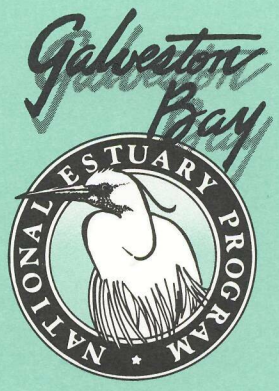
Galveston Bay Data Inventory



Galveston Bay
National Estuary Program
GBNEP-40
December 1991

Galveston Bay
Data Inventory

printed on recycled paper using soy-based ink

Galveston Bay
Data Inventory

Prepared by

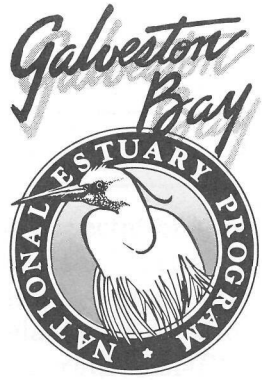
George H. Ward
Neal E. Armstrong

Interagency Contract 14-09013-0933
Center for Research in Water Resources
The University of Texas at Austin
Austin, Texas 78712

The Galveston Bay National Estuary Program

Publication GBNEP-40
December 1991

This project has been funded in part by the United States Environmental Protection Agency under assistance agreement # CE-006550-01 to the Texas Natural Resource Conservation Commission. The contents of this document do not necessarily represent the views of the United States Environmental Protection Agency or the Texas Natural Resource Conservation Commission, nor do the contents of this document necessarily constitute the views or policy of the Galveston Bay National Estuary Program Management Conference. The information presented is intended to provide background information for Management Conference deliberations in drafting of official policy in the Comprehensive Conservation and Management Plan (CCMP). The mention of trade names or commercial products does not in any way constitute an endorsement or recommendation for use.



Policy Committee
The Honorable Rodney Ellis, Chair
Texas Senate

Mr. John Hall
Chair, Texas Natural Resource Conservation Commission

Ms. Jane Saginaw
Regional Administrator, EPA Region 6

Mr. James Blackburn
Galveston Bay Foundation

Ms. Terry Hershey
Commissioner, Texas Parks and Wildlife Commission

Mr. Charles Miller
Past Chairman of the Board of Directors, Greater Houston Partnership

Mr. Charles W. Jenness
Chair, Texas Water Development Board

Local Governments Advisory Committee
The Honorable Ray Holbrook, Chair

Management Committee
Mr. Myron O. Knudson, Chair
Mr. Tom Remaley, Vice-Chair

Scientific/Technical Advisory Committee
Dr. Robert McFarlane, Chair
Ms. Teresa Battenfield, Vice-Chair

Citizen's Advisory Steering Committee
Ms. Sharron Stewart, Chair
Mr. Ron Embry, Vice-Chair

Galveston Bay Public Forum
Dr. Don Bass, Chair

Program Director
Dr. Frank S. Shipley

The Galveston Bay National Estuary Program

Texans increasingly express their expectations for a clean environment in terms of entire ecosystems. Until recently, our tendency was to view environmental problems in isolated pieces we could understand—indeed this view was institutionalized (and seemingly immortalized) in an elaborate mosaic of fragmented jurisdictions. The Galveston Bay National Estuary Program (GBNEP) is a forerunner in elevating hands-on management of coastal environments to the level of the ecosystem; and in doing so, is encouraging an integration of traditionally disparate institutions.

The GBNEP was established under the authority of the Water Quality Act of 1987 to develop a Comprehensive Conservation and Management Plan (CCMP) for Galveston Bay. The purpose of the CCMP is to address threats to the Bay resulting from pollution, development, and overuse. To address these threats, five years of work commenced in 1990, consisting of three phases: (1) Identification of the specific problems facing the Bay; (2) A Bay-wide effort to compile data and information to describe status, trends, and probable causes related to the identified problems; and (3) Creation of the CCMP itself to enhance governance of the Bay at the ecosystem level. The GBNEP is accomplishing this work through a cooperative agreement between the U.S. EPA (Region 6) and the State of Texas (administered by the Texas Natural Resource Conservation Commission.)

The structure of the GBNEP reflects a strong commitment to consensus-building among all Galveston Bay user groups, government agencies, and the public. The GBNEP "Management Conference" consists of six Governor-appointed committees with broad representation, totaling about one hundred individuals. Meetings of these committees are also open to the public, and public participation in policy-setting and in Bay management are considered strengths of the program. When submitted to the Governor of Texas in late 1994, the CCMP will reflect thousands of hours of involvement (much in the form of volunteer time) by individuals who in various ways use, enjoy, or help govern this vital coastal resource.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

The Galveston Bay National Estuary Program Data Inventory is not, and hopefully will never be, a finished product. Rather, this represents the first installment of Work-in-Progress. As the system is routinely used, other refinements and capabilities should be added to enhance its function and utility to researchers. As new data is acquired from Galveston Bay, and older historical data sets are (hopefully) discovered, the data base itself should increase.

This report documents the philosophy and initial effort in creating the Data Inventory. The deliverables for the project include this report and its companion documents, the Galveston Bay Data Inventory System software, and the data base itself. One companion report, the GBDIS User's Manual, describes the structure and operation of the Data Inventory System in much more detail than is appropriate here. While the data resource is summarized in this volume, individual data sets are described in the Data Set Report series in the Appendix (bound separately).

The Principal Investigators wish to acknowledge the efforts and energies of the many individuals, both private citizens and the employees of agencies and companies, who offered their assistance in this compilation. These persons generally share our perception of the historical and scientific value of raw observations and our concern for preservation of such data sets, and generously contributed their time to help search for the information compiled here.

Finally, with the reader's indulgence, we must anticipate and respond to a criticism of this report which will occur to fully half the readership. Please be assured that we realize that the Latin word *data* is plural for *datum*, and therefore should take a plural verb. On the other hand, *data* as an English word has acquired some currency as a collective noun, with a singular or plural construction according to whether it refers to the entities taken together or taken separately. These Principal Investigators are themselves divided on this point, but we have followed this latter usage rather than the strict Latin syntax. If the reader is offended by the occasional singular verb, as is one of the Principal Investigators, we would invite him/her to pencil in "data set" for "data" and all will be well. For the reader whose mind remains open, unlike one of the Principal Investigators, we suggest consulting the Second College Edition of the *American Heritage Dictionary* (Boston: Houghton Mifflin Co.)

GALVESTON BAY DATA INVENTORY

Principal Investigators:

Neal E. Armstrong
George H. Ward
The University of Texas at Austin

EXECUTIVE SUMMARY

In the past, a wealth of data has been collected from the Galveston Bay system, relating to the movement and quality of water, the biology of the bay, navigation, socioeconomics and fisheries, some of this information dating back more than a century. Therefore one of the early tasks in the Galveston Bay National Estuary Program (GBNEP) was to locate and inventory these data. The specific objectives of this project were to: (1) survey local, state, and federal agencies and other organizations for data sets; (2) prepare in a standard format written descriptions of data sets; and (3) compile the data inventory in an electronic, searchable, microcomputer-based information base. The product of the project should enable a researcher with a specific data requirement to determine where (and whether) the historical data exist and how to access them.

At the outset of this work, a data management system (DMS) had to be chosen to form the software basis for the inventory. There is a phenomenal selection in DMS products presently on the market. The most recent buyer's review available in 1989 listed 80 PC-based relational database management software packages. To sharpen the choice, thirteen criteria were formulated that the GBNEP DMS must satisfy (five of which had been specified by the GBNEP management and were required by the contract), dictated by the anticipated characteristics of the entries and the probable requirements of the users. Ultimately, dBase IV was adopted as the data inventory software.

The data resources for the Galveston Bay system take many forms, including point observations (such as grab samples), time series (streamflow records, tide scrolls), line series (cross-sectional profiles, scanner imagery), areal delineations (maps, aerial photography), anecdotals (event descriptions), regional statistics (bird rookeries, population profiles, economic activity). This variety of forms demanded a considerable flexibility--and therefore complexity--in the electronic Galveston Bay Data Inventory System (GBDIS). One of the major features of the GBDIS data-base structure is the use of multiple files, elements of which are "related" (i.e., logically identified for access and retrieval purposes). Because different types of data have different properties, their logic structures are different. Retrieval is accomplished by searching on field variables, perhaps constrained by user-specified relations, and by keyword textual searches of the title and abstract information in the data entries. This dual approach to retrieval allows both quantitative sorting of the information, as well as qualitative searching.

One of the more important retrieval fields is that specifying the locations within the system at which the data observations/measurements were made. We anticipate one use for the data inventory to be retrieval of data of a specified type pertaining to a specific region of Galveston Bay. Latitude-longitude coordinates were adopted as the basic position specification. This decision entailed a considerable effort in the data entry process: because relatively few data sets have the measurement positions specified by latitude and longitude, it was necessary to map these points and determine the coordinates ourselves. However, the generality and flexibility of this approach we believe justifies its employment.

There are many sources for data on Galveston Bay, including open literature, grey literature, file documents, transient literature, formalized data tabulations, organized data archives, and raw data. The task of location proceeded simultaneously on several fronts: review of bibliographies and indexes; *direct* review of journals and reports; visits and contacts with likely sources. All of this work was carried out by the project principal investigators, personally; no student help was employed. It is important to differentiate this project, whose objective was to inventory extant data from Galveston Bay, with the companion GBNEP project, to compile information on the Bay. The former (i.e., this project) focuses upon raw measurements, while the latter focuses upon the technical literature.

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 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.

In summary, the project proved to be far more complicated and time-consuming than originally envisioned. Several factors contributed to this:

(1) There proved to be a large number of data sources for Galveston Bay, but only a minority could be described as major projects (e.g., the TWC Statewide Monitoring Network, the Galveston Bay Project, the TWDB Bays & Estuaries Program, etc.), i.e. the data resource can be described as a few large projects and a great many small projects, which served to multiply contact time and logistics;

(2) The point-of-contact approach failed, requiring much greater time and effort of the PI's to find and gain access to agency data holdings;

(3) In general, the response of the data sources to our inquiries has been poor, necessitating multiple letters or calls, and requiring months (at best)

to finally gain access to data. In 1991, after the project was technically over, many were only then responding.

However, the dominant reason is that the management of older data--and by this we mean any data taken prior to 1980--is by-and-large a shambles.

The principal conclusions regarding the data resource for Galveston Bay drawn from the experience of this project are:

(1) Most of the data sets for Galveston Bay taken prior to 1980 are presently unavailable. The majority of this data appears to be irrevocably lost.

(2) When one considers that the data prior to 1980 comprises the vast majority of data taken in Galveston Bay ever, this implies that most of the data resource has vanished.

(3) The factors which have led to this loss of data are still operating today.

These conclusions apply primarily to data on the biological, water quality and hydrographic features of the system, which are the most important insofar as the GBNEP objectives are concerned, however they probably apply to other categories of data as well.

These conclusions of course must be qualified for specificity. For example, sediment quality data is of more recent concern, and has benefited from advances in analytical technology, so is in relatively good shape. Also, specific data collections with national archival procedures are well-managed, e.g. the historical mapping of the National Ocean Service and its predecessor agencies, and the data collection efforts of the U.S. Geological Survey. On the other hand, for many major and fairly recent data collection projects implemented by federal, state and regional agencies, the data are *totally missing*. Additionally, data sets, which had been entered on digital media, now only exist as one or a few *hard-copy* tabulations. Thus the utility of the data is severely truncated, and the effort invested in putting the data in a utilitarian format is lost. The situation is worse for research data of individuals.

The factors that contribute to this data loss include:

(1) problem-specific operation of most agencies, and the valuation of older information as "obsolete";

(2) low priority assigned to archiving and preservation of older data, and the general perception of archiving of information as an unwarranted expense;

(3) personnel turnover in the agencies combined with little or no documentation;

(4) agency instability, i.e. dissolution, merging and reorganization of an agency, as well as frequent displacement and relocation;

(5) natural calamities (fires, floods, hurricanes) in poorly protected housing;

(6) changes in data management technology, without upgrading of historical files;

(7) proprietary attitude toward data by individual investigators.

All of these are mutually exacerbating. All of these are continuing to operate and permit continued loss of data. In our view, the problem is critical.

1. INTRODUCTION AND OBJECTIVES

1.1 Significance and rôle of historical data in estuarine science

In estuarine management, as in science in general, we need basic information on cause-and-effect relationships operating in the estuary. How water quality affects the presence or absence of an organism, how that water quality is affected by injection of contaminants and by tide- or wind-induced transport, and what kind of controls on water quality can be instituted to promote or eliminate a particular organism are all examples of the application of cause-and-effect relationships to a specific management problem. In a very real sense, estuarine management is applied science: the effectiveness of management of Galveston Bay is dependent upon how well we understand the operative cause-and-effect relationships.

Two of the key elements of formal scientific method are *controlled experiments*, in which the effect of one (or a few) causal factors are isolated by holding constant all other sources of variation, and *replicability*, in which the experiment of one scientist is verified through duplication by other scientists working independently of the first. In the natural sciences applied to an estuary, in particular Galveston Bay, we are denied these.

An estuarine system, its water quality, fluid motion, sedimentary processes and biological communities, is an extensive, complex watercourse, governed by *extrinsic* factors of meteorology, hydrology, oceanography, waste discharges and the activities of man, and by the *intrinsic* interactions among all of these variables. Few of these are subject to any kind of control, for scientific or other purposes. Further, there is practically no prospect of replicating a set of measurements, because the precise combination of extrinsic and intrinsic conditions is never the same. One consequence of this is a degree of apparent "noise" in measurements under the same gross conditions. For example, the measured salinity at Atkinson Island, say, under superficially constant freshwater inflow, tides and meteorological conditions varies over a wide range, due to the myriad of extrinsic and intrinsic factors influencing the time signal of salinity at that point.

The rôle of a controlled experiment in a natural system like Galveston Bay is approximated (at best) by a combination of analytical methods applied to observations from the real system. These methods have the general objective of quantifying the association of one variable with another by procedures of statistical modeling and deterministic modeling (themselves each a formulation of hypotheses about the relation between the variable of concern and other potentially controlling variables). The success of these analytical methods is dependent upon two factors: the ingenuity and insight of the scientist, and an extensive observational data base from the estuary. The latter is indispensable for the former; observational data is the limiting nutrient, so to speak, for scientific analysis. What is lost by not being able to carry out a controlled experiment must be made up in multiplicity of observations. This multiplicity is necessary both to improve statistically the "signal-to-noise ratio" of whatever associations are to be detected, and to achieve a larger range of variation in the controlling factors, allowing a more reliable determination of quantitative dependencies.

In a system like Galveston Bay, observations are dearly won. Whether the object is fishery population or salinity structure or sediment quality, a considerable investment is needed to get an observer to where he needs to be and to perform the necessary field measurements. None of the principal management problems of Galveston Bay is amenable to a project of special-purpose data collection: such data collection would be far more costly than is feasible to support, especially in these days of competing demands for limited research monies. We are dependent upon the *collective enterprise* of data acquisition, to synthesize a comprehensive data base from the efforts of numerous agencies and researchers.

This situation is of course neither novel nor unique. Much of the progress in any of the natural sciences, especially the earth sciences and the ecological sciences, has developed from sifting and re-analysis of accumulated data. This accumulation of information requires three things: (1) documentation of methods of measurement, (2) integrity of the individual scientists, (3) preservation of the basic measurements. In estuarine science, (1) is generally achieved through the scientific literature. Fortunately, in estuarine science we have been blessed with (2), a high level of scientific integrity. The reasons for this are worthy of exploration, but would be a digression here. The limiting factor is frequently (3), a preserved and accessible base of measurements over a range of conditions and period of time. Ironically, as the information age has matured, and techniques of intense data collection have developed, the ability to preserve these observations has declined. Earlier, the professional journals served this purpose, but the cost of journal pages and the impetus for presenting only summarized or pre-digested data in technical papers has all but eliminated the journals as a repository of raw observations.

The need to locate and compile data resources for an estuary has been recognized as an essential, even critical prerequisite for development of the comprehensive management plan for a National Estuary Program. Therefore an early task in the Galveston Bay NEP (GBNEP) is this Data Inventory project. Specific requirements in a data base for cause-and-effect analyses are:

Long period of record to exhibit variation

Long enough monitoring period to encounter a range of conditions or configurations (e.g., streamflow monitoring)

Observations under different controlling conditions, to allow separation of cause and effect

Sufficient observational base to suggest and examine new hypotheses

Sufficient observational frequency to resolve temporal responses and controlling periodicities

Sufficient spatial density to resolve spatial variability

Rarely, if ever, will any single program of data collection satisfy these requirements. It will in general be necessary to combine the results from several or many programs, which in turn means that a researcher or manager will need some ready access to sources of data. Moreover, given the bewildering variety of measurements and observations included in the general term "data," some means of sorting and retrieving is mandatory.

1.2 Data inventory objectives for Galveston Bay

In the past, a wealth of data has been collected in the Galveston Bay system, relating to the movement and quality of water, the biology of the bay, navigation, socioeconomics and fisheries. Some of this information dates back more than a century. However, most of this data has been collected for specific purposes, by agencies or individuals with a narrow objective, sparsely in time or in limited areas of the system. The data have great potential value to the Galveston Bay National Estuary Program if they can be combined into a comprehensive data base yielding a historical depiction of the bay. The purpose of this project is to locate and inventory these data.

The specific objectives of the project are to:

(1) survey local, state, and federal agencies and other organizations for data sets related to the Galveston Bay Priority Problems List;

(2) compile and publish in a standard format a written report of complete descriptions for existing data sets; and

(3) compile the data inventory in an electronic, searchable, microcomputer-based data set index.

The general approach followed in this project was formalized in the Project Work Plan (Armstrong and Ward, 1990) submitted to and reviewed by GBNEP management, the Scientific and Technical Advisory Committee and the Management Committee. The Work Plan, delivered in draft form on 2 January 1990, included:

(1) list of agencies/institutions to be surveyed for data sets;

(2) outline of the proposed format and content of entries in the Inventory;

(3) recommended software for the electronic searchable index;

(4) proposed search logic for index entries, with example;

(5) completion schedule for the project.

Specifics of the project strategy and how these objectives were approached are given in Section 2.2 below. The structure and content of the Data Inventory and specific findings on the data resources for Galveston Bay are addressed in the remainder of

this report. The prospective user of the Data Inventory System should also consult the Data Inventory System User's Manual (Armstrong and Ward, 1991), a companion document to this project report, and should browse through the digital data base itself. Finally, one additional companion report, the Appendix, bound separately, contains a listing of the Data Set Reports generated by the project, which summarize the methods, extent, source and status of the data sets inventoried in this project.

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, anecdoctals 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, anecdoctals.

(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: ___

Anecdotals: ___

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

	DISCIPLINES					
PRIORITY PROBLEMS	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	■	■	■		■	

3. DATA MANAGEMENT SYSTEM

To implement the Galveston Bay Data Inventory System (GBDIS) portion of this study, it was necessary: to select software that matched the requirements of the system to be developed; to develop the structure of the database to accomplish the various types of retrievals desired; to write the code to permit data entry into the database not only by the Principal Investigators and their staff, but by the Galveston Bay National Estuary Program staff or others as well; and to write the code to permit data retrieval from the database using menus easily understandable to the lay public. These steps are described below.

3.1 Software

One product of this project is an extensive listing of available information on the Galveston Bay system. The character and treatment of this "information" base are discussed in the following section. It was formatted in a digitized data base, most efficiently accessed and manipulated with a data management system (DMS). At the outset of this work, it was necessary to select a DMS to form the software basis for the Data Inventory System.

There is a phenomenal selection in DMS products presently on the market. Further, the available software is reviewed regularly by various periodicals and ranked according to generally desirable properties, including speed of execution of basic operations such as reads and sorts, simplicity of execution, mathematical function capability, and data fields accessible. The most recent buyer's review available at the outset of the project (Badgett et al, 1989) lists 80 relational PC-based database management software packages. Many of the DMS product features are not relevant to the intended use in the Galveston Bay Data Inventory System (such as speed, since we do not anticipate the GBDIS as being time-bound), while others are very relevant, such as ease of learning and documentation, implicit in (9) above.

The approach to evaluation followed here was to enumerate the properties needed or desired for the specific application of the GBDIS. Thirteen criteria were formulated that the DMS must satisfy, some of which were required by the contract (1-5 below), and others were dictated by the anticipated characteristics of the entries and the probable requirements of the users.

1. The DMS must be electronic, microcomputer-based software, and retail below $950.

2. The DMS must allow searching of the data base, based on key descriptors and/or fields related to the content of the data set descriptions.

3. The DMS must be sufficiently flexible to allow future updates with descriptions for new agency and project data, and future long-term monitoring data.

4. The DMS software should be standard and generally available for data base management.

5. It is desirable that the software be suitable for IBM-compatible equipment, available throughout state agencies. However, consideration should be given to the possibility of integration of the data base with COMPAS or related Macintosh systems.

6. The selected software should allow several (3-5, say) files to be accessed simultaneously.

7. The DMS software should be relational, i.e. permitting cross-comparisons between different files of information.

8. In addition to standardized retrievals, the DMS software should allow the construction of special-purpose retrievals, e.g., logical conjunction and disjunction of different key descriptors. Therefore, the DMS should possess a programmable capability. This should include the ability to construct custom menus.

9. The DMS should be user-oriented and have a high level of acceptance among PC users.

10. The DMS should allow variable-length records (to conserve storage space) and permit multiple index fields (for efficient retrieval).

11. The DMS should be capable of employment for other applications, and should allow easy interfacing with standard, readily available data processing software, e.g. spread sheets and statistical packages.

12. The information-retrieval system is expected to be a permanent, sustained entity, to be continuously updated and provided to various entities of the state and public requiring its use. Therefore, the selected DMS software should evidence a potential for longevity, i.e., to be supported and supplied by its manufacturer into the foreseeable future.

13. While not an immediate requirement of the data inventory system, the ability for networking will prove important in maintaining an updated data base at a central location that is accessible to users at other locations and in different agencies. Therefore, the DMS should include provision for networking.

Some discussion of these criteria is warranted. The first five, of course, are requirements delineated by the Management Conference and made a requirement of this contract. (The cost limitation in the first criterion was been added by the Principal Investigators.) Criteria (3) and (4) together imply the need for longevity, but for emphasis this is stated separately as Criterion (12). The rapid changes in PC

capabilities and available software will render any software choice "obsolete" within a matter of months. However, so long as the basic software structure is maintained by the manufacturer and is "upward compatible," the effort in structuring the data base and its retrieval logic will not be lost. The quality of longevity is not trivial. Of the relational DMS software reviewed by Jacobsen (1984) five years ago, 45% are no longer marketed. Of the five packages given detailed consideration in the 1989 *Personal Computing* review (Badgett et al., 1989), two did not exist five years ago. Probably the best measure of longevity is the demonstration of the manufacturer to upgrade and maintain support for its basic DMS product. This clearly introduces a bias into the selection against newer companies (or companies newly entering the DMS market), but the importance of the Galveston Bay Data Inventory System is too great to risk on future viability of an unproved company.

Cost, it will be noted, was not a criterion *per se*, apart from the retail ceiling in (1) of $ 950 (which was set because there seemed to be no logic in paying several thousand dollars for capabilities that could be acquired much less expensively). This is because the costs of various software packages satisfying the other criteria were in the range of $ 300-900. Differentials in this range become miniscule compared to the expense of personnel time, and therefore do not comprise a decision variable (except in the case of two otherwise identical packages--which, in any event, did not apparently exist). However, economy of investment in a broader sense *is* a criterion and represents the motivation for (11), in that a DMS that is used solely for manipulation of the GBNEP Data Inventory is a poor investment if another, more flexible system could be used for other purposes as well.

One potential additional use of concern to the GBNEP is the manipulation of the digitized data sets themselves, which in many instances will be the next step of an investigator studying the Galveston Bay system. The requirement (11) that the DMS permit interfacing with spread-sheet software addresses this concern specifically. (There are DMS packages with graphics capability, but this is only a part of the numerical manipulation that a researcher may desire.) To a certain extent the ability to export and import ASCII files will satisfy this requirement, but the anticipated application would be greatly facilitated by direct export/import of a standard spread-sheet format, specifically Lotus 1-2-3. As we became involved with the input of information, this criterion proved to be additionally important to simplify keyboard entry of large data files.

Criterion (5) regarding transferability to the Macintosh environment can be accommodated by export/import of ASCII files, so this did not prove to be a discriminating property. The reference to COMPAS, the NOAA Coastal Ocean Management Planning and Assessment System, is somewhat misleading, since COMPAS is a system for the assimilation and display of estuarine data *per se*, not entries in a data inventory, which is the present objective. However, a researcher's next step upon locating data sources will be to manipulate this data, which might include importing into COMPAS. Moreover, some of the information in the GBDIS can be of value in COMPAS, e.g. the file of sampling station coordinates. The flexibility implicit in criteria (8), (9) and (11) will permit incorporation in a Macintosh-based system, including COMPAS. (Indeed, in the development of data base files, we

frequently performed data entry in a Macintosh environment then exported to dBase on the PC. The reverse is easily accomplished.)

These criteria were applied to the selection of the DMS for the Galveston Bay DIS. The first broad criterion applied is (5) that the system be PC-based, given the wide usage of IBM-type equipment in the Texas Water Commission and other concerned state agencies, and (7) that the system be relational (i.e., permitting access to more than one file simultaneously, and allowing logical linking, comparison and sorting of elements from separate files). The eighty products listed by Badgett et al. (1989) satisfy these constraints, and, while others may exist, we considered this list sufficiently exhaustive for the purposes of this selection. Note that criteria (2), (4), (6) and (trivially) (1) are immediately satisfied as well. Next, in order to pare this list to something more manageable, we applied (8), that the system be programmable, and (11), that the system specifically export/import Lotus 1-2-3 formatted files. The software systems satisfying these criteria are listed in Table 6. In this table are presented the extent to which the remaining criteria are satisfied by each DMS. It turned out that all candidates satisfy (10), so this criterion did not form a basis for discrimination. Application of the criterion of longevity (12) was simple but perhaps a bit brutal: was the software of sufficient significance five years ago to appear in the review of Jacobsen (1984)?

Two candidate systems emerged from this screening, viz. dBase IV and DataEase 4.01. The basic differences between these systems is that DataEase emphasizes user simplicity at the sacrifice of programming power, while dBase, though providing through its sophisticated programming capability more user-custom flexibility, is considered more difficult to learn and apply (see Sander, 1988, and Blackford et al., 1988, as well as Badgett et al., 1989). This seems to have been reputation of dBase products residual from the previous versions, e.g. dBase III and dBase III+. To a large extent, this weakness of dBase has been remedied in dBase IV through the addition of user-friendly menus.

It was our recommendation that, of these two, dBase IV be adopted as the Data Inventory software. This decision was based upon the additional observations that dBase IV seems to be more readily available, both in local retailers and in mail-order houses, and (more significantly), few, if any, of the TWC staff that rely upon database-manager software, use DataEase while about half of this staff use dBase. Additional advantages accruing to the choice of dBase were (1) dBase seems to be the preference for EPA data base systems, as typified by the Clean Lakes Clearinghouse DMS, (2) the contractor for the Gulf Initiative data inventory, Sverdrup Technology Inc. (1989), has selected dBase III+ for that system.

One of the final choices in Table 6, Paradox, has become quite popular with some of the TWC staff, equal or a close second to dBase. Therefore, some additional comment is warranted as to the elimination of this software from consideration. Examination of Table 6 shows that Paradox was screened out for failing the

Table 6: Dms Software Considered For Galveston Bay Data Inventory System

Software	*Criterion*			
	(3) flexible	(9) user-friendly	(12) longevity	(13) net-working
dBase IV	Y	Y	Y	Y
dBase III+	Y	N	Y	Y
Paradox 3.0	?	N	N	Y
DataEase 4.01	Y	Y	Y	Y
Informix-SQL	?	Y	N	Y
Knowledge-Man/2	?	Y	Y	Y
R:Base DOS	Y	N	Y	Y
Ramis/PC 3.0	?	?	N	Y
Oracle	Y	?	N	Y
Super-base 4	?	?	N	Y
Key/500	?	?	N	N
XDB-SQL	?	?	N	Y

criteria for longevity and for user-simplicity. As noted above, the former can certainly be faulted for bias against newer software; certainly the present prominence of Paradox in the market would suggest that this product has achieved as much claim to permanence as its competitors. However, we believe the requirement for longevity to be valid, even if difficult to formulate as an objective criterion; further, once an objective criterion has been stated, it should be applied uniformly and unilaterally. The fact is that Paradox fails that criterion. The failure of Paradox in user-simplicity is based upon the large suite of functional options, which are intimidating to a first user, and the complex of manuals that must be penetrated in order to apply Paradox. Certainly, this deficiency is eliminated if the user has access to instructional seminars, as is the case for the TWC at Austin. Clearly, however, this cannot be generally assumed for Data Inventory users.

Finally, we must observe that this entire selection process is specious. This is because the immediate anticipated requirements of the GBNEP Data Inventory System are modest and will not tax or exhaust the capabilities of any of these data management systems. Indeed, any of the systems of Table 6, and many more, would accommodate our immediate needs. Therefore, the choice of dBase IV is in fact rather arbitrary, being based upon extraneous concerns, such as networking capability, or surmountable "deficiencies," such as user-friendliness. This arbitrariness is compensated by the fact that this decision is not really locked in. Because of the import/export capability of dBase, as well as our recognition of the need for generality in structure, the GB Data Inventory System data base should be capable of later transfer to another DMS, should the GBNEP or TWC deem appropriate. Therefore, the use of dBase IV in this project can be viewed as an expedient to act as a basis for retrieval structuring and data input, but not an irreversible software commitment.

3.2 Data Inventory Structure

The design of the data inventory database structure was driven by several constraints. First, the contract called for certain items to be included in the database, namely:

1. Agency/institution name
2. Data file name
3. Agency data base manager or contact, with telephone number and mailing address
4. Data description paragraph including name of the program resulting in the data collection, and objectives and use of the data collection
5. Period of duration for the collection of the data
6. Description of any technical data collection procedures utilized, including:
   a. data collection locations
   b. sample frequency
   c. methods and materials
   d. sample preparation/preservation
   e. laboratory procedures and methods
   f. results (data) entry and editing methods

g. data scrubbing/error trapping procedures
h. period of record, by parameter
i. parameters/information collected and units of measure

7. Complete technical specification for any computer storage media utilized for the data, including file formats with field layouts, software applications, and accessibility; for spatial data, aerial coverage, scale or resolution, digital vs. other forms of storage, units stored, and methods and coordinate types for location determinations.
8. Citations of any publications which have used or reported the data
9. Complete description of quality assurance and control for data collection, editing, and storage
10. Other information specific to data set.

Second, even though the general user of the database would not need to be aware of the structure of the database, the types of user searches envisioned dictated some of the structure. For the person unfamiliar with the GBDIS system, searches were designed to include the following:

by agency performing the work;

by data file name;

by principal investigator(s);

by keywords descriptive of the data set or present in the title of papers or reports written using the data;

by duration (year or range of years);

by location (latitude and longitude, water quality segment, segment name, and other identifiers);

by parameter (physical, chemical, or biological);

by priority problem; and

by combinations of two or more descriptors in fields.

Third, the GBNEP or group that would update the database would need to be able to input data to the database without difficulty. And fourth, the more sophisticated user of the system may want to do more detailed searches of the database at the "dot prompt" and would need to be able to search the proper portion of the database without difficulty.

All of these items except Items 6b through 6i could be easily accommodated in a single database, which is characterized as the "general information" collected about each data set. Item 6a was included initially in the general database as it was anticipated that many searches would be on location of data. However, because the number of sampling stations used in some sampling efforts became extremely large,

it became necessary to create a separate file linking station numbers and locations to individual data sets. As the types of data to be gathered began to be examined in detail, it was clear that Items 6b through 6i would have to be subdivided into classes of data type and databases created for each type with a connecting reference to link each of these data type databases. This was a marked departure from what had been envisioned in the Work Plan, but resulted in the only workable solution to accomplish the goals of the GBNEP. For Item 6, the subdivisions of data type used were:

Morphology
Hydrography
Hydrology
Water Quality
Sediment Quality
Biological
Public Health
Pollutant Loading
Sociologic
Economic

and for each of these data types, two databases were created: an information database to include Items 6b through 6g and a data database to include Items 6h and 6i. These are described in more detail in Armstrong and Ward (1991).

One of the prime features of the GBNEP data-base structure is the use of multiple files, elements of which are "related" (i.e., logically identified for access and retrieval purposes). This is because different types of data have different properties (called "fields" in the data-base management patois), so their logic structure must be different. An element (i.e. entry) of the GBNEP data base is a "project", referring to a uniform, systematic, autonomous data-collection enterprise. A "project" might be a one-time collection of sediment samples, a one-year study of shrimp communities, or a routine collection of water quality data at regular intervals over many years. A project might concentrate upon one geographical region of the bay, or might involve samples throughout the bay. Retrieval is implemented by searching on field variables, perhaps constrained by user-specified relations, and by keyword textual searches of the title and abstract information in the data entries. This dual approach to retrieval allows both quantitative sorting of the information, as well as qualitative searching.

The information contained in the databases allow all the information called for in the contract list above to be entered as well as to allow for the various types of data to be entered, data such as:

point observations/measurements,
analog time series,
analog line series,
continuous or discrete areal delineations,
anecdotals.

Likewise, the sources for these data took several forms as follows (with examples of each):

open literature (books, journals)
grey literature (technical reports, project studies)
file documents (unpublished manuscripts, internal memoranda)
transient literature (newspapers, diaries, historical collections)
formalized data tabulations (publications, computer-readable media)
organized data archives (indexed maps, aerial photos)
raw data (field sheets, strip charts, cassette logs)

and could also be entered. For those entries from the open literature, a citation in the usual scientific format sufficed to uniquely identify the source of data and permitted a researcher to access the source. Thus, field elements obviously included author, title, and journal or book bibliographic identifiers. Any of the above properties will constitute a retrieval parameter (or field), e.g., "author", "aerial photograph" or "tide record". Other retrieval fields would include types of information or measurement, such as "salinity", "*Callinectes* sp", or "water depth". Additional retrieval parameters, incorporated as fields for direct retrievals or descriptors for textual searches, include: specific chemical measurements; geographic location; date (of sample or of fundamental information in citation); date of publication

A GBDIS response query to the specific field parameters activated during the data entry operation has been developed, offering a selection of qualifying information to the entry clerk. This is the means of entering information on the type of measurement or analysis employed, the Q/A procedures, and so on. As an example, a retrieval on "salinity" should produce all of the parameters conventionally used as measures of salinity, including conductivity, chlorinity, and density (hydrometer). Fortunately, there are a small number of potential methodologies applicable for a given variable, so it was feasible to build up a file (or files) of these as part of the data system structure. Like the ADS files described below, these files will be "transparent" to the user. Their contents would be accumulated during the initial data base formulation and entries; we anticipate that after a short period, these files would become essentially static, and would serve from then on to prompt the data entry technician for more detail. The philosophy is that at this point in the process--as the data entry is made from the primary source--it is easiest to search out and input such relevant details as Q/A and analytical methodologies.

The linkage among all the database files is a unique reference number assigned to each data set. Thus, once a data set is identified as the one from which information is desired, then information from any of the associated database files may be retrieved.

3.3 Data Entry

Data entry into the inventory database is facilitated by a program written in dBase with menus and data forms to allow the user to select the type of data to be entered (i.e., morphological, water quality, etc.) then to enter that data via a form on the screen into the database. Some error checking is done during the data entry process (to avoid duplication of reference numbers for example) and helps are provided in terms of lists of parameter names corresponding to those used by the TWC and EPA. Again, the data entry program and procedures are described in Armstrong and Ward (1991).

3.4 Data Retrieval

Data retrieval from the GBNEP Data Inventory System is achieved through a program written in dBase using menus and help screens so that the lay user as well as the experienced user may search the database in a number of ways. As noted above, the types of retrievals now possible using the system are:

1. by federal, state, or local government agency or private corporation performing the work using an acronym for the agency or corporation;

2. by data file or program name (i.e., maintenance dredging);

3. by principal investigator(s);

4. by keywords descriptive of the data set or present in the title of papers or reports written using the data;

5. by duration (month/day/year or period from one month/day/year to another);

6. by location (latitude and longitude, water quality segment, segment name, and other identifiers);

7. by parameter (physical, chemical, or biological);

8. by GBNEP priority problem; and

9. by combinations of the above.

With such search capabilities, for example, one can determine all the studies inventoried which were conducted by a federal, state, or local governmental agency or other groups (search 1) or those studies carried out by that agency in a specific period of time and/or location (search 9). It will also be possible to locate specific data sets, for example, the previous Galveston Bay Study data set (search 2). If it is desired to know the studies performed by a particular report author (search 3) conducted in particular parts of the bay system at particular points in time (search 9), that can be done. Searches for studies in which particular types of water quality constituents and biological components and processes were sampled again in space and time can be done. Finally, searches for studies with information pertaining to particular GBNEP priority problems are possible.

One of the more important retrieval fields is that specifying the locations within the system at which the data observations/measurements were made. One anticipated use for the data inventory was retrieval of data of a specified type pertaining to a specific geographical subarea of Galveston Bay. After much consideration and review, latitude-longitude coordinates were adopted as the basic position specification. This decision entailed a considerable effort in the data entry process; because relatively few data sets had the measurement positions specified by latitude and longitude, it was necessary to map these points and determine the coordinates ourselves. However, the generality and flexibility of this approach justified its employment. Further, latitude-longitude coordinates and geographical descriptors can be cross-referenced, thereby facilitating searches given only the geographical name of a feature in the Bay.

3.5 Computer Hardware Required

The Galveston Bay Data Inventory System at the University of Texas is implemented on a dedicated 386-based microcomputer of PC architecture, operating at 20 kHz, and equipped with high-density disk drives and an 80 MB hard drive. The actual system, in its present form, requires some 10 MB of hard drive storage including the 2.5 MB needed for dBase IV software, so the system can be accommodated on a more modest machine. The size of the data files and the complexity of logical searching do, we believe, mandate a short cycle time. We recommend therefore that the system be installed on at least a 286-based machine (i.e., AT equivalent) with at least 20 MB hard drive. Clearly, if the machine is to be used for any other purposes requiring hard drive access, then a larger capacity drive may be necessary.

The original project scope assumed that any user of the Galveston Bay Data Inventory System would separately purchase dBase IV software. Upon reconsideration, we have invested in the Developers Version of dBase IV, which allows the production of compiled, executable codes that obviate separate software (and will free some of the hard drive storage as well). Therefore, a potential user no longer needs a separate purchase of dBase IV.

4. DATA LOCATION AND APPRAISAL

4.1 Project Prosecution

The GBNEP Data Inventory project entailed an enormous amount of agency and individual contacts. These included over 200 letters of inquiry, many meetings, and innumerable telephone conversations. A continuing log of these contacts was maintained through the course of the project to allow immediate appraisal of the status of given data sets and/or agencies. The original schedule was for a six-month study to begin 1 December 1989 (with the Work Plan to be delivered one month later). More time proved to be necessary to implement the Interagency Contract through which the study was performed, and the project did not formally begin until 1 April 1990, at which time the completion date was revised to 31 December 1990. The completion date was then further extended to 31 May 1991, so that the project duration has now more than doubled the original schedule. The reasons for this extended period are examined below.

One of the prime reasons for the extended schedule is the nature of the data resource itself. There are a relatively small number of "big" projects, which amassed considerable information individually, but a large number of small projects that cumulatively account for perhaps 50% of the total information base. We did not anticipate such a large number of small projects. Clearly, the time and effort required to identify and locate a set of data is largely independent of the information content of that data, or, if anything, might vary inversely, i.e. the larger projects being easier to identify and locate than the smaller. Further, since about 1970, there has been a decline in intensity of data collection. (Of course, this qualitative statement must be tempered according to the specific data. For example, sediment chemistry data collection has increased from the late 1960's up to the present, with most of the information resource being accumulated since 1980. The same can be said of many exotics and toxic substances, for which data collection has improved with the development of analytical methods. Further, since about 1970, we have seen the burgeoning of remote sensing, and the vast information potential of that technology. Both the magnitude and time distribution of the data resource are examined further in the following section.) Therefore, the older data sets take on a relatively greater importance in the synthesis of a comprehensive data base, hence the need for redoubled effort to locate them. The totality of many small projects, most of which are older and difficult to locate, greatly compounded the personal efforts of the Principal Investigators.

A second factor affecting prosecution of the work was the generally unsatisfactory response to the inquiries of the Principal Investigators. Some agencies simply did not reply to our inquiries. This was especially true of private industries and individual researchers. Those that did reply frequently did so many months after the original inquiry (and then only after follow-up letters and telephone calls). In fact, we have

had as much response and new access to information *since* 31 December 1990 as we had prior to this date from the inception of the project.

The principal agencies and individuals contacted in the course of this study are summarized in Table 7, along with known or suspected data holdings, the nature of the response of the agency to our inquiries, and the status of data holdings. (A listing of agency acronyms is provided at the end of this report.) These are grouped according to the major sectors of federal agency, state agency, local and regional agency, universities (organized research units only), and private agencies. Those agencies directly involved with the Galveston Bay National Estuary Program are marked with an asterisk (*). A marginal response means that little or no effort was made on the part of the responder to determine available holdings and/or to make those holdings available to the GBNEP. The timeliness of the responses is indicated by a Q (quick) or L (tardy), the latter referring to responses generally after 31 August 1990 (by which time, according to the original project plan, we should have had all responses in and been in the process of completing the inventory process). Purely negative responses are divided into three categories: N for simply no response to our inquiries, R for a response with a promise of a follow-up that never materialized, and D for a flat denial of access to information (including the denial that data existed, when we knew the opposite to be true).

As can be seen from this table, the overall response was disappointing. Of the non-University agencies contacted, over 60% gave negative or marginal responses. The least responsive category was private industry, with over 70% nonresponsive--as we might have anticipated. What is most surprising is the poor response from GBNEP agencies, with over 67% marginal-or-poorer response, including 38% nonresponse. While this poor showing was dominated by the private and local GBNEP participants (theoretically the principal beneficiaries of the project), it should be noted that some federal and state agencies also were nonresponsive. If we include those GBNEP agencies which provided good but tardy cooperation, more than 82% of the GBNEP agencies failed to expedite the prosecution of this work. Generally, we do not regard this as obstructionism or antipathy to the project, but rather, perhaps, passive resistance. Much of the tardy response was due to the numerous agency employees who were contacted directly and had to make room in their normal duties to accommodate the inquiries of the PI's. Most of these staffers had never heard of the GBNEP and regarded the PI's as still another interruption of their work, which they courteously but resignedly sustained.

Table 7 of course does not list the many individuals contacted. This class of contact includes most universities, where research is usually performed independently under the direction of faculty, who must be contacted on an individual basis. Each faculty is responsible for the storage and tracking of his own materials, and there is little or no provision for preservation of those of a retired or resigned faculty. (In fact, most such faculty take their materials with them or arrange for disposal.) Only organized research units of universities are listed on Table 7, because these research units offered the best chance for some continuity over time, and some means of permanent storage of data and results. The individual researchers, as a class, exhibited the highest level of nonresponse, and where there were responses, the lowest proportion of data recovery.

Table 7: Agencies, Response To Nep Inquiries, Type And Status Of Data Holdings

Under HOLDINGS, ■? indicates suspected holdings in the discipline category, -?- indicates no knowledge of holdings.
Under RESPONSE, a GOOD or MARGINAL response is characterized by Q for quick (timely), or L for late (tardy).
A response of NONE is characterized by N for no response at all, R for reply promising follow-up but no further action, and D for denial of access to holdings, or denial of existence of holdings when the holdings in fact exist.
An asterisk (*) marks an agency associated with the National Estuary Program.

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						STATUS
	GOOD	MAR-GINAL	NONE	MOR-PHOLOGY	HYDROG-RAPHY	HYDROL-OGY	BIOCHEM-ISTRY	BIOLOGY ECOLOGY	SOCIO-ECONOMICS	
FEDERAL:										
*USCE/Galv	Q,L			■	■	■	■	■	■	vast, varying preservation
USCE/Ft-W		Q		■	■	■	■	■		
USCE/SW Div		Q								
USCE/WES	Q			■	■		■			much material lost
USCE/CERC		L	R	■	■					BEB data lost
USGS/Aus	Q			■	■	■				well archived
*USGS/Hous			R		■	■	■	■		
USGS/Reston	Q			■		■				well archived
USGS/GHRC			D			-----?-----				Trinity River Proj lost?
USGS/NCIC		Q		■						
*USFWS/Houston			N		■?		■?	■?		
USFWS/Slidell		Q			■?	■?	■?	■?		no archival procedure
*USDA/SCS		L		■		■	■?			
USCG	Q				■					data at NCDC
NOAA/NESDIS		Q		■						
NOAA/NODC		Q	R	■			■	■		
NOAA/NOS/NCD	Q			■						
NOAA/NOS/PSSS	Q			■						
NOAA/NOS/Tides		L			■					still awaiting response
NOAA/NOS/NS&TP	Q						■	■		
NOAA/NCDC		Q			■					good archival procedures
*NOAA/NMFS (USBCF)		L	D,R		■		■	■	■?	much data apparently lost
*EPA/Reg VI		Q			■	■	■	■		no archival procedure

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR-GINAL	NONE	MOR-PHOLOGY	HYDROG-RAPHY	HYDROL-OGY	BIOCHEM-ISTRY	BIOLOGY ECOLOGY	SOCIO-ECONOMICS	
FEDERAL (Continued)										
EPA/Hous	L				■					data sent to Reg Vi
NASA/JSC	Q		R	■						Trinity Bay Study lost
NASA/EROS			R	■	■?	■?				
NASA/Langley	Q									
EOSAT		Q								
National Archives	L				■					
STATE:										
*TWC/Austin	Q				■		■	■?	■?	SMN archived, GBP lost
*TWC/District 7	L				■		■	■?	■?	
*TWDB/B&E	Q			■	■	■	■	■	■	
*TWDB/TNRIS		Q		■					■	
*TPWD/Coastal/Aus	Q			■?	■		■	■	■	
*TPWD/Seabrook	Q				■		■	■		much data lost
*TPWD/Rockport	Q				■		■	■		much data lost
*GLO	L			■	■					
*TSDH			R		■		■	■		much data lost
*RRC			D			■			■	
TACB	Q				■					meteorology only
TENRAC						-----?-----				defunct, no records
TSDA			N				■?		■?	
TDA		Q							■	
REGIONAL:										
TRA										in estuary, mainly derivative
CLCND			D		■	■				no archival procedure
*GCWDA			D				■?			
HGCSD				■					■?	
*HGAC		Q							■	

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR-GINAL	NONE	MOR-PHOLOGY	HYDROG-RAPHY	HYDROL-OGY	BIOCHEM-ISTRY	BIOLOGY ECOLOGY	SOCIO-ECONOMICS	
LOCAL:										
*Harris Cnty HD		L			■?		■			much data lost
*Harris Cnty PCD		Q			■		■			extensive data, limited access
*Galv Cnty Engr		Q		■						
*Galv Cnty Parks		Q		■						
*City of Houston	L					■	■			vast data on point sources
*POHA			N	■	■?	■?			■	
*Galv Wharves		Q								mainly EH&A POG reports
*Galv Cnty HD	L						■			much data on point sources
*Chamb Cnty HD			N				■			
UNIVERSITY										
UT:										
CRWR	Q			■	■	■	■	■		mainly in project reports
*BEG	Q			■	■	■	■	■		much data well archived
*MSI	Q			■						Beasley collection
Barker Cntr				■	■				■	
BBR		L								
UTMB		Q								no archival procedure
TAMU:										
TAES		Q								no archival procedure
Sea Grant	Q									project reports only
TWRI				■	■	■	■			project reports only
TREC		Q								no archival procedure
CCSU:										
Blucher Institute	Q				■					recent, excellent archiving
PAS:										
Limnology	Q				■		■	■		

Table 7 (Continued)

AGENCY	RESPONSE TO INQUIRIES			DATA HOLDINGS BY DISCIPLINE						COMMENTS
	GOOD	MAR-GINAL	NONE	MOR-PHOLOGY	HYDROG-RAPHY	HYDROL-OGY	BIOCHEM-ISTRY	BIOLOGY ECOLOGY	SOCIO-ECONOMICS	
PRIVATE:										
BJI			N	■	■	■?	■?	■?	■	major studies in bay
Bovay			N		■		■?	■?	■?	
*TCB			N			-----?-----				many studies in bay area
WWC			N			-----?-----				
B&R		L	N			-----?-----				
LGL			N		■		■	■		many studies in bay
TI		L		■	■					
LAN			N	■	■					major studies in 1950s
Dannenbaum			N			-----?-----				
Flour			N			-----?-----				
Law			N			-----?-----				
CESI					■?		■	■		Apparently defunct
Tobin			N	■						vast photographic holdings
Geomarine	Q									
Parker Bros		L								
EHA	Q				■	■	■	■	■	vast holdings, mainly reports
*HL&P	Q				■		■	■		vast holdings, no archival
*Enron Corp		Q								
*Dow Chem			N							
*Exxon/Humble			R	■	■		■	■		much data 40-60s t
*Amoco			N							
*Sterling Chem			N							
Shell			N	■?	■?	■?				
Goldston Oil			N			-----?-----				
Ethyl Copr			N			-----?-----				

Table 7 does not communicate the number of agencies which responded only after personal visits and persistent searching of the PI's. These are indicated as a Good response (in the sense that the bulk of the agency's data holdings were made accessible to the project). Also, in some instances, the initial response was to deny existence of information. When we personally knew of past data programs, we pressed for specifics, and usually were able to exact a more cooperative response. It helped therefore to be specific and knowledgeable in our request. In many cases, this preliminary negative response was due to personnel turnover, and the respondent being truthfully ignorant of the agency's past work. Sometimes, the data still did not materialize, but not because of uncooperativeness by the agency staff.

The point-of-contact approach generally failed. With a few, laudable exceptions, the agency points-of-contact did little more than provide a list of names and phone numbers of individuals in the agency that might have some information. As noted above, lack of knowledge of the GBNEP and its program contributed to tardy access to some agency holdings. This situation derived more from the failure of the point-of-contact approach than lack of staff cooperativeness. Both of these factors, the generally poor response, and the need for the PI's to directly contact various agency personnel and personally search through agency files, translated directly to a great investment of project time in the Three R's, writing, ringing, and rooting.

One additional, significant contributor to the increased project effort and schedule was the poor state of data/information management that seems to be ubiquitous. An unexpectedly large effort had to be expended in searching for data sets, which were expected to be, and should have been, readily available. Major data collection projects, which, as of 31 December 1990, were *totally missing* included:

the Galveston Bay Project High-Frequency Program,
the USBCF biological program of 1958-67,
data collections by the Harris County Pollution Control Department from the 1960's, the intensive studies of the Houston Ship Channel by Humble Oil during the 1950's and 1960's, and practically anything done by the Corps before 1960, (See Section 4.2.)
the older Texas State Health Department surveys of Galveston Bay, from the 1950's and early 1960's,
the four-year program of water sampling on the Houston Ship Channel carried out by Texas A&M University in the late 1960's under sponsorship of the Federal Water Pollution Control Administration.

Data sets which had been entered on digital media but, as of 31 December 1990, only existed as one or a few *hard-copy* tabulations include:

the Galveston Bay Project Routine Program (1968-1972),
the USBCF 1958-67 water quality data,
the joint TSDH/Galveston County Galveston Bay Project of 1963-67,
the USCE Trinity marsh biology and chemistry data 1975-76, and
the HL&P Cedar Bayou studies of Trinity Bay in the early 1970's (but incomplete).

We are compelled to note that many of the successful contacts and data location were due to the long experience of the PI's in this area, acquaintance with individuals in the various agencies and companies, and specific knowledge of those entities' past project work. This is not an expression of how great the PI's are (a matter of considerable debate, unfortunately), but of the general poor state of data management and the difficulty of the inventory task. Much of the data reflected in this inventory, especially that rescued from the brink of oblivion, could not have been accessed through a normal discovery procedure, even with the considerable auspices of a National Estuary Program. Without the initiate's ability to call a specific person in an agency and ask for a specific piece of information, this data could not have been located. At this writing, though the project is formally over, information on data sets is continuing to dribble in, data sets are continuing to turn up, and we are still trying to track down leads for missing data.

As the project progressed, the scale of the problem began to be manifest, and difficulties in time and scheduling were encountered, it became necessary to prioritize the data sought. The touchstone for this prioritization was the array of GBNEP Priority Problems, as well as the pragmatic judgment of how the PI time could be best invested. With respect to the latter, we began to de-emphasize inventorying of data from programs in which there was in place a readily accessible and high-technology data management system. For example, for the USGS streamflow data, the NOS navigation charting products, standard census and economic data compilations, and the National Climatic Data Center (including National Weather Service data), information is readily available elsewhere concerning the retrieval and application of these kinds of data. Also, we placed first emphasis upon primary data sources. For example, many of the holdings of TNRIS were not cataloged in the system, because the same information is available from the primary data collection entity. Finally, the press of time led to prioritization of information according to the GBNEP Priority Problems. Thus, for instance, much socioeconomic data and information on bird and terrestrial organisms could not be sought. Also, attorneys as a general category were downgraded to lower priority because most of these were unlikely to retain data files, would have only derivative data, and would be generally resistant to release of information.

Fortunately, the data management system is capable of expansion *ad infinitum* so later researchers or GBNEP staff can augment the data inventory with additional information. Indeed, we are continuing, at this writing, to add data sets to the system.

4.2 Summary of Data Resource

In Table 8 is presented a summary of data holdings by various agencies according to the more detailed discipline breakdowns of Table 5. Of perhaps more significance, Table 9 summarizes the *major* data-collection projects in the Galveston Bay system, with some indication of the data base content and extent, and the present status of that data base. (This table should be cross-compared with Tables 7 and 8.) These data sets are grouped according to the basic nature of the data, in these categories:

CST - Conductivity and/or salinity, and temperature

CHM - Water chemistry

SED - Sediment quality

HYDG - Hydrography, including current measurements, physical processes and transports

BIO - Biology and ecology

There are obviously other categories of data, and other ways to categorize the data summarized in this table. Some projects are counted in several categories when more than one type of data was collected. The time period of data collection is given for each project. It should be noted that "agency" and "project" are not equivalent. When the procedures or objectives of an agency activity change at some point in time, we regard this as separate programs, e.g. the sediment data collections of the Galveston District Corps for the periods 1971-72, 1974-80 and 1980-90.

In Table 9, each data set is further characterized by a rough estimate of the information content of the data. This is difficult to quantify in any absolute sensible fashion, may be impossible, and is probably silly to try. Ideally, it should include some measure of the time and space intensity of the data collection, and the difficulty of the analysis, but also the "value" of the information. For present purposes, we define an "observation" to be the measurement of one parameter at one point in space-time. Therefore a measurement of surface salinity and temperature at a station in the bay comprises two "observations." A four-level profile of these same parameters at that station would constitute eight "observations." One sediment sample analyzed for 20 constituents would represent 20 "observations." For biological data, we consider an "observation" to be the *count* of one species in a collection, but not the identification of individuals. If the individuals (or a subsample) of that species are subjected to further measurement, e.g. length or weight, then that comprises a separate "observation." There are obviously many deficiencies to this kind of simple-minded measure, but it does serve as a relative indicator of the magnitude of a data set and therefore its relative importance among other similar data sets. By separating the data sets into the above discipline categories, we duck the question of the relative value of a measurement of BOD, say, versus a count of menhaden.

Table 8: Agency Data Holdings By Sub-Discipline

■? indicates suspected holdings in the sub-discipline category.
An asterisk (*) marks an agency associated with the National Estuary Program.

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	MORPHOLOGY					
	bathy-metry	shoreline morph-ology	siltation	erosion	dredging	subsidence
FEDERAL:						
*USCE/Galv	■	■	■	■	■	
USCE/Ft-W		■	■			
USCE/WES	■	■			■	
USCE/CERC		■	■			
USGS/Aus			■			■
USGS/Reston	■	■	■			
USGS/NCIC	■	■				
*USDA/SCS	■	■		■		
NOAA/NESDIS		■				
NOAA/NODC	■	■				
NOAA/NOS/NCD		■				
NOAA/NOS/PSSS	■	■				
NASA/JSC		■	■			
NASA/EROS		■				
STATE:						
*TWDB/B&E	■	■	■			■
*TWDB/TNRIS		■				
*TPWD/Coastal/Aus	■?	■?	■?			
*GLO	■	■	■	■	■	
REGIONAL:						
HGCSD						■
LOCAL:						
*Galv Cnty Engr		■	■?			
*Galv Cnty Parks		■				
*POHA	■	■		■?	■?	

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	MORPHOLOGY (Contiuned)					
	bathy-metry	shoreline morph-ology	siltation	erosion	dredging	subsidence
UNIVERSITY						
UT:						
CRWR	■	■	■			
*BEG	■	■	■	■	■	■
*MSI		■				
Barker Cntr	■	■				
TAMU:						
TWRI		■				
CCSU:						
Blucher Institute	■	■				
PRIVATE:						
BJI	■	■	■?	■?	■?	■?
TI		■				
LAN	■	■	■	■		
Tobin		■				
EHA	■	■			■	
*HL&P	■	■				
*Exxon/Humble	■	■	■	■		
Shell	■?	■?	■?	■?		

	HYDROGRAPHY					
	tides	currents	circul-ation	winds	waves	meteorology
FEDERAL:						
*USCE/Galv	■	■	■	■	■	
USCE/WES	■	■	■			
USCE/CERC					■	
USGS/Aus	■					
*USGS/Hous	■?	■?				
USCG				■	■	■
NOAA/NOS/Tides	■	■	■			
NOAA/NCDC				■	■	■
National Archives	■	■	■			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	HYDROGRAPHY (Continued)					
	tides	currents	circul-ation	winds	waves	meteorology
STATE:						
*TWDB/B&E	■	■	■			
*TWDB/TNRIS	■	■	■	■		■
*TPWD/Coastal/Aus					■	■
*TSDH		■?				
TACB				■		■
UNIVERSITY						
UTA CRWR	■	■	■			
*UTA BEG			■			
CCSU Blucher Institute		■				
PRIVATE:						
BJI			■			
LAN		■?	■?			
EHA	■	■	■	■		■
*HL&P			■	■		■
*Exxon/Humble	■	■?	■		■?	
Shell	■?	■?	■?	■?		

	HYDROGRAPHY (continued)				
	salinity	temper-ature	turbidity	tracer studies	spills
FEDERAL:					
*USCE/Galv	■	■	■	■	■
USCE/Ft-W	■	■	■		
USCE/WES	■	■	■	■	
USCE/CERC	■	■			
USGS/Aus	■	■	■		
*USGS/Hous	■	■	■		
*USFWS/Houston	■?	■?			■?
USCG					■
NOAA/NOS/Tides	■	■			
*NOAA/NMFS (USBCF)	■	■	■		
*EPA/Reg VI	■	■			■
EPA/Hous	■	■			■
NASA/EROS			■?		
National Archives	■	■			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE				
	HYDROGRAPHY (continued)				
	salinity	temper-ature	turbidity	tracer studies	spills
STATE:					
*TWC/Austin	■	■	■	■	■
*TWC/District 7	■	■	■		■
*TWDB/B&E	■	■	■		
*TWDB/TNRIS	■	■	■		
*TPWD/Coastal/Aus	■	■			
*TPWD/Seabrook	■	■			
*TPWD/Rockport	■	■			
*GLO					■
*TSDH	■	■	■		
REGIONAL:					
CLCND	■				
LOCAL:					
*Harris Cnty HD	■?	■?			
*Harris Cnty PCD	■	■			
*POHA	■?	■?			■?
UNIVERSITY					
UTA CRWR	■	■	■	■	
*UTA BEG	■	■	■		
TAMU TWRI	■	■			
CCSU Blucher Institute			■		
PAS Limnology	■	■			
PRIVATE:					
BJI	■	■			
Bovay	■?				
LGL	■	■			
EHA	■	■	■	■	
*HL&P	■	■	■		
*Exxon/Humble	■	■			
Shell	■?	■?			

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	HYDROLOGY					
	inflow	runoff	sediment	water use	diver-sions	return flows
FEDERAL:						
*USCE/Galv	■	■	■	■	■	■
USCE/Ft-W		■	■		■	
USGS/Aus	■	■	■	■		
*USGS/Hous	■	■	■	■?	■?	■?
USGS/Reston	■		■			
*USFWS/Houston						
USFWS/Slidell		■?	■?			
*USDA/SCS		■	■	■?	■?	■?
*EPA/Reg VI	■			■	■	■
NASA/EROS			■?		■?	
STATE:						
*TWC/Austin				■	■	■
*TWC/District 7						■
*TWDB/B&E	■	■	■		■	
*RRC						■
REGIONAL:						
CLCND					■	■
*GCWDA						■
LOCAL:						
*Harris Cnty HD						■
*Harris Cnty PCD						■
*City of Houston		■				■
UNIVERSITY						
UTA CRWR		■	■			
*BEG			■			
PRIVATE:						
BJI		■?		■?	■?	■?
EHA		■		■	■	■
*HL&P					■	■
*Exxon/Humble						
Shell		■?		■?		

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE					
	BIOLOGY & ECOLOGY					
	micro-biology	phyto-plankton	zoo-plankton	micro-benthos	veget-ation	shell-fish
FEDERAL:						
*USCE/Galv	■	■	■	■	■	■
USCE/Ft-W					■	■
USCE/WES				■		
*USGS/Hous		■	■			
USGS/Reston		■	■			
*USFWS/Houston	■?	■?	■?		■?	■?
USFWS/Slidell						■?
NOAA/NODC						■
NOAA/NOS/NS&TP						■
*NOAA/NMFS (USBCF)	■?	■	■	■?	■	■
*EPA/Reg VI		■	■			■
STATE:						
*TWC/Austin		■	■	■		
*TWC/District 7	■	■	■		■	■
*TWDB/B&E		■	■			
*TPWD/Coastal/Aus					■	■
*TPWD/Seabrook					■	■
*TPWD/Rockport					■	■
*TSDH	■					■
UNIVERSITY						
UTA CRWR		■	■		■	
*BEG				■	■	■
PAS Limnology		■				
PRIVATE:						
BJI						■?
Bovay						■?
LGL		■?	■?			■
EHA		■	■			■
*HL&P		■	■			■
*Exxon/Humble		■	■			■

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE				
	fish	benthos	marshes	wetlands	oysters
FEDERAL:					
*USCE/Galv	■	■	■	■	■
USCE/Ft-W	■		■	■	
*USFWS/Houston	■?	■?	■?	■?	■?
USFWS/Slidell	■?			■?	
NOAA/NODC	■?			■	■
NOAA/NOS/NS&TP					■
*NOAA/NMFS (USBCF)	■		■	■	■
*EPA/Reg VI				■	
STATE:					
*TWC/Austin	■	■	■	■	
*TWC/District 7	■	■	■	■	■
*TWDB/B&E	■	■	■	■	
*TPWD/Coastal/Aus	■	■	■	■	■
*TPWD/Seabrook	■	■	■	■	■
*TPWD/Rockport	■	■	■	■	■
*GLO			■	■	■
*TSDH					■
UNIVERSITY					
UTA CRWR			■	■	
*BEG			■	■	■
TAMU TWRI			■	■	
PRIVATE:					
BJI	■?		■?	■?	
Bovay	■?				
LGL	■	■			■
EHA	■	■	■	■	
*HL&P	■	■			■
*Exxon/Humble	■				

Table 8 (Continued): Agency Data Holdings By Sub-Discipline

AGENCY	DATA HOLDINGS BY SUBDISCIPLINE	
	SOCIOECONOMICS	
	demography	economics
FEDERAL:		
*USCE/Galv	■	■
USCE/Ft-W		■
STATE:		
*TWC/Austin		■
*TWDB/B&E	■	
*TWDB/TNRIS	■	■
*TPWD/Coastal/Aus	■	■
*GLO	■	
*TSDH		■
*RRC		■
TDA	■	
REGIONAL:		
*HGAC	■	■?
LOCAL:		
*City of Houston		■
*POHA		■
PRIVATE:		
BJI		■?
Bovay		■?
EHA		■

Table 9: Principal Large-Scale Data Sets from Galveston Bay and Their Status as of 31 December 1990

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal sampling & special studies	TGFOC (now TPWD)	ca 1936 -ca 1950	CST	10,000?	field sheets	Operations out of Marine Lab at Rockport	Lost, probably in 1971 Rockport fire	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950 -1961	CST	20,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	CST	500	field sheets	only generalized results in 4 journal papers	data lost, probably in 1961 hurricane	0/-
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	CST	1,000	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Galveston Bay Chemistry Survey	USBCF	1958-67	CST	14,500	digital	a few poor quality print-outs remain as data report	cards destroyed	2/0
Coastal Fisheries Sampling	TPWD	1962-75	CST	15,000	digital subset	10-40 routine stations	Index cards at Sea-brook Lab TPWD	1/0
Coastal Studies Data	TSDH	1963-67	CST	7,000	digital	few printouts exist	tape lost	2/0
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	CST	7,000	field sheets	Some open-bay stations	available Galveston Dist.	1/-
PH Robinson SES Surveys	TAMU	1968-69	CST	700	hard copy	most data in project reports & theses	see reports	3/-
Galveston Bay Project, Routine	TWQB	1968-72	CST	6,500	digital	printouts in project reports	tape lost	2/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Estuarine Systems Project	TAMU (Roy Hann)	1968-71	CST	15,000	digital	data never published	cards destroyed data lost	0/0
Galveston Bay Project, High-frequency	TWQB	1968-72	CST	6,000	digital	no hard copy	tape lost	0/0
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	CST	15,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Galveston Bay Project, Ecological Survey	UT MSI (Copeland)	1969	CST	500	hard copy	data published only in summary form	data lost	0/1
Statewide Monitoring Network, bay + tribs	TWC	ca.1970-present	CST	70,000	digital	clumsy downloading capabilities, usually hard copy	available TWC	3/3
Estuarine Water Quality	TSDH	ca.1970-	CST	30,000	digital		available TSDH	0/2
O&M Dredging Project	USCE/ Galv	1974-80	CST	6,500	hard copy	comprehensive water sampling	available at Galveston Dist.	2/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	CST	1,000	digital	raw field sheets exist	cards destroyed	1/0
Coastal Fisheries Sampling	TPWD	1975-pres	CST	35,000	digital	randomly selected stations	available from TPWD	1/4
Intensive inflow study entire bay system	TWDB	1976	CST	6,500	digital		CDS of TWDB	0/3
Bays & Estuaries Program	TWDB	1976-89	CST	20,000	digital	Coastal Data System	available TWDB	0?/3
O&M Dredging Project	USCE/ Galv	1980-pres	CST	4,500	field sheets	comprehensive water sampling	available at Galveston Dist.	1/-

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Harris County Stream Pollution Surveys	HCHD	1949-51	CHM	8,000	hard copy	stations in upper bay & tributaries	data lost	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950 -1961	CHM	10,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	CHM	500	field sheets	only generalized results in 4 journal papers	field data lost	0/-
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	CHM	5,500	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Galveston Bay Chemistry Survey	USBCF	1958-67	CHM	6,000	digital	a few poor quality print-outs remain as data report	cards destroyed	2/0
Coastal Fisheries Sampling	TPWD	1962-75	CHM	7,000	hard copy	10-40 routine stations	Index cards at Sea-brook Lab TPWD	1/0
Coastal Studies Data (Galveston Bay Project)	TSDH	1963-67	CHM	25,000	digital	few printouts exist	tape lost	2/0
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	CHM	3,000	field sheets	Some open-bay stations	available Galveston Dist.	2/-
PH Robinson SES Surveys	TAMU	1968-69	CHM	700	hard copy	most data in project	see reports	3/-
Galveston Bay Project, High-frequency	TWQB	1968-72	CHM	6,000	digital	no hard copy	tape lost	0/0
Galveston Bay Project, Routine	TWQB	1968-72	CHM	35,500	digital	printouts in project reports	tape lost	2/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Estuarine Systems Project	TAMU (Roy Hann)	1968-71	CHM	7,000	digital	data never published	cards destroyed data lost	0/0
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	CHM	15,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Statewide Monitoring Network, bay + tribs	TWC	ca.1970-present	CHM	150,000	digital	poor downloading capabilities, usually hard copy	available TWC	3/3
Estuarine Water Quality	TSDH	ca.1970-	CHM	30,000	digital	mainly coliforms	available TSDH	0/2
Houston Ship Channel	HCPCD	1970-81	CHM	40,000	hard copy	kept on file at lab	destroyed in 1981 fire	0/-
O&M Dredging Project	USCE/ Galv	1971-72	CHM	2,000	hard copy	comprehensive water sampling	available at Galveston Dist.	1/-
O&M Dredging Project	USCE/ Galv	1974-80	CHM	12,500	hard copy	comprehensive water sampling	available at Galveston Dist.	2/-
Galveston County near-shore & tributaries	GCHD	1972-pres	CHM	175,000	hard copy	field & lab sheets	on file at GCHD	1/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	CHM	3,000	digital	raw field sheets exist (1 copy USCE warehouse)	cards destroyed	1/0
Intensive inflow study entire bay system	TWDB	1976	CHM	5,500	hard copy		TWDB files	1/-
Bays & Estuaries Program	TWDB	1976-89	CHM	15,000	digital	Coastal Data System	available TWDB	0/3

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal Fisheries Sampling	TPWD	1975-pres	CHM	35,000	digital	randomly selected stations	available from TPWD	1/4
Houston Ship Channel	HCPCD	1981-pres	CHM	32,000	hard copy	some digitization since October 1988	on file at lab	1/1
O&M Dredging Project	USCE/ Galv	1980-pres	CHM	11,500	digital post 88	comprehensive water sampling	available at Galveston Dist.	2/3
O&M Dredging Project	USCE/ Galv	1971-72	SED	2,000	hard copy	comprehensive sediment sampling	available at Galveston Dist.	1/-
O&M Dredging Project	USCE/ Galv	1974-80	SED	7,500	hard copy	comprehensive sediment sampling	available at Galveston Dist.	2/-
Statewide Monitoring Network, bay + tribs	TWC	ca.1975-present	SED	7,000	digital	poor downloading capabilities, usually hard copy	available TWC	3/3
Submerged lands survey	BEG	1976	SED	6,000	hard copy	on digital WP file, but nontransportable	published by BEG	4/0
O&M Dredging Project	USCE/ Galv	1980-pres	SED	11,000	digital post 88	comprehensive sediment sampling	available at Galveston Dist.	2/3

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Siltation Study of Galveston (Hydrodynamic Survey)	USCE/ Galv	1936-37	HYDG	30,000	field sheets	some graphs exist in project report (1 copy extant)	data lost	0/-
Houston Ship Channel Model Study, Prototype Data	USCE/Galv	1964-65	HYDG	7,000	hard	Some open-bay stations	available Galveston Dist.	1/-
Galveston Bay Project, High-frequency	TWQB	1968-72	HYDG	500	digital	no hard copy	tape lost	0/0
Littoral Environment Observation	USCE/CERC	1974-80	HYDG	6000	digital	Surf-zone obs on Galveston & Bolivar Gulf shoreface	Available from CERC	0/1
Intensive inflow study entire bay system	TWDB	1976	HYDG	7,500	hard copy	most velocity profiles lost	TWDB files	1/-
Intensive inflow study entire bay system	TWDB	1989	HYDG	40.000	digital		available TWDB	1/2
Coastal sampling & special studies	TGFOC (now TPWD)	ca 1936 -ca 1950	BIO	20,000?	field sheets	Operations out of Marine Lab at Rockport	Lost, probably in 1971 Rockport fire	0/-
Coastal sampling, oyster studies, GB Survey	TGFOC	ca 1950 -1961	BIO	100,000?	field sheets	Operations out of Seabrook Lab	Lost in Hurricane Carla, 1961	0/-
East Bay/Rollover Pass Survey	TGFOC	1954-55	BIO	20,000	field sheets	only generalized results in 4 journal papers	field data lost	0/-
Survey of buried & exposed shell, Galveston Bay	Turney for SSOCA	1954-58	BIO	n/a	maps		lost	0/-

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Houston Ship Channel Ecological Survey	Humble Oil Co.	1957-58	BIO	4,500	field sheets	most data in project report (TWC library)	see report, field sheets unavailable	1/-
Clear Lake Shrimp Survey	USBCF	1958-59	BIO	23,000	hard copy	summary data in Chin dissertation	lost	0/-
Galveston Bay Fishery Survey	USBCF	1958-67	BIO	160,000	digital	no printouts remain	cards destroyed data lost	0/0
Coastal Fisheries Sampling	TPWD	1962-75	BIO	100,000?	field sheets	10-40 routine stations	data lost Olmeto warehouse?	0/-
Biology of sand seatrout	TPWD	1966-68	BIO	2,000	hard	reduced data in journal paper only	not available	0/-
PH Robinson SES Surveys	TAMU	1968-69	BIO	8,000	hard copy	most data in project reports & theses	see reports	3?/-
Cedar Bayou SES Studies, sponsored by HL&P	TAMU	1968-73	BIO	100,000?	digital	hard copy sent to EPA some printed in reports	tapes unreadable, paper destroyed	3?/0
Galveston Bay Project, Ecological Survey	UT MSI (Copeland)	1969	BIO	25,000	hard copy	some data in project reports & theses	field sheets lost	1/-
Dickinson Bayou Study	UTMB (Faget)	1972	CHM, BIO	480	hard copy	some data in journal paper	data lost	0/-
Wallisville EA Study, Trinity Delta Marsh	USCE/ Galv	1975-76	BIO	8,000	digital	one print-out exists	cards destroyed	1/0

Table 9 (Continued)

Project Name or Description	Agency	Period	Data type(s)	Number of obs	Format	Notes	Status	Code*
Coastal Fisheries Sampling	TPWD	1975-pres	BIO	400,000	digital	randomly selected stations	available from TPWD	1/2
Ecological survey of Trinity Bay	TWDB/ EH&A	1975-76	BIO	8,000	digital	field sheets lost	tape undecodable	0/0

*Status Code p/d p=paper format: 0 - lost, 1 - rare (e.g., one copy extant), 2 - scarce, 3 - exists but inconvenient, 4 - available
d=digital format: "-" - never existed, 0 - digital form lost, 1 - hard-to-acquire or poorly maintained, 2 - well-maintained but rarely disseminated, 3 - well-maintained but inconvenient format, 4 - well-maintained & transportable

By scanning the "observations" column of Table 9, one obtains a rough impression of the intensity and importance of the data programs to the scientific study of Galveston Bay as a whole. (Of course, this table does not communicate the content of a data set for a particular region of the bay, or for a special problem, e.g. decline of oyster abundance.) There are many other data collection projects not listed here because their total content falls below the general threshold of this scale of project, of 500-1000 or so observations. Most academic studies and many special studies of federal and state agencies fall below this threshold. Fig. 3 displays the relative ranking of project data-set content for two categories, water chemistry and biology, of all data collection projects in the bay. Most of the figure is drawn from Table 9, with the smaller projects estimated (and probably underestimated). It is apparent that the data resource for the bay as a whole (in contradistinction to a particular region or a specific period of time) is dominated by a few large-scale collection activities, with numerous much smaller projects. This does not imply, however, that the smaller projects may be ignored. The *cumulative* information in these smaller studies exceeds that in most larger projects. Further, these smaller projects may fill important gaps in the space-time record.

The time history of data collection in the bay is also of great interest. This history since 1950 is roughly indicated in Figs. 4-6 for salinity/temperature, water chemistry and biological data. (Again, this was drawn from Table 9 by assuming a uniform rate of data production over the course of each project.) Generally, data collection intensity peaked about 1970, and has been declining since. Some of the programs, such as the Texas Water Commission Statewide Monitoring Network (a.k.a. Stream Monitoring Network) and the county health departments monitoring, sample the tributaries of the system as well as the bay. On Figs. 4 and 5, the specific sampling restricted to the bay and Houston Ship Channel is shown separately, to give a better indication of data collection in the bay system *per se*. Also in Fig. 5 the marked increase in data generation subsequent to 1970 is perhaps misleading. This was due to increased interest in a wide spectrum of parameters such as metals and organic toxicants coupled with analytical methodologies (e.g., mass spectrometry) that permit a large generation of parameters from a single sample/procedure. The intensity of data collection in terms of water samples pulled from separate stations has in fact declined sharply since 1970. In Fig. 6, the collection of biological data since about 1975 has become dominated by the activities of the Texas Parks and Wildlife Department. The importance of this data collection enterprise, in terms of the raw numbers of observations made, cannot be overstated.

One potentially significant program is missing from the CHM category of Table 9 and from Fig. 5, viz. the water sampling in the Houston Ship Channel and upper Galveston Bay performed by Dr. Walter Quebedeaux of Harris County Health Department. This program was in place since the 1950's, but its results were jealously guarded by Dr. Quebedeaux, who frequently described the intensity of the program but rarely released any data. We can find no record of the data, and there is even doubt that intense systematic data collection really took place, in that a few long-term employees of the department were ignorant of the program or avouched that it never existed. In any event, if the data existed, it is probable that it was lost in the 1981 fire at the Pollution Control Department lab.

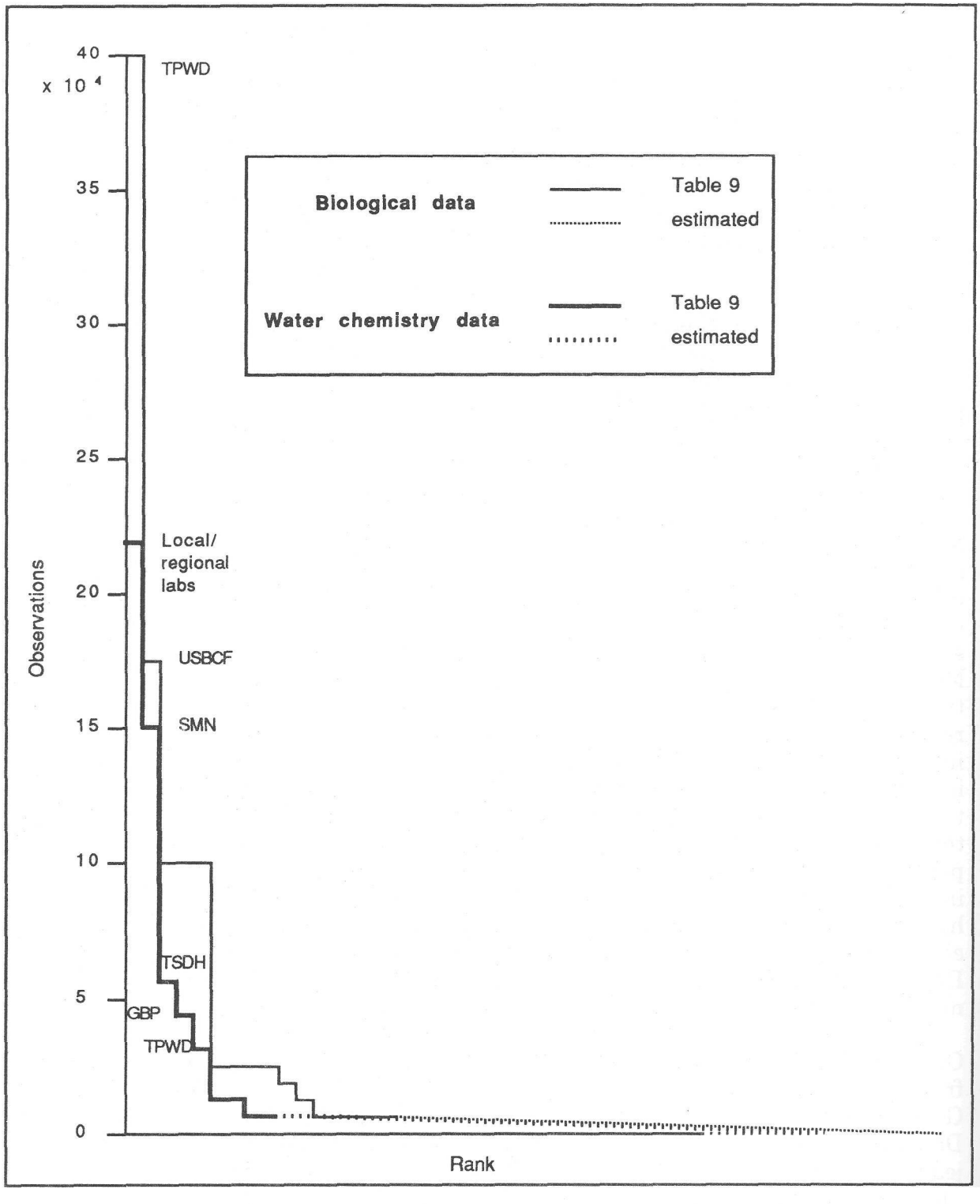


Figure 3. Galveston Bay data sets rank ordered by number of observations

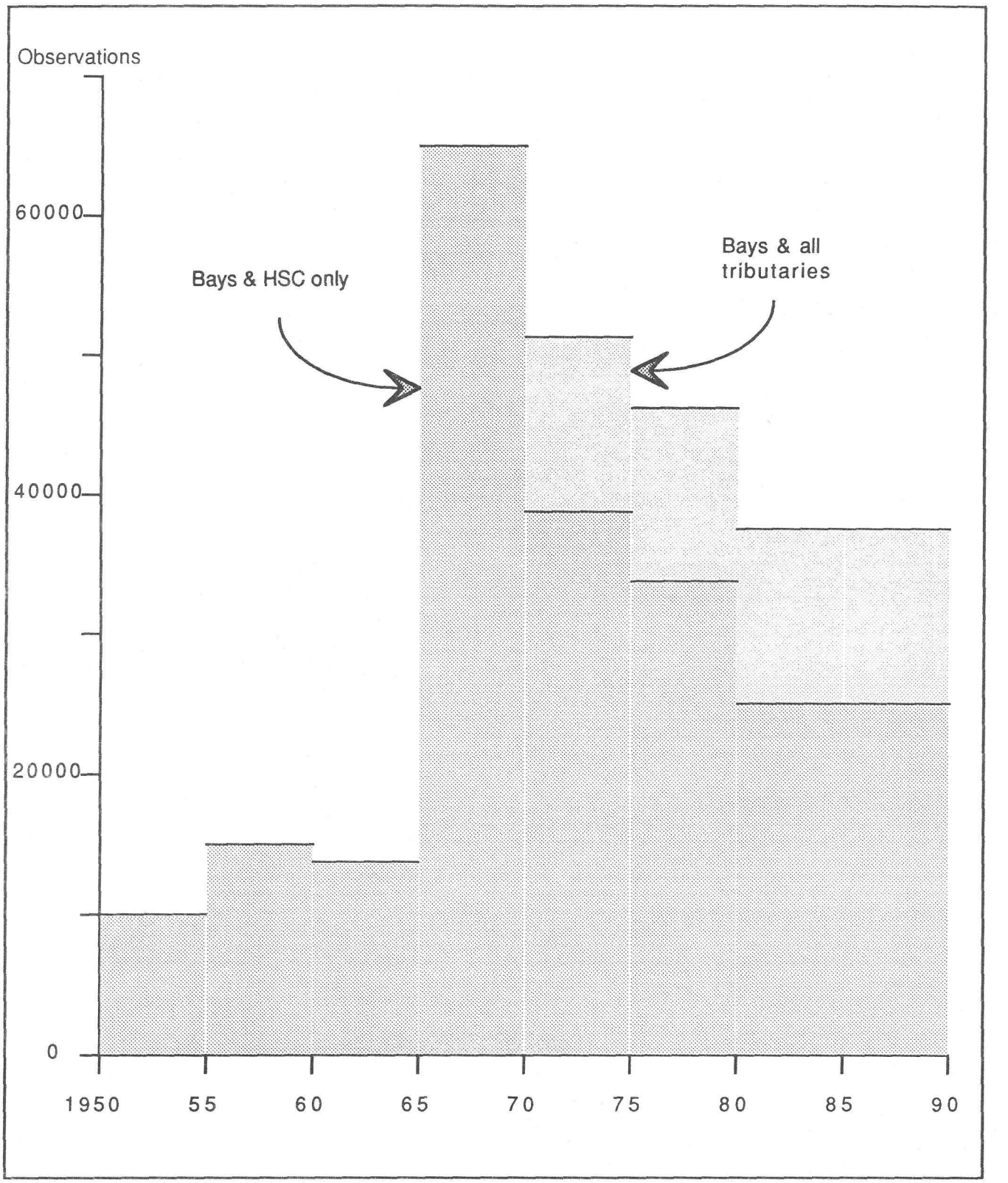


Figure 4. CST sampling intensity in Galveston Bay 1950-1990

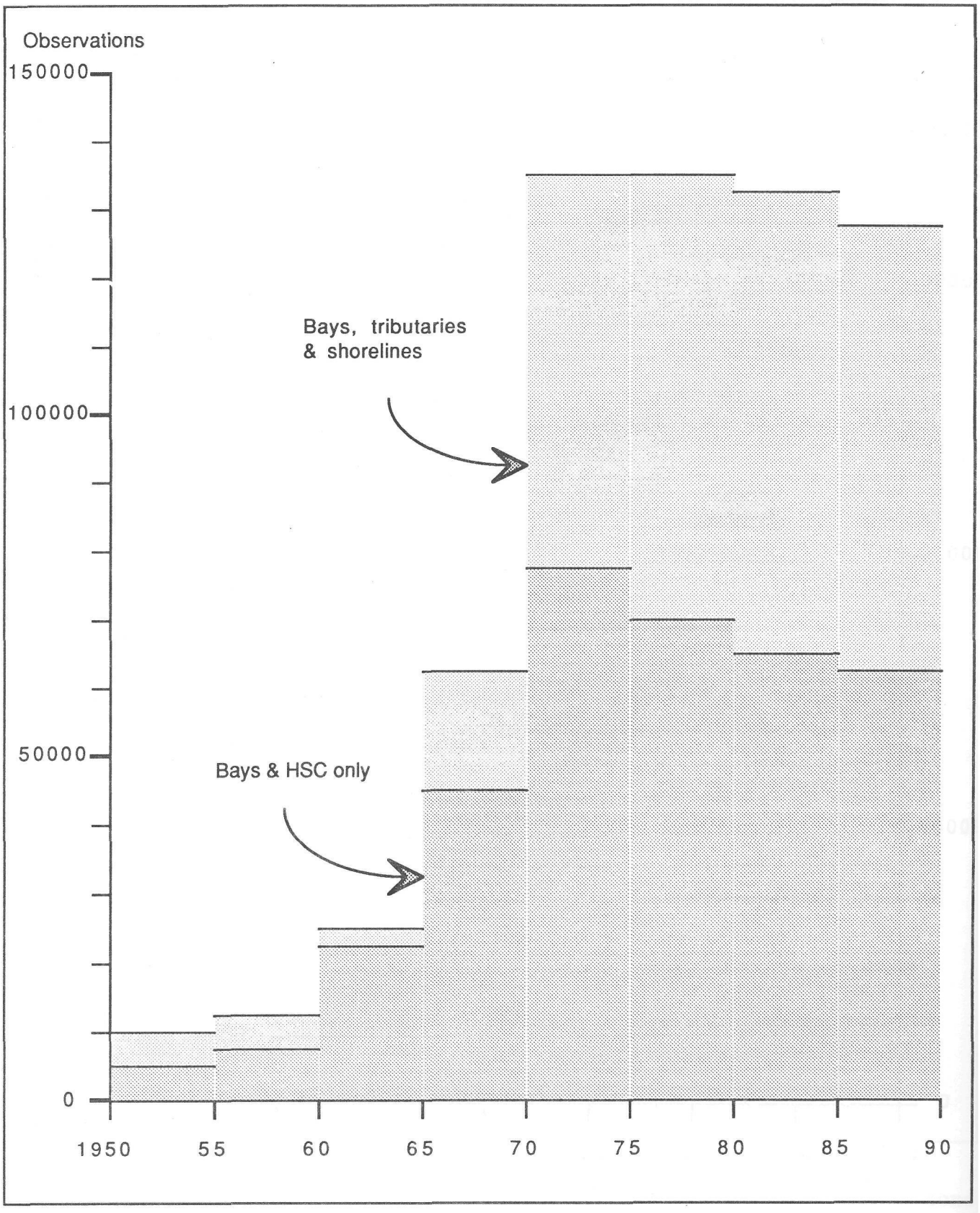


Figure 5. Water chemistry sampling intensity in Galveston Bay 1950-1990

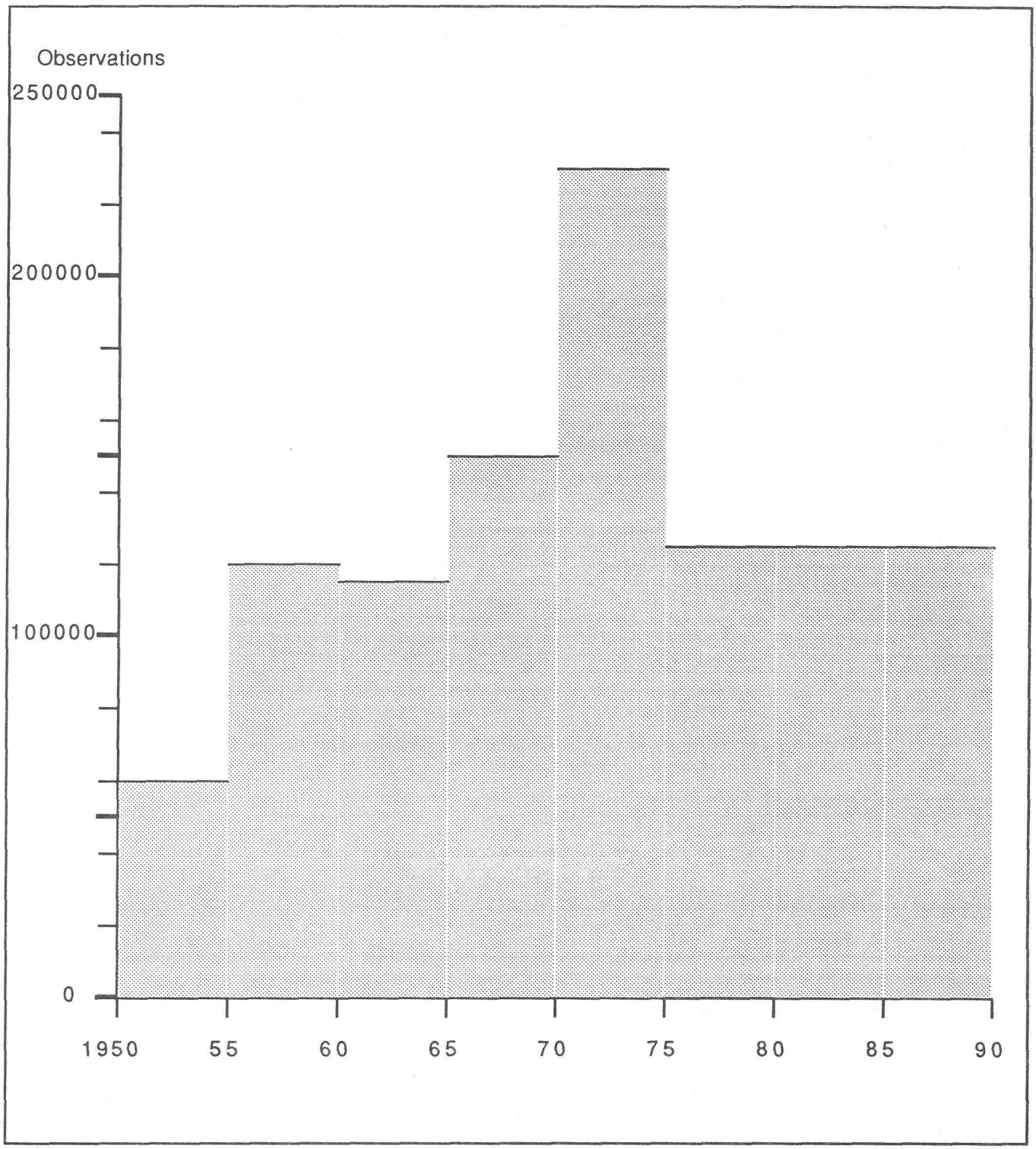


Figure 6. BIO sampling intensity in Galveston Bay 1950-1990

One must recognize that if a data set with tens of thousands of observations exists only in some hard-copy form, for practical purposes a researcher will view that data set, as inaccessible for serious analysis, as a small child viewing a puppy in a pet store window. The resources rarely exist in a research project, even in a state or federal agency, to undertake the keyboarding of such vast data sets. Thus when a data program includes provision for a digital record of the data, it is nearly as devastating to have that digital record lost as to lose the raw data itself. The occurrences of codes in Table 9 such as 1/0 or 2/0, indicating loss of the digital record--even though some hard copy form remains--should be especially noted.

Figures 7-9 summarize data accessibility for salinity/temperature, chemistry and biological data, as a function of the age of the data (i.e., the dates when collected), expressed as a percent of the total data resource for the same time period. The immediate impression one obtains from these figures is an appalling rate of data inaccessibility that approaches 100% for data older than the 1960's. Inspection of Table will confirm that the majority of the "inaccessible" data is in fact lost. The many smaller projects not reflected in Table 9 exhibit, if anything, a higher rate of data loss. For practical purposes almost everything prior to 1950 has been lost, and this includes some substantial data collection efforts. The reasons contributing to this high loss of data are examined in the following chapter.

Table 9 summarizes the status of extant data in several technical categories effective 31 December 1990. We are pleased to report that since that date several major data sets have in fact been located. This resulted from a combination of serendipity and persistence. Specific major data sets located by this project, in fact rescued from the edge of the abyss, include:

The Galveston Bay Project data sets, i.e. the Routine Monitoring, BOD data, and High-Frequency Programs, in *digital form*
The 1936-37 hydrographic data of the Corps of Engineers
The 1972-74 Joint EPA/TWQB Waste Source Survey
The Trinity Delta survey of the USCE
The Trinity Bay Study of HL&P associated with Cedar Bayou SES in digital form
The Texas State Department of Health Galveston Bay Project of 1963-67.

For example, the 1936-37 hydrographic data of the Corps were collected to provide supporting and verification data for a physical model to be built at the Waterways Experiment Station (the first of four physical models of Galveston Bay eventually built at Waterways). Apparently the only remaining trace of this project is one copy (marked DRAFT) of the project report in the Galveston District library:

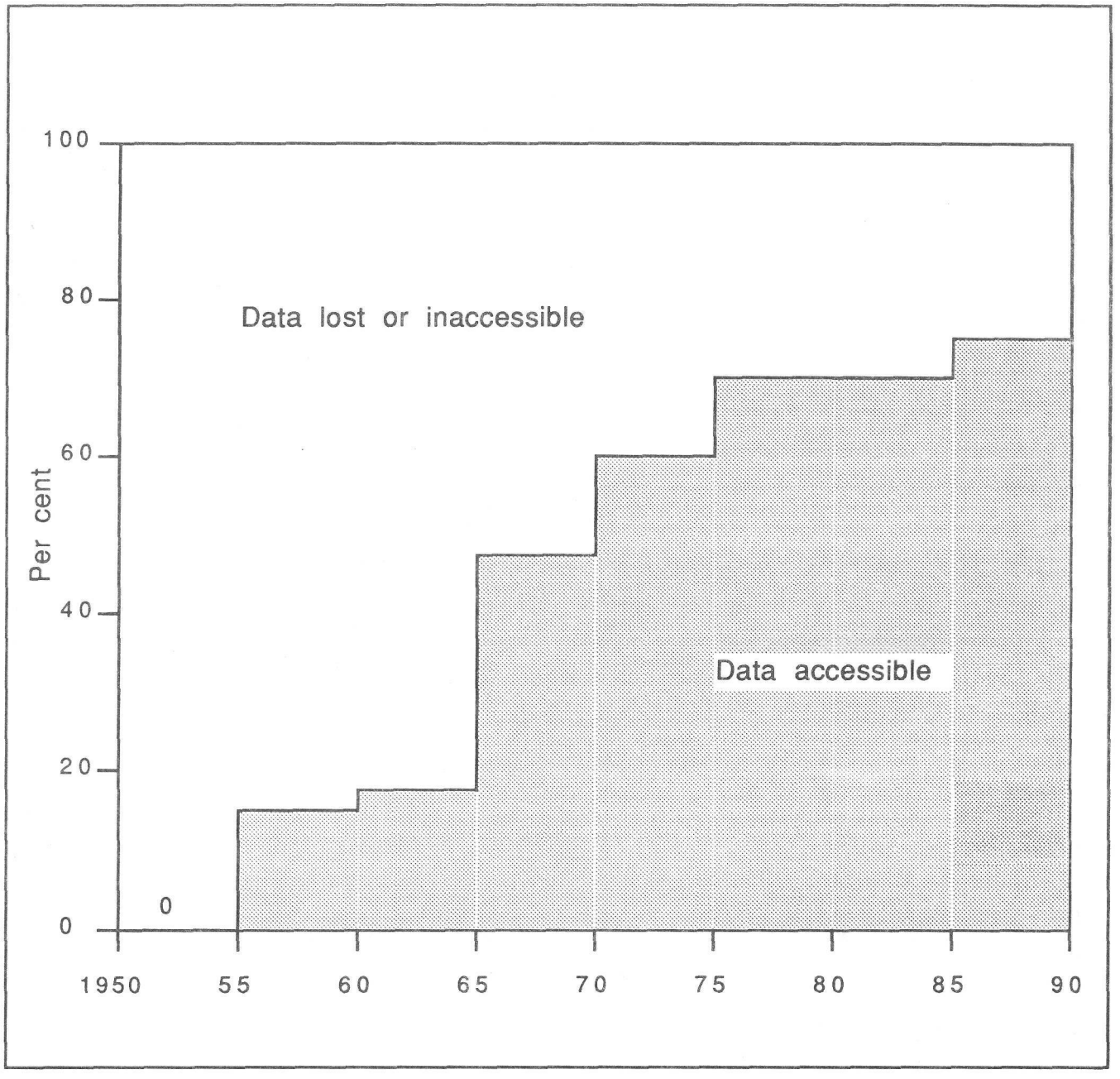


Figure 8. Availability of chemistry data in Galveston Bay (Bays & HSC) 1950-1990

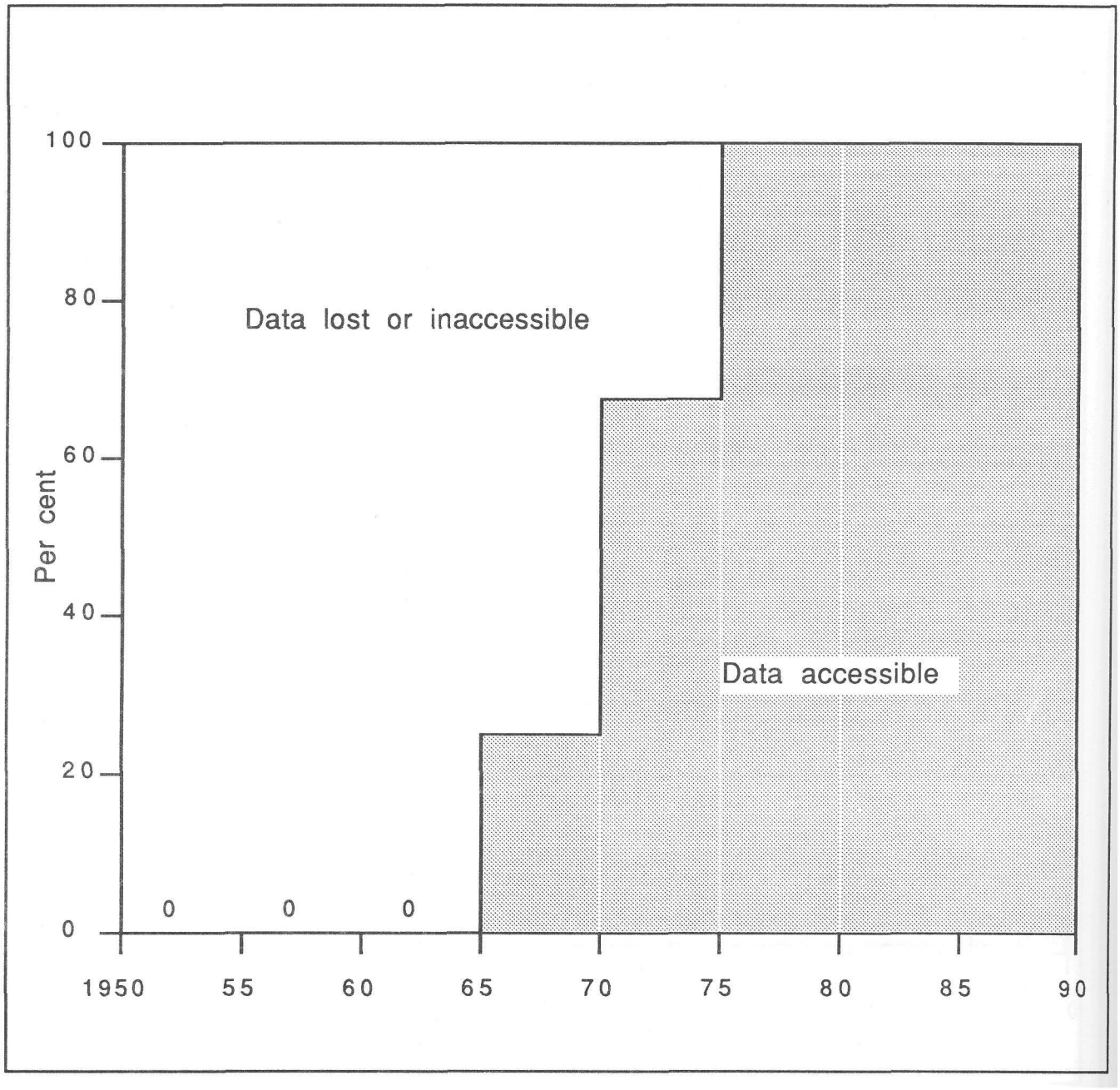


Figure 9. Availability of biological data in Galveston Bay 1950-1990

USCE, 1942: Report on Galveston Bay, Texas, for the reduction of maintenance dredging. U.S. Engineer Office, Galveston, Texas.

This is one of the most extensive hydrodynamic surveys ever performed in Galveston Bay, including detailed current profiling over extended periods and under a range of conditions. (Indeed, one may appreciate from the summary of Table 9 how pitifully small the data base on hydrography is for Galveston Bay.) Its value is augmented by the fact that the data were from a period when the ship channels were at a substantially shallower project depth than present. Only reduced current data appear in the report. The actual field sheets of current measurements have been "lost" for half a century. In the GBNEP Data Inventory project, the original field sheets were finally tracked down at the National Archives branch in East Point, Georgia. Unfortunately, the corresponding tide scrolls are still lost, but nonetheless these current data will form a valuable resource for future investigators.

Location of the TWQB Galveston Bay Project data sets is a good indication of the level of effort invested in this data inventory. During the Galveston Bay Project, all field data were entered into digital records and the originals discarded. When the Galveston Bay Project was concluded in 1972, Tracor, Inc., the company responsible for the data management aspect of the study, transmitted all data to the Texas Water Quality Board on a digital tape. Inexplicably, all copies and records of this tape have vanished. Even former TWQB/TDWR/TWC employees were sought in this project, but to no avail. The Routine Monitoring data were preserved in hard copy (i.e., reproductions of printouts) in the project reports--though to redigitize would be a major effort--but the High-Frequency data existed only in digital form, so this was a major loss. As a last resort, we sought copies of the data from the Austin consulting firm of Espey, Huston & Associates which was the last to have worked with the GBP data base back in 1973. It developed that EH&A had the data set at one time on punched cards. Years ago, the card holdings were purged, but copies were thought to have been transferred to tape. The EH&A computer center kindly provided digital copies of many pregnant-appearing files from these old tapes. These files proved to be intermixed segments of many unrelated card decks, containing model data, hydrologic data, accounting files and surveying logs from the company operations during the early 1970's. Among these, separated, interspersed and generally jumbled, could be recognized records from the Galveston Bay Project. We sorted through all of these records manually (which required writing several special-purpose codes, e.g., to decode old BCD characters, to read past imbedded end-of-files, and to copy off selected records), separating out the GBP data, and finally, after many tedious hours and the inspection of over 5 million characters, succeeded in reconstructing the data sets. While the rescue of this data set is certainly important, it should also be noted that this was one of the data sets which we assumed at the outset would be readily available to the project.

Additional information on individual data sets and their present disposition are given in the Data Set Reports, in the Appendix.

5. CONCLUSIONS AND RECOMMENDATIONS

In summary, the Data Inventory Project proved to be far more complicated and time-consuming than we originally envisioned. The original schedule was doubled by extensions, and it is apparent that work on the project will continue after the project's completion marked by this report. Several factors contributed to the time expansion:

(1) There proved to be a large number of data sources for Galveston Bay, but only a minority could be described as major projects (e.g., the TWC Statewide Monitoring Network, the Galveston Bay Project, the TWDB Bays & Estuaries Program, etc.), i.e. the data resource can be described as a few major projects and a great many small projects, which served to multiply contact time and logistics;

(2) The point-of-contact approach failed, requiring much greater time and effort of the PI's to find and gain access to agency data holdings;

(3) In general, the response of the data sources to our inquiries has been poor, requiring multiple letters or calls, and requiring months (at best) to finally gain access to data: many are only now responding.

However, the dominant reason is that the management of older data--and by this we mean any data taken prior to 1980--is by-and-large a shambles. Much more effort was needed to locate and retrieve this data than expected. For all of these reasons far more PI time has been dissipated by searching and agency communication, than has been invested in actual data inventory and data base creation. The frustration of the time-consumption of tracking down misplaced data sets has been compensated (somewhat) by the conviction that this work had to be done, and, as we began to realize as the project developed, the sooner the better.

Some definite conclusions regarding the data resource for Galveston Bay can be drawn. These conclusions apply primarily to data on the biological, water quality and hydrographic features of the system, which are the most important insofar as the GBNEP objectives are concerned.

1. Most of the data sets for Galveston Bay taken prior to 1980 are presently inaccessible. The majority of this data appears to be irrevocably lost .

2. When one considers that the data prior to 1980 comprises the vast majority of data taken in Galveston Bay ever, in terms of sampling intensity (though this is compensated somewhat by the greater number of observations per sample due to modern metrological and analytical technology), this implies that most of the data resource has vanished.

3. The factors which have led to this loss of data are still operating today.

These conclusions of course must be qualified for specificity. For example, sediment quality data is of more recent concern, and has benefited from advances in analytical technology, so is in relatively good shape. Also, specific data collections with national

archival procedures are well-managed, e.g. the historical mapping of the National Ocean Service and its predecessor agencies, and the data collection efforts of the U.S. Geological Survey. Further, several important data sets have been recovered since 31 December 1990, including the Galveston Bay Project High-Frequency Program and the USCE 1936-37 program, and some which were previously unavailable have now been provided. On the other hand, there are enough major data sets that remain lost, including the USBCF biological program of 1958-67, data collections by the Harris County Pollution Control Department from the 1960's through 1981, the intensive studies of the Houston Ship Channel by Texas A&M, and the intensive sampling performed by the City of Houston and by Harris County in the 1940's and earlier, that the above conclusions still hold. Further, these are examples of *agency* programs; the situation is worse for research data of individuals.

We have identified seven principal factors that contribute to this data loss, as follows:

1. Low priority assigned to archiving and preservation of older data.

This is a reflection of human psychology. Once a project or survey is completed, there is a tendency to stack the results out of the way and move on to the next challenge. Many agencies operate under a pressure of time, which conspires against good archival practices. Some agencies have some form of data management currently in place. While this is encouraging, it is also precarious, in that these programs are sensitive to shifts in organizational emphasis. An office purge is forever.

2. Mission-specific agency operation: perception of old data as "obsolete" and archiving as an unwarranted expense.

The Corps collects hydrographic or water quality data to support, e.g., navigation projects in place or in planning. Once a condition survey has been used to determine the need for dredging, once a decision on spoil disposal is made, once a project design is completed, the data sets employed in those activities are no longer needed. The mission of Texas Parks and Wildlife is to monitor the state of the coastal fisheries. The present condition is always primary. The Texas Water Commission and EPA are concerned with the present loadings of contaminants and the enforcement of water quality standards. The level of loadings a decade ago, or even last year, are rarely pertinent to that mission. And so it goes. The value of data diminishes quickly with age in these kinds of problem-specific operations. Yet it is these agencies that are largely responsible for the bulk of data collection within the Galveston Bay system.

3. Personnel turnover, combined with little or no documentation.

Only a handful of people in an agency generally has immediate familiarity with a data base. If the data base is not currently in use, this number will decline due to turnovers. When the last of these leave, the institutional memory goes with them. This was apparently the fate of the Galveston Bay Project data tape, described above, as well as numerous other programs in both the public and private sector. In some instances we had agencies deny that a sampling program ever took place (despite historical documentation to the contrary). This problem is most acutely manifested in the case of a single principal investigator at a university. Most of the

rare data sets we succeeded in locating for this project resulted from contacting (finally) the one or two persons remaining in the agency that knew something about the data. In one instance, the sole remaining contact died shortly after locating and transmitting the data to this project.

4. Agency instability, i.e. dissolution, merging, reorganization, displacement & relocation.

Some data sets have survived by dint of being undisturbed, until this Data Inventory Project located them. With an office move, as parcels, files and boxes are shifted about, the exposure to loss or discard is greatly increased. The disarray and haste usually typifying such moves contribute to a "clean-the-house" mentality, exacerbated by snap judgments on the part of personnel in no position to appraise the value of information. The decision is forced to consider data sets whose retention is already tenuous. Clearly, any sort of instability that leads to such shifting of materiel increases the probability of data loss. Aerial photography is particularly exposed to such loss because it has a monetary value as salvage, due to its silver content, which further conspires against its preservation.

5. Natural calamities (fires, floods, hurricanes) in poorly protected housing.

This problem speaks for itself. We have had a surprisingly large number of losses to such events, cf. Table 9. Ironically, it is the large, centralized, difficult-to-duplicate sets that are most exposed. The usual problems of water leakage, faulty wiring, and deterioration operate everywhere, but the Texas coastal zone--where most of the Galveston Bay data is housed--is exposed to extraordinary hazards. The human tendency is to disregard the risk of extreme hazard: we cannot help but note that the new Galveston Bay Information Center is located on Pelican Island.

6. Changes in data management technology, without upgrading of historical files.

This is a surprising factor, at least to these authors. There are several forms of this technological hazard. The first is simple technological obsolescence. At the time of data entry, punched cards and 8-track formats seemed to be fixed technology. Now, they are virtually unreadable. There is a transition period, of course, when newer technologies replace the old, but the task of upgrading formats of large, rarely used data files is onerous and of low priority. Then, with the same apparent suddenness of the demise of the LP and the Magcard, the technological hardware support is no longer available. At this writing, many data sets are being "stored" on floppy disks. In five years, they could be as unreadable as 8-inch floppies are today.

A second variety of this hazard is software obsolescence, in which the encoding is no longer readable. This ranges from discontinuation of a proprietary software, to loss of the description of coding formats. The prominent example of the former is System-2000 data bases. There are several examples of the latter, in which there exist tapes containing numerical data which can be read but whose meaning is no longer documented.

The third form of this hazard is due to the increasing information density of digital storage. As large data bases are compressed into smaller physical dimensions, the possibility of physical loss is increased: an errant electromagnetic field, small fire, or simple mislaying can wipe out the equivalent of reams of data. Probably the most prevalent form of this hazard is the acquisition of parity errors on an archival tape, and data garbling by stray magnetic fields. (Desk-top speakers seem to be an inviting flat surface upon which to stack floppies.) As new high-density media begin to appear, e.g. the compact disc, the possibility of simple physical loss becomes greater.

7. Proprietary attitude toward data by individual PI's.

This has been an endemic problem in academia, but it is also too frequently manifested in federal and state agencies. We will not propose to analyze the causes of this mentality, which may be rooted in the publish-or-perish environment, the paranoia of being "scooped" in some great insight gleaned from data analysis, the notion that "information is power," the view that one's data is valuable, and the view that one's data is worthless. We will observe, however, that many data sets exist in only their original form, in the possession of the person or agency which originally collected it, and are unavailable for the use of other investigators. This is a major source of the category of "inaccessible" in Table 9 and Figs. 7-9.

The important observation about all of these factors is that they are self-exacerbating and mutually reinforcing. Low priority of data management implies poor housing and careless data management practices, and increases the exposure to discard due to agency instability. The existence of only one or a few copies of a data set, and its possession by one or a few individuals increase the potential of loss due to natural or technological hazards. All of these factors are continuing to work at present, and are creating the potential for further loss of data, which will be lamented in the future. In our view, the problem is critical.

The facile--and fatuous--recommendation to correct the situation would be to eliminate the above seven factors. We would proffer the following specific recommendations, which we believe to be more pragmatic and to lie within the purview of the Galveston Bay National Estuary Program or its participating agencies.

1. All sponsored research projects (including consulting contracts and interagency contracts) should include a *requirement* for preparation of a data report documenting the *raw* measurements of the project. If a digitized version of the data base is part of the project, transmittal of a copy on an appropriate digital medium should also be required, with written (hard-copy) documentation of formats and software operation. Compliance with this requirement should be a condition for any future contracts. For public agencies, the data so transmitted should then be subjected to the requirement of public distribution given in (3) below.

2. All projects internal to an agency, performed by an agency staff, involving observations and measurements should require preparation of a data

report. If a digitized version of the data base is part of the project, a copy on the appropriate digital medium should also be required, with written (hard-copy) documentation of formats and software operation. For public agencies, the data so transmitted should then be subjected to the requirement of public distribution given in (3) below.

3. In public agencies, the release of the data report and digital copy should be made mandatory after a certain calendar period, e.g., six months. (If the data is still under review, it should be so marked, but being under review should not be used as a reason for delaying release.) Reimbursement for the expense of copying is appropriate, but the price should be reasonable. After all, the public has already paid for it once.

4. All agency files and materials should be marked with a destruction schedule by its originator. For measurements and raw data, at least, the files should be marked "permanent storage, not for destruction." In some agencies, smaller but equivalent words may be desirable.

5. At least one hard-copy record of every data set should be maintained. This might be raw data sheets, or might be a print-out of a digital data record. Also, even when a data set exists in a digitized data-management format (e.g., a data base management software form such as Lotus or dBase), a separate version in general encoding format (e.g., ASCII) should be maintained.

6. Data Inventory and Acquisition Projects should be sponsored as soon as practicable, either internal to an agency, or through external contract, to extend the present activity for Galveston Bay, and to secure similar data sets for the other Texas embayments and for the Texas coast. In particular, holdings in the following agencies and sites should be retrieved, organized and, where appropriate, digitized:

   the Texas Parks and Wildlife Olmeto warehouse

   the U.S. Corps of Engineers: Galveston District, the Texas area offices and the Waterways Experiment Station

   the National Marine Fisheries Service laboratories in Galveston

   the major research universities in the Texas coastal zone

   private engineering firms, surveying companies and aerial photographic services

   the U.S. Fish and Wildlife Service offices in Houston and Slidell

7. Some centralized, cooperative data storage and management facility is needed, one which is divorced from the separate mission-oriented state and federal agencies. Emphasis should be on competence of staffing and an

appropriate delineation of scope. The Texas Natural Resources Information System could become this entity, but it suffers from many problems, not the least of which is adequate and stable funding, which presently prevent its serving this function. This recommendation, of course, exceeds the jurisdiction of the GBNEP agencies, but could profit from the strong unanimous support of these agencies. It is, however, the only long-range solution that is evident to us.

REFERENCES

Armstrong, N.E. and G.H. Ward, 1990: Galveston Bay Data Inventory Work Plan. Report to National Estuary Program, Center for Research in Water Resources, University of Texas at Austin.

Armstrong, N.E. and G.H. Ward, 1991: Galveston Bay Data Inventory Users Manual. Report to National Estuary Program, Center for Research in Water Resources, University of Texas at Austin.

Badgett, T., J. Devlin, J. Pepper and M. Bryan. 1989. Buyer's guide, database managers. *Personal Computing*, December 1989, 151-180.

Blackford, J. et al. 1988. Buyer's guide, relational data bases. *Personal computing*, November 1988, 173-228.

Galveston Bay National Estuary Program. 1991. Fiscal Year 1991 Workplan. GBNEP, University of Houston-Clear Lake, Houston, Texas.

Hightower, M. 1989. Galveston Bay Priority Problems List. Memorandum, TAMU Sea Grant, 9 November 1989.

Jacobsen, Bill. 1984. Choosing and Using a Data Base Management Program. Special Data Base Management Section, *Creative Computing*, September 1984, S-1 - S-15.

Sander, E.J. 1988. DataEase 2.5, Release 3. *PC World*, July 1988, 110-112.

Sverdrup Technology, Inc. 1989. Implementation Plan for Catalog of Data Bases, Directory of Investigators, and Electronic Bulletin Board. Draft Report to EPA, STI, Stennis Space Center Group, November 1989.

Ward, G.H. and N. E. Armstrong, 1991: Two moves equal one fire: the historical data base for Galveston Bay. Galveston Bay Characterization Workshop, National Estuary Program, Houston, 21-23 February 1991.

ACRONYMS & ABBREVIATIONS

AM	Texas A&M University (College Station)
AM/Civ.Engr	Texas A&M Dept of Civil Engineering
AM/Oceanog	Texas A&M Dept of Oceanography
AM/Ocn Engr	Texas A&M Dept of Ocean Engineering
AM/Wldlf Fish	Texas A&M Dept. of Wildlife & Fisheries
AM/ZooBot	Texas A&M Depts of Zoology & Botany
API	American Petroleum Institute (Environmental Committee)
BCF	U.S.Bureau of Commercial Fisheries (Now NMFS)
BJI	Bernard Johnson, Inc. (Houston)
CCA	Clean Channel Association
CCSU	Corpus Christi State University
CERC	Coastal Engineering Research Center, USCE (Ft. Belvoir & Vicksburg)
CESI	Coastal Ecosystems, Inc. (Dallas)
City of Galv	City of Galveston, various deparments
City of Houston	Various departments, especially Public Works
CLCND	Chambers-Liberty Counties Navigation District (Anahuac)
EHA	Espey Huston & Associates (Austin)
EPA/DC	Headquarters, USEPA
EPA/Hous	USEPA, Houston area laboratory
EPA/Reg VI	USEPA Region VI office (Dallas)
Galv Cnty HD	Galveston County Health Department
Galv Wharves	Galveston Wharves (Port Authority)
GBF	Galveston Bay Foundation
GCWDA	Gulf Coast Waste Disposal Authority
Geomarine	Geomarine, Inc. (Plano)
GLO	General Land Office
Harris Cnty HD	Harris County Health Department
HBOI	Harbor Branch Oceanographic Institute
HGCSD	Houston-Galveston Coastal Subsidence District
HL&P	Houston Lighting & Power
LAN	Lockwood, Andrews & Noonam (Houston)
LGL	LGL,Inc., Bryan
NASA	National Aeronautics & Space Administration
NASA/EROS	EROS Data Center, NASA
NASA/JSC	NASA, Johnson Space Center
NASA/Langley	NASA Langley Space Center
NAVOCEANO	Naval Oceanographic Office (Stennis)
NCSU	North Carolina State University
NCWQ	National Council on Water Quality (Washington, D.C.)
NMFS	National Marine Fisheries Service
NOS	National Ocean Service (Rockville, MD)
NWS	National Weather Service
O&G	Oil & Gas Industries, Houston-Galveston area
Pilots Assn	Pilots' Associations, Houston & Galveston
POHA	Port of Houston Authority
RRC	Railroad Commission
SJRA	San Jacinto River Authority
SWRI	Southwest Research Institute (San Antonio)
TAMU	Texas A&M University (see AM)
TAMUG	Texas A&M University Galveston Campus
TCB	Turner, Collie & Braden (Houston)

TCC	Texas Chemical Council
TENRAC	Texas Energy & Natural Resources Advisory Council (now defunct)
TI	Texas Intruments
TIA	Trinity Improvement Association
TNRIS	Texas Natural Resources Information System
TPWD/Aus	Texas Parks & Wildlife Dept., Headquarters
TPWD/Seabrook	Texas Parks & Wildlife Dept/Seabrook lab
TRA	Trinity River Authority (Arlington)
TSDA	Texas State Dept of Agriculture
TSDH	Texas State Department of Health
TWC	Texas Water Commission
TWCA	Texas Water Conservation Association
TWDB	Texas Water Development Board
TWRI	Texas Water Resources Institute, TAMU
U of H	University of Houston, various departments
U St T	University of St. Thomas, especially Storm Research Institute
USCE	U.S. Army Corps of Engineers
USCE/BEB	Beach Erosion Board (now CERC), USCE
USCE/Ft-W	Fort Worth District, USCE
USCE/Galv	Galveston District, USCEZ
USCE/SW Div	Southwest Division Office, USCE (Dallas)
USCE/WES	Waterways Experiment Station, USCE, Vicksburg
USCG	U.S. Coast Guard (Houston & Galveston)
USDA/SCS	Soil Conservation Service of the US Dept Agriculture
USFWS	U.S. Fish & Wildlife Service
USFWS/Galv	U.S.FWS Galveston Lab
USFWS/Slidell	National Coastal Ecosystems Team, USFWS (Slidell)
USGS/Aus	U.S. Geological Survey, Austin office
USGS/GHRC	Gulf Hydroscience Research Center, USGS (NASA Stennis)
USGS/Hous	U.S. Geological Survey, Houston Office
USGS/Reston	Reston (VA) Office, U.S. Geological Survey
UTMB	UT Medical Branch (Galveston)
UTSPH	UT School of Public Health (Houston)
UT/ Barker Cntr	UT Barker History Center
UT/BEG	University of Texas Bureau of Economic Geology
UT/Civ Engr	UT Dept of Civil Engineering
UT/CRWR	University of Texas Center for Research in Water Resources
UT/Geol	University of Texas Dept of Geology
UT/LBJ	UT LBJ School of Public Policy
UT/Mar Biol	UT Dept of Marine Biomedicine (Galveston)
UT/MSI	UT Marine Science Institute (now Laboratory), Port Aransas
UT/Zool & Bot	UT Depts of Zoology & Botany