computerized _____

Database Name: _____

File formats: _____

Field Layout: _____

Software Applic: _____

Accessibility: _____

d. Data Products _____ Descrip: _____

e. Other _____ Descrip: _____

GENERAL INFORMATION (Cont'd)

1 GBNEP Data Ref. No:

10 Citation

a. Author _____

b. Year: _____

c. Title: _____

d. Journal/Report: _____

e. Volume (Number): _____

f. Pages: _____

g. Document location: _____

h. NTIS Number: _____

i. EPA document no: _____

j. Library call no: _____

k. Acc. no. in GBP Libr: _____

l. Other identifying nos: _____

m. Abstract: _____

n. Publication date: _____

11 Priority Problem(s) _____

**Figure 2. Example Data Inventory Forms,
Water Quality and Sediment Quality**

GALVESTON BAY DATA INVENTORY PROJECT

WATER QUALITY

1 GBNEP Data Ref No:

30 Sample, Survey Type

a. Frequency

Frequ. ____ Units ____ / Irreg. ____

b. Vertical Resolution

No. samples over depth: One ____ > One ____

31 Sample Handling

32 Lab Proc and Methods

33 Data Entry/Edit. Methods

34 Data Scrubb/Error Trap.

35 Parameters/Information

Use TNRIS codes attached

GALVESTON BAY DATA INVENTORY PROJECT

SEDIMENT QUALITY

1 GBNEP Data Ref No:

36 Sample, Survey Type

a. Frequency

b. Vertical Resolution

37 Sample Handling

38 Lab Proc and Methods

39 Data Entry/Edit. Methods

40 Data Scrubb/Error Trap.

41 Parameters/Information

Use TNRIS parameter codes

Data organization focused on the specific priority problems identified by the Conference. Strategy element (8) above approached this by assembling a matrix organization relating these priority problems to specific types of data required to address each problem. Table 4 displays this matrix by broad topic areas, based upon the statement of the GBNEP priority problems (Hightower, 1989, GBNEP, 1991). Each of these priority problems can be subdivided much more specifically, as has already been done by the subcommittees of the STAC (Hightower, 1989, GBNEP, 1991). Further, each of the broad disciplinary categories is itself subdivided into more specific areas, as shown in Table 5. Thus Table 4 is a highly compressed summary of a much more detailed matrix (which was further extended and revised as the project progressed). The holdings of target agencies were continuously related to the data areas in these arrays. This approach in no way delimited the needs and intention of the GBNEP to compile a comprehensive data base, but rather served as a means of ensuring that the data requirements of each priority area are addressed and continually considered throughout the project

We endeavored to follow every lead uncovered with resolve and persistence. This was tempered of course by (9), weighing of the principle of diminishing returns versus criticality of the data. For example, we have invested many hours in searching for the U.S. Bureau of Commercial Fisheries 1958-67 biological/water quality collections. On the other hand, we abandoned the search for the 1964-69 Texas A&M questionnaires of Houston Ship Channel industries when the files could not be located from either the research institute or the principal investigators. For reasons to be presented shortly, the principle of diminishing returns had to be invoked as the work progressed and the recovery effort increased.

**Table 4: Matrix Of Data Requirements By Discipline
Versus N.E.P. Priority Problem Areas**

PRIORITY PROBLEMS	DISCIPLINES					
	MORPH- OLOGY	HYDRO- GRAPHY	HYDRO- LOGY	BIO- CHEM- ISTRY	BIOLOGY & ECOLOGY	SOCIO- ECONO- MICS
RED/ALT OF LIVING RESOURCES	■	■	■	■	■	■
PUBLIC HEALTH ISSUES		■	■	■	■	
RESOURCE MANAGEMENT ISSUES		■		■	■	■
SHORELINE EROSION	■	■	■		■	