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.