RPWG
JJ

MONITORING RECOVERY FOLLOWING THE *EXXON VALDEZ* OIL SPILL: A CONCEPTUAL MONITORING PLAN

PRELIMINARY DRAFT

Prepared for

RESTORATION TEAM WORK GROUP
Exxon Valdez Oil Spill Group
645 "G" Street
Anchorage, Alaska 99501

Prepared by

PARAMETRIX, INC.
5808 Lake Washington Blvd. NE
Kirkland, Washington 98033

In Association With

ABA Consulting
PO Box 1151
Capitola, California 95010

Goldstream Consulting
PO Box 83090
Fairbanks, Alaska 99708

Darryell Olsen
2980 Glenmorrie Drive
Lake Oswego, Oregon 97034

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

LIST OF TABLES

Tables

LIST OF FIGURES

Figures

SUMMARY

TO BE COMPLETED

1.0 INTRODUCTION

1.1 Background

The Exxon Valdez Oil Spill Trustee Council (herein referred to as the Trustee Council) is developing a Restoration Plan for the spill injured resources and damaged services. One option under consideration during development of the Restoration Plan is to implement a comprehensive monitoring program. The purpose of the monitoring program is:

- to assess the adequacy of natural recovery,
- to evaluate the effectiveness of restoration activities,
- to document long-term trends in the condition of resources and services affected by the oil spill, and
- to contribute to existing baseline data.

The Trustee Council initiated a planning effort to develop the first phase of a comprehensive and integrated monitoring program for resources and services injured by the *Exxon Valdez* oil spill. Phase 1 is development of a conceptual monitoring plan. The conceptual monitoring plan will provide the framework for the more detailed technical planning during Phase 2. The framework will be used by the Trustee Council to make decisions involving the selection and implementation of monitoring activities. The conceptual monitoring plan, or elements thereof, will be incorporated into the Restoration Plan.

1.2 Why Monitor Recovery?

The question, "why monitor recovery?", requires a two part answer. First, monitoring is key to determining if recovery has occurred. Recovery of resources and use of services depends in large part on the public's perception that the resources and services have recovered. This perception can only be based on reality if monitoring occurs. Likewise, decisions in managing the resources are largely influenced by their perception of resource recovery. These perceptions should also be based on information that can only be derived from monitoring recovery.

The second part of the answer to this question is that the credibility of the Trustee Council in making decisions regarding recovery also requires monitoring. The general public, special interest groups (e.g., subsistence, commercial fisherman), and agency technical staff can not be expected to support decisions of the Trustee Council in the absence of data documenting the status of resources and services.

Thus, monitoring is an essential component of recovery. Only through an adequate degree and duration of monitoring can the Trustees fulfill their responsibility to provide stewardship in the recovery of the injured resources.

1.3 What is a Conceptual Monitoring Plan?

A conceptual monitoring plan is an instrument identified by the National Research Council (NRC) (1990) in *Managing Troubled Waters* as a means to logically direct our nation's environmental monitoring. Its ultimate goal is to guide the planning and decision making process in any monitoring program to produce information that is useful in making management decisions and to communicate the status of natural resources to various interest groups. To reach this goal there must be considerable two-way communication between scientists generating information and users of the information (management agencies and public).

The NRC describes a conceptual monitoring plan as:

- a tool for developing and refining monitoring systems,
- a means for identifying elements to be considered for an optimum monitoring plan, and
- a guide for decisions on what to monitor, when, how, and where.

A conceptual monitoring plan is a means for establishing a relationship between those who require monitoring information and those who provide this information. It is a generic plan for establishing criteria and procedures desirable for implementing specific monitoring plans. It is a guide to decision making regarding monitoring activities. It provides guidance in dealing with variability and uncertainty in monitoring. The plan also provides a map for coordinating various monitoring activities.

As with any such tool, it is both how well the tool is constructed and how well the tool is used that determines its effectiveness. Our basic precept in constructing this conceptual monitoring plan is that it be the product of contributions by as many involved parties as possible. Thus, we have actively sought the participation of a large number of individuals through telephone interviews, a technical workshop, and by review of previously prepared materials.

1.3.1 Monitoring Plan Principles

There are two basic principles inherent in the conceptual monitoring plan. These principles follow:

- whenever possible, monitoring designs should reflect cause-effect relationships while accounting for variability and uncertainty, and

- specific design decisions (e.g. the number of stations, number of replicates, monitoring procedures, etc.) can be made only after objectives and related information needs are clearly established.

The goal of producing information that is useful in making management decisions will only be met if these basic principles are followed.

1.3.2 Essential Elements of a Conceptual Monitoring Plan

There are a number of elements essential to a conceptual monitoring plan as identified by the NRC (Figure 1). These elements include:

Needs:	To be successful, a conceptual monitoring plan must take its direction from the needs of the eventual users of the information produced by the plan.
Users:	Those who require monitoring information for management or use of the natural resource.
Environmental: Conditions	Knowledge of the existing basic features of the environmental resources and services these resources support.
Objectives:	Clear statements of the needs and expectations the users have for the monitoring program.
Investigators:	Those who will develop and implement specific monitoring plans, analyze results, and communicate monitoring information.
Sampling Design:	Technical approach for the hypotheses to be tested; what, how, where, and when to monitor;, and how data will be analyzed.
Implementation:	Strategy for establishing and maintaining monitoring activities and communicating information.
Conclusions:	Judgements reached by investigators and managers as to the status of resources and services based on monitoring information.

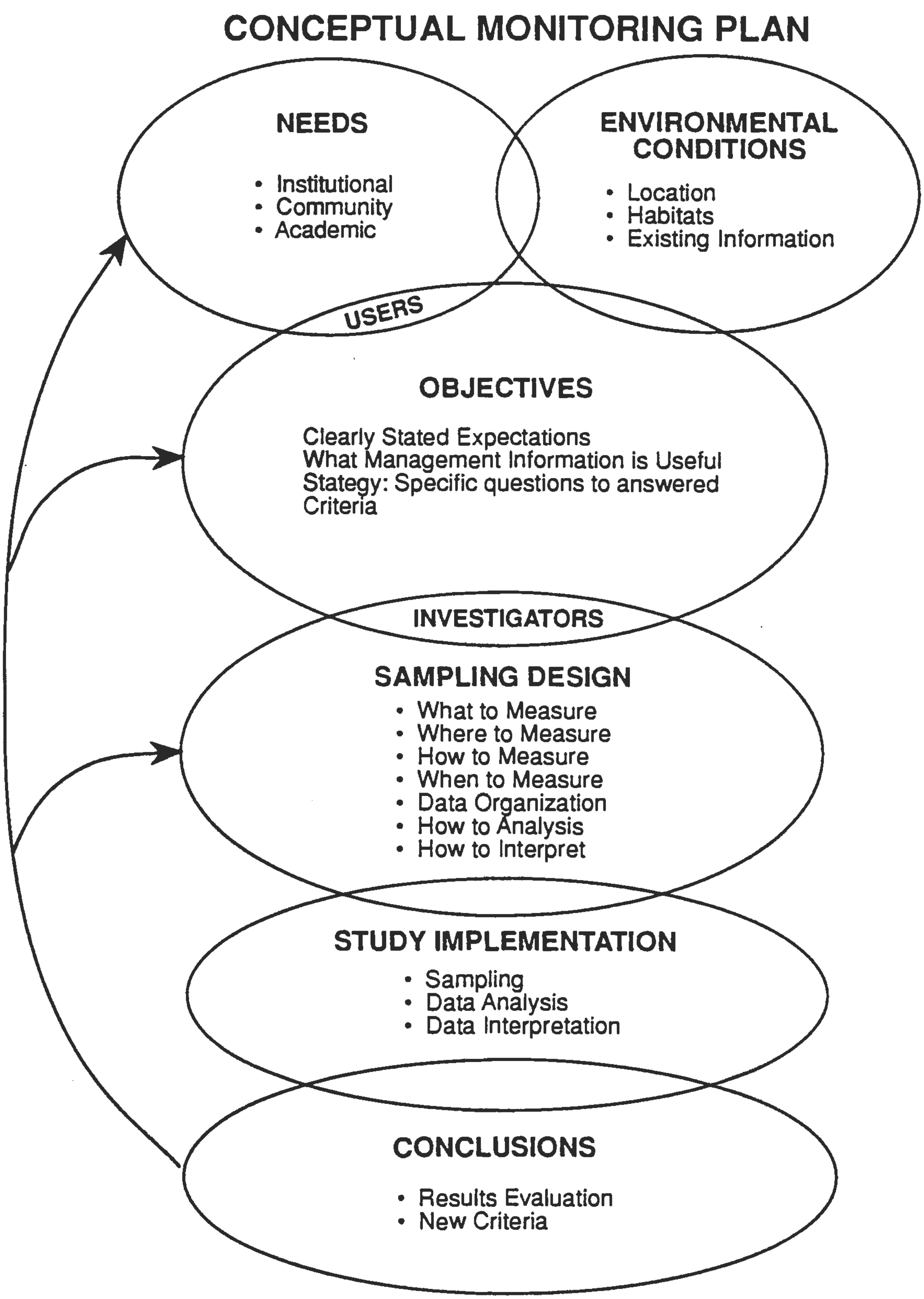

Figure 1.
Essential Elements of a Conceptual Monitoring Plan

1.3.3 How a Conceptual Monitoring Plan Can Be Used

A primary objective of the conceptual monitoring plan is to serve as a tool in designing specific monitoring plans for the resources and services to be monitored. Figure 2 shows how this tool can be applied to recovery monitoring in the oil spill area.

1.3.4 Goals and Expectations

The goals and expectations will define what information [objective(s)] is useful to the Trustee Council and investigators attempting to determine when resources have recovered or at what rate they are recovering. Development of the objectives requires two-way communication between the users of monitoring information and investigators who will produce this information. Development of the objectives also requires integration of public concerns and expectations together with the legal framework (Settlement Agreement).

These objectives should be unambiguous statements defining what constitutes useful information. They should require a cumulative assessment approach to provide a synoptic view of the injured resources. This synoptic view should:

- identify the recovery of multiple resources as well as cumulative recovery,
- describe levels of certainty anticipated,
- provide a framework for synthesizing monitoring information.

1.3.5 Study Strategy

The objective of developing a study strategy is to narrow the focus of monitoring efforts on questions and parameters of the resources and services that are most likely to produce the needed information. The study strategy identifies the resources (species and services) at risk or sufficiently in need of recovery monitoring. It also involves development of a conceptual model (not conceptual plan) that clearly states testable questions.

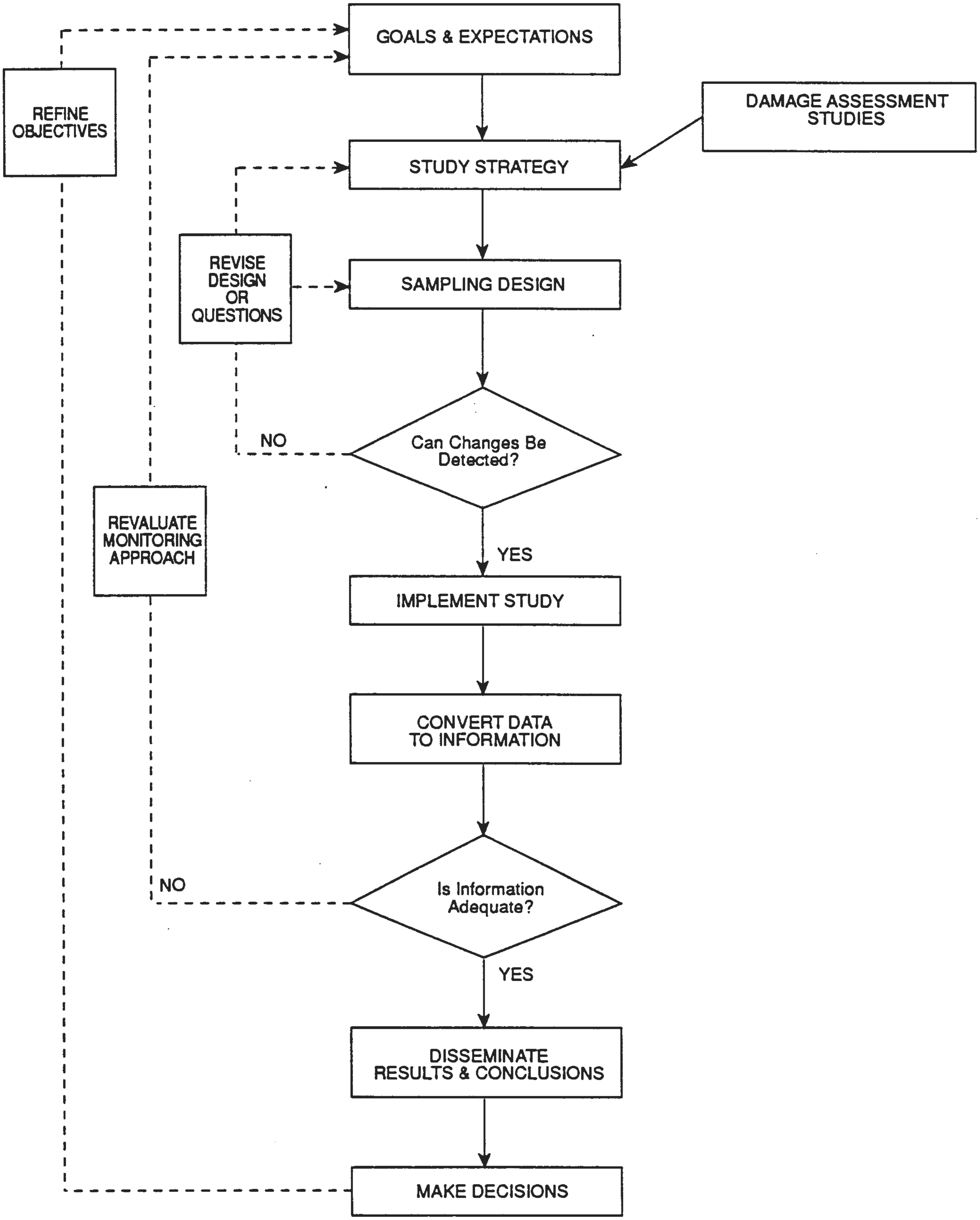

Figure 2.
Design and Implementation Elements, Recovery Monitoring Studies

Figure 3 illustrates the basic elements of such a conceptual model (not conceptual plan) for recovery monitoring. It illustrates that the Trustee Council, together with the investigators and interested public, should be involved in developing expectations. This conceptual monitoring plan involves the development of Trustee Council and investigator expectations. The plan indirectly includes public participation through the Public Advisory Group's review and comments provided to the Trustee Council. This participation has lead to the development of the goals and objectives of this conceptual monitoring plan.

Participation by the various parties has also led to the formation of general strategies to reach the identified goals and objectives. These strategies attempt to:

- define conceptual models,
- establish boundaries,
- develop predictions with estimates of uncertainty.

The conceptual models include testable questions. Clearly stated testable questions are a part of developing specific monitoring plans. These questions should identify links among ecosystem attributes. The questions must be testable within the constraints of the ecosystem, scientific techniques, and financial resources, as well as institutional constraints.

Boundaries established in the conceptual models include spatial, temporal, biological, and physical/chemical boundaries. These boundaries are based on information derived from the damage assessment investigations and restoration activities. Additional boundaries for the recovery monitoring are the legal constraints imposed by the Settlement Agreement and the practical, but undefined, boundary established by available funding.

Conceptual models should identify quantitative and qualitative changes in the resources expected during recovery. These predictions should attempt to identify the effects of resource and service management actions on the targeted resources and services. They should also identify the uncertainty likely to exist in measuring or estimating these changes.

Finally, the conceptual models should provide for review of predictions and testable questions during the coarse of investigations. These reviews should lead to refinement and reformulation where appropriate.

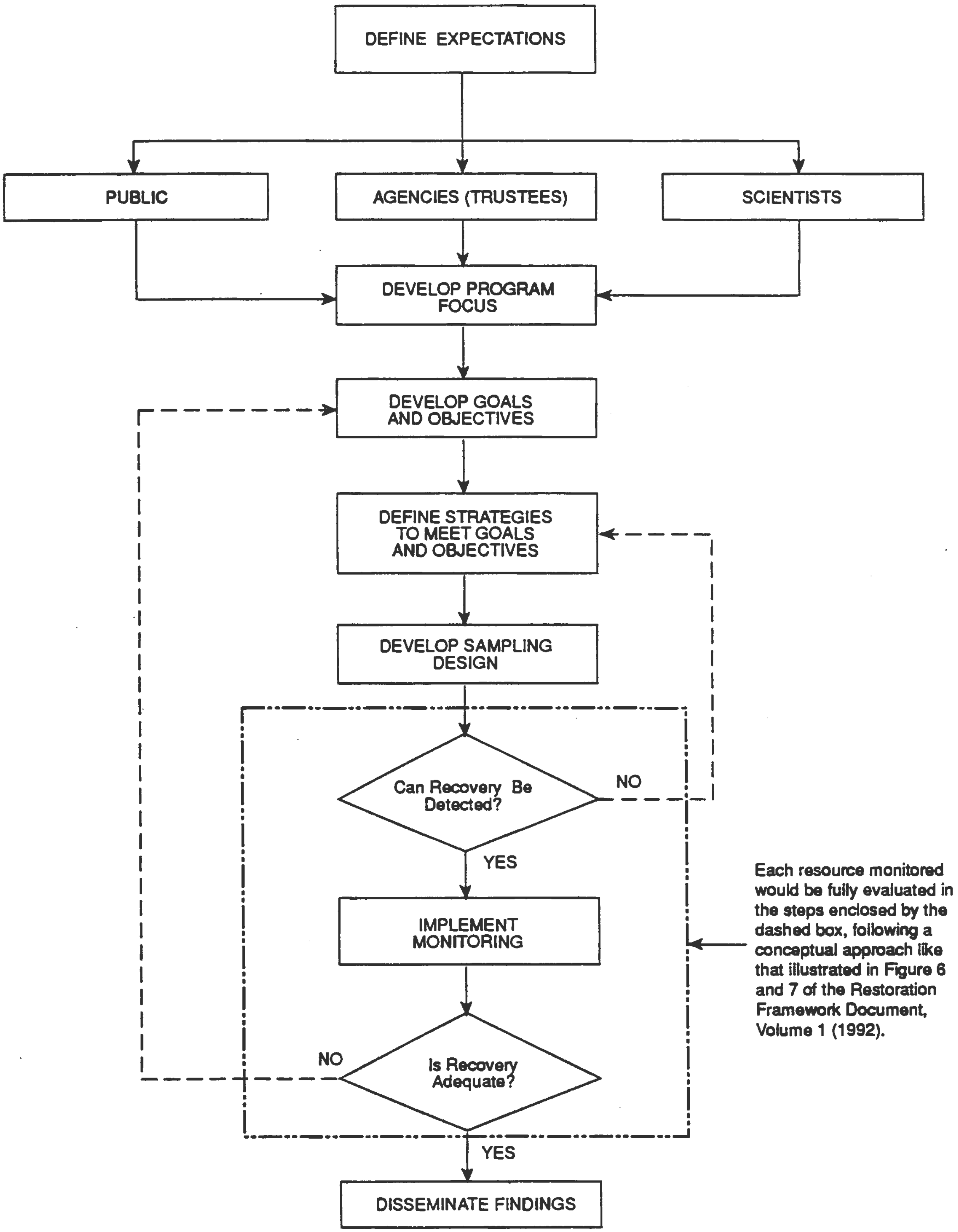

Figure 3.
Schematic of Conceptual Model

1.3.6 Preliminary Studies

The conceptual monitoring plan developed by the NRC (1990) identifies preliminary research as a key step to developing specific questions. In the case of the *Exxon Valdez* oil spill, the preliminary studies have taken place in the form of damage assessment studies (Figure 2). These studies are generally far more extensive, in both areal extent and degree of investigation, than is likely to be considered in most preliminary studies. The damage assessment studies are more than adequate to fill the role of preliminary investigations for the purposes of recovery monitoring.

1.3.7 Sampling Design

The key component in the sampling design for specific studies is the link to testable questions. Many of our nation's past monitoring studies have failed to meet expectations because they failed to link monitoring to testable questions. Key elements of a sampling design are discussed below. Information on development of sampling designs for specific monitoring activities in the oil spill area will be prepared as part of Phase 2 of the monitoring program.

Sampling designs have a number of key elements, most of which are obvious to investigators. The key elements were reviewed by the NRC (1990) to ensure that they are included in any well-planned monitoring activity. These key elements are:

- identification of meaningful kinds and amounts of change,
- identification and quantification of sources of variability,
- specification of how variability will be partitioned,
- decisions of what to measure,
- statistical models for selection of kinds and numbers of measurements,
- optimization and power analyses to ensure detection of meaningful levels of change,
- quality assurance objectives.

Meaningful change is based on the testable questions derived in the preceding step. Both the users and the investigators must contribute to defining what types and levels of change can be measured and how they will identify recovery of the resource.

There are many sources of natural variability that should be considered in developing a monitoring plan. Seasonal, cyclic, successional, and biological interaction are major sources of variability to be considered by investigators in developing specific monitoring plans. Although characterizing variability is difficult because of the many sources, incomplete understanding of the sources, and the scale of the marine environment, it is essential to develop a capacity to detect meaningful levels of change. Other man-caused sources of variability should also be considered.

Selecting variables to measure should focus on those most likely to provide information on recovery of the resources. The variables should address the testable questions identified in the preliminary studies. The choice of variables should be based on knowledge obtained in the damage assessment and restoration investigations and on the statistical capability to reflect change. Variables can include:

- early warning indicators (those most likely to detect recovery),
- sensitive indicators (those which have had the greatest damage),
- process indicators (those reflecting complex system interactions),
- high information indicators (those representing a number of different parameters or resources).

Statistical models should identify the more precise estimates that can be derived for the smallest sampling effort. This will help to select variables with information-to-noise ratios that are adequate to test the identified questions. The statistical models should define how questions will be evaluated and how variations from other sources will be interpreted. Statistical comparisons should be evaluated to consider the capacity to compare baseline conditions or reference areas, both of which are commonly difficult.

Sampling optimization and power analyses are means of both ensuring that objectives are met and that appropriate levels of effort are employed. These statistical techniques require quantitative estimates of the major sources of anticipated variability. Their application will lead to appropriate allocation of limited resources.

Quality assurance activities include quality control and quality assessment. Quality control is included within specific monitoring plans to ensure standardization of sample collection, processing protocols, analytical techniques, and technician training. It should provide a means to correct or remove erroneous data and resolve inconsistencies that degrade data integrity.

Quality assessment requirements are also incorporated into specific monitoring plans. They quantify the effectiveness of quality control procedures by instituting repetitive measurements, internal test samples, interchange of operators and equipment, independent verification of findings, and audits.

To be effective, quality assurance must be included in initial planning of the monitoring program. It must continue as an integral component of the total monitoring system through implementation and dissemination of information.

1.3.8 Data Conversion to Information

The objective of monitoring is to produce useful information rather than volumes of data. Through the organization, processing, and synthesis of data, together with knowledge, the data are endowed with reference and purpose, to become useful information. This useful information provides additional knowledge to be used in making decisions. Conversion of data to useful information involves planned data management, as well as planned data analysis and modeling.

Data management should be planned to provide easy access to data and related information by all users. Because of the amount, complexity and inter-relationships of data, it is essential to establish a computer-assisted data management system. It may be that all data should be stored in a central location or library. The data management system should consider data quantities, relationships of various data, quality assurance requirements, and types of analyses to be performed. "Data management activities are as important to the success of monitoring programs as the collection of data." (National Research Council 1990).

The objective of data analysis is to summarize and simplify data, to test hypotheses, and to measure change (recovery). These analyses should be planned as part of development of specific monitoring plans. To be successful, the analyses should summarize results, deal with linkages among data, use standard modeling approaches, evaluate assumptions, and evaluate sensitivity of analyses.

1.3.9 Dissemination of Results and Conclusions

It is obviously important that results and conclusions be disseminated to the users. Mechanisms and timing of reports to accomplish dissemination should be included in development of the monitoring plans. Status reports should be included to allow evaluation of monitoring efforts and adjustments where appropriate.

Management information is only produced when it is actually conveyed in a usable and accessible form.

1.4 MONITORING PLAN APPROACH AND DESIGN

Development of this conceptual monitoring plan relied, in part, on the report, *Managing Troubled Waters, The Role of Marine Environmental Monitoring,* produced by the NRC in 1990. This report describes the role of a conceptual monitoring plan in guiding monitoring efforts and provides guidance in preparation of a conceptual plan.

This draft plan also relies heavily on the input and advice from resource experts, principal investigators, agency representatives, and Restoration Team and Restoration Team Work Group members. The various components of the conceptual plan are, in large part, a synthesis of ideas and contributions we obtained by interviewing these many individuals. The value of the conceptual monitoring plan is derived primarily by their contribution.

Recovery of all of the resources and services injured, damaged or potentially injured/damaged by the *Exxon Valdez* oil spill cannot be monitored equally in time and space. The approach employed by this conceptual plan is to design a recovery monitoring program that accomplishes the following:

- defines the goals, objectives, and strategies for monitoring recovery,
- identifies resources and services, or elements thereof, that should be considered for monitoring,
- provides a mechanism to monitor on an annual basis and throughout the life of restoration funding (i.e., 10 years),
- provides guidance on considerations of sampling design
- identifies opportunities for comparing monitoring activities to and integrating them with other programs,
- identifies potential mechanisms to guide implementation of the recovery monitoring program.

The draft conceptual monitoring plan specifically addresses the questions stated in the Request for Proposal for the plan. These are stated below with a summary response. Throughout the plan each question is discussed in more detail.

1. What process or mechanism would best assist the Trustee Council in determining monitoring, damage assessment, restoration, and related project priorities?

The settlement from the *Exxon Valdez* oil spill resulted in money set aside specifically to restore, replace, enhance, rehabilitate, or acquire equivalent resources that were injured due to the spill and the reduced or lost services. The prioritization of activities funded by these monies should be driven by scientific endpoints and public concerns. As stated in the settlement, public involvement is to be an integral part of the restoration process. In order for the public to feel that recovery has been successful and that the settlement monies have been used properly, they need to be involved in the process. In order to gain the maximum knowledge and perform a scientifically credible program, resource and service experts must be involved throughout program development, implementation, and review.

The process or mechanism that would best assist the Trustee Council in establishing project priorities is one that utilizes both scientific and public review. Priorities should be based on those activities that have the greatest potential for addressing the objectives set forth by the settlement; specifically those that address restoration, land acquisition, enhancement, and rehabilitation of spill-damaged resources and services. In order to determine if these activities have been successful it will also be necessary to monitor. These categories, or types of activities, should also be prioritized and given weights, or percentage allocations of effort, that can be translated to funding. It may be useful to obtain feedback not only in overall perception of how best the monies can be allocated by activity, but also in a prioritization of how activities and monies should be allocated by resource and/or service.

For instance, initially, it would be useful to know how the public would like to see monies spent and what resources and services they are most concerned about. Then, to gain a scientific perspective, the resources and services and potential monitoring activities should be prioritized to determine what activities will provide the most information. Finally, the public feedback and scientific perspectives must be integrated. If, for instance, the public feel that land acquisition is important, this activity must be compared to the prioritization of injured resources and services, to determine which can benefit from such an activity.

A matrix can be used for assisting in the prioritization and linkage between type of settlement activity and the resources and services that would benefit from the action. In addition to prioritizing settlement activities, it is necessary to prioritize monitoring activities. Priority should be given to activities that are most likely to address the objectives of monitoring. A matrix, similar to that described above, can be useful in determining monitoring priorities.

2. What are realistic goals and objectives for monitoring?

Monitoring is essential to understanding if settlement activities have been successful at restoring, enhancing or replacing resources/services. The overall goal of the monitoring is stated in the draft Restoration Work Plan (reference). The over all goal is to develop a comprehensive and integrated monitoring program that will:

- follow the progress of natural recovery,
- evaluate the effectiveness of restoration activities, and
- establish an ecological baseline from which future disturbances can be evaluated.

This goal has been further broken down into specific objectives in Section 4.0 of the draft conceptual monitoring plan. The objectives have been reviewed by many individuals including Restoration Team and Restoration Team Work Group members, peer reviewers and principal investigators. The objectives in this draft of the plan reflect the input received from this review process. The objectives, as stated in this draft are ambitious and need to be further refined in Phase 2 of the monitoring program, when the bounds of the monitoring program are set. The extent of monitoring is monetarily driven, and "realism", other than technical feasibility, scientific merit (including how well the program element addresses the objectives), and public concerns/interests, cannot be ascertained by those other than the Trustee Council. The objectives stated in Section 4.0 are realistic in terms of technical feasibility and scientific merit. Cost ramifications or economic feasibility of monitoring alternatives will not be determined until Phase 2 of the program. With each subsequent phase of the monitoring, as well as during proposal review and throughout the actual monitoring, the progress and specific elements should be reviewed to determine how well objectives are being met.

From the interviews conducted with peer reviewers, principal investigators, and Restoration Team and Restoration Team Work Group members, the expectations for what the conceptual plan should accomplish follow:

- Identify what recovery monitoring should and should not attempt to accomplish.
- Identify monitoring goals.
- Establish criteria for selecting resources and services to be monitored.
- Identify strategies to ensure recovery monitoring is effectively implemented.
- Ensure natural and sample variation is taken into consideration and provide guidelines.
- Identify and prioritize resources and services to be monitored and why.
- Identify appropriate monitoring approach (e.g., indicator species, population level, trophic level, ecosystem characteristics) and provide rationale.
- Provide mechanisms for integration with other monitoring and management activities.
- Define why there is a need for monitoring.
- Establish requirements for dissemination of information.
- Provide baseline data for assessing future perturbations.
- Define "recovery".
- Establish a plan or framework to guide long-term monitoring.
- Recommend a mechanism for managing recovery monitoring.

Most of these expectations are addressed in Section 4.0 by the stated objectives and strategies for meeting the objectives.

3. What resources and services should be monitored and why, given the goals and objectives of the monitoring?

Through a process of interviews with resource/service experts, principal investigators, and Restoration Team and Restoration Team Work Group members, a set of criteria were developed to prioritize the resources and services to be monitored (see Section 5.0). From the interview process the criteria were also given weighting factors as to their relative significance in meeting the objectives of the monitoring program (workshop activity). The criteria will be applied to the identified injured resources and damaged services and the scores tabulated to allow prioritization of the resources/services based on how well they met the criteria (workshop activity). Additionally, the criteria will be applied to resources not directly injured by the spill, but those identified as ecologically linked to the injured resources and damaged services, such as forage fish. The result of this process is described in Section 5.0.

The process of prioritizing what resources and services should be monitored described above only takes into account technical versus economic aspects of monitoring resources and services. During Phase 2 of monitoring program development, the economic factors will be introduced within the technical design of each monitoring element (i.e., with proposal submittal for monitoring alternatives for each resource and service identified). The cost effectiveness of monitoring options as well as the application of the technical criteria will again be applied, this time to each monitoring option, to determine an overall prioritization of monitoring activities.

4. **Which clean-up, damage assessment and restoration science studies contain elements that would best serve the purposes of the intended monitoring program, and what are these elements?**

Damage assessment and restoration science studies that, to date, contain monitoring elements that address the overall goal of the monitoring program should be given added weight in prioritizing monitoring activities since a database already exists. The programs that are continued or supplemented with monitoring, should remain consistent with the earlier studies (with standardized units of measurement, overlap of the parameters measured, and study of the same locations and populations, etc.) so that recovery is not measured differently than injury, and the data are useful in comparing to pre-spill or control area data.

Once the resources/services to be monitored have been prioritized (as described above), the clean-up, damage assessment, and restoration studies can be reviewed to determine which of these contain elements that would best serve the purposes of the monitoring.

5. **Which surveys of services (e.g., recreation subsistence, aesthetics, etc.) provided by natural resources contain elements that would best serve the purposes of the intended monitoring program, and what are these elements?**

In other words, of the services directly linked to natural resources, which surveys would best serve the monitoring program? From synthesizing responses to this question from our interviews, the subsistence monitoring of shellfish tissue concentrations and consumption levels by Alaska Fish and Game, and the sport and commercial catch data collected by the state and U.S. Fish and Wildlife Service are thought to be the most useful survey of services to the monitoring program. Of course, the usefulness of these or other surveys of services, may change according to the prioritization of resources/services for monitoring. However, both of the programs mentioned are the responsibility of resource management agencies, thus their continuation may not be dependent upon spill settlement funds.

Additionally, surveys of people's perceptions on recovery would be useful, since in the final outcome the public must feel that the activities funded by the settlement have yielded information on recovery of the injured resources and damaged services important to them.

6. **What consideration should be given to the relationships among different monitoring components (e.g., sediments, shellfish, fish, mammals, birds, etc.), and how should they be integrated?**

Part of the overall goal of the monitoring is to follow the progress of natural recovery. Because the recovery of one resource may be directly linked to the recovery of another resource or physical/chemical parameter, it is critical to look at the linkages between resources to measure and interpret recovery. Therefore, in evaluating monitoring priorities, the linkages between resources and between resources and services should be considered. That is why some of the criteria developed to aid in prioritizing monitoring resources concern linkages between the resources and services.

To facilitate review of the linkages between resources, a matrix has been developed of the injured resources and damaged services, including also resources not directly affected by the spill but linked to the resources and services that were affected (Section 5.0, Table 1). The damage assessment and restoration studies and results of interviews with experts and principal investigators were reviewed to construct the matrix.

It is equally important that the linkages between resources and services that are monitored be clearly presented in terms understandable to the public in order that they can understand the worth and value of the program.

7. **What relationships need to be established with other monitoring programs within the spill area and how should they be integrated?**

There is value in identifying monitoring programs within and outside the spill area for several reasons

- monitoring may already be planned or underway that can provide answers to some of the objectives of this monitoring program, thus representing a savings of effort and money,
- monitoring programs may provide information on methods, natural variation, and the usefulness of monitoring particular elements,

- dove-tailing of programs may provide information on a global versus regional level,
- the monitoring in one program may influence the results obtained in another program, (i.e., through disturbance or enhancement of a site or population being studied), and
- to learn lessons from the experience obtained in other programs.

Several programs were identified through the interview process that may prove useful for some or all of the reasons mentioned above. Table 2 presents a matrix of the spill injured resources and damaged services juxtaposed to various elements of several monitoring programs. The monitoring elements are linked to the various monitoring programs through the listing in Table 2. A matrix such as this provides a tool for establishing links between this monitoring plan and those already in existence. It is also a tool for selecting parameters to measure to meet the needs for monitoring indicators of future perturbations.

Once the matrix table has been completed, the linkages established, and a selection of parameters made of those that will help in meeting the overall goal of this monitoring program, contact should be made with each of the programs to provide an opportunity for integration and/or coordination. The methodologies used in the programs should be evaluated by experts to determine their strengths and weaknesses. Any that prove suitable should be incorporated into this monitoring program by maintaining and/or requiring comparability in methodologies, reporting units, etc. Contact with the programs should be established to coordinate activities and ensure that the data are accessible to the Trustee's monitoring program.

8. What process (including infrastructure) should be considered to guide implementation and management of monitoring?

Depending on the funding level agreed to by the Trustee Council, monitoring may involve aspects covering each element of the overall goals listed under 2 above. Some monitoring of natural recovery and the documentation of long-term trends may fall outside the scope of restoration activities and as such, should be conducted by an independent party.

Monitoring the effectiveness of restoration activities should be a part of individual restoration activities, with perhaps, some information generated by other aspects of the monitoring program. For instance, if a restoration activity involves installation of a fish ladder, part of the restoration should include monitoring of activity at the fish ladder to ascertain whether or not the ladder is effective in allowing fish passage and improving the rate of passage. However, recovery of an overall fisheries resource, regardless of location of the ladder, may be the responsibility of the recovery monitoring program.

The two should be integrated so that the overall monitoring can account for the effect of the fish ladder on overall recovery.

In order for the information from the three overall goals to be useful to each other, it will be necessary for all monitoring programs to follow guidelines on standardized reporting units, data format, QA/QC, etc.

The process used to guide implementation and management of monitoring should also include frequent involvement of an independent, rotating pool of resource and monitoring experts. The same reviewers should not have responsibility for repeatedly reviewing a program. Review of monitoring efforts should include review of the monitoring protocol prior to implementation, review of restoration activities that have monitoring elements, ensuring that the objectives are addressed and the program is technically sound; and review of draft and final products resulting from both these efforts.

1.5 PLAN ORGANIZATION AND CONTENT

TO BE COMPLETED

2.0 WHY MONITOR?

2.1 VALUE AND USES OF MONITORING

Why should the Trustee Council devote funds from the *Exxon Valdez* oil spill settlement to monitoring recovery? Given the many demands for these funds, this is an essential question to answer.

Monitoring will allow the Trustees to:

- measure the success and rate of recovery of resources and services in the spill area,
- determine the effectiveness of natural recovery and of selected restoration projects,
- facilitate resource and service restoration, and
- generate new data and improve on existing baseline information to aid in detection of and response to effects of future oil spills or other perturbations.

Monitoring's greatest value to the Trustee Council is the public assurance and documentation that the injured resources and services are recovering. The *Exxon Valdez* oil spill and subsequent spills world wide have produced concern and fear among many people. This fear includes the perceptions that resources will never recover, the Trustee Council is not capable of ensuring recovery of resources and services, restoration will occur whether recovery demands it or not, and that settlement funds are being used to support activities that will not yield results. Well-designed monitoring activities are a vital element as they will assure the public that recovery is occurring.

The Trustee Council's responsibility for stewardship of the natural resources requires them to ensure that resources and services injured by the oil spill are recovering. This can only be accomplished through monitoring. Monitoring must be sufficiently rigorous and scientifically defensible to provide confidence to the public and the scientific community that recovery is documented.

Part of the overall goal for monitoring is to identify previously undocumented injuries that may exist. Through recovery monitoring the Trustees will provide a vehicle that may detect such injuries.

Perhaps more important, the information gather through recovery monitoring and through monitoring indicator parameters will provide a baseline, a lack of which proved a significant and overwhelming detriment to determining the extent and magnitude of the spill effects. Establishing a baseline for the future, along with documenting natural recovery and the effectiveness of restoration activities, should be perhaps the highest priority of the Trustee Council in their stewardship of the natural resources and services.

2.2 CONSTRAINTS ON MONITORING

The main constraint on monitoring is monetary. To monitor each of the injured resources and damaged services along the entire geographic area of the spill and throughout several generations would be cost prohibitive. This necessitates making selections of what resources and services to monitor in the hopes that the information can be extrapolated to other resources and across geographic areas.

Other constraints on monitoring include the general lack of baseline information for the resources and services. This lack of information limits the ability to statistically compare population changes and estimate variation.

3.0 DEFINITIONS OF RECOVERY, RESTORATION, AND LONG-TERM MONITORING

3.1 RECOVERY

Recovery is a term that means something different to different people. Recovery of the various natural resources, and the services they support, following the *Exxon Valdez* Oil Spill will occur at variable rates for different resources and will likely vary geographically across the spill area and between populations. Thus various degrees of recovery will be present in different resources and services and at different locations in the future.

It is necessary to define the term recovery. For the purposes of the conceptual monitoring plan, the term recovery means a return to "normal" or estimated levels of what current populations/conditions would be had the spill not occurred ("no spill conditions"). Recovery can occur through natural biotic and geomorphic processes as well as through restoration or manipulation of existing conditions to facilitate recovery. Recovery may also include replacement or enhancement of affected resources and services, or elements thereof.

For specific resources and some services, recovery to predicted "no spill" levels may not occur for many generations, if ever. Additionally, other factors (stresses) influence resources and ecosystems, both natural and anthropogenic, and return to pre-spill or no spill conditions may not be realistic or feasible. Recovery will most likely be the development of some steady state of conditions that may differ from those that existed before the spill.

Ideally, "complete recovery" of resources and services would include:

- at the original locations
  - in the original abundances
  - with the original population age-class structure
  - and the original biomass
  - with the original linkages with other resources/parameters (i.e., same prey items)

- use of the injured area
  - by the original user groups
  - to the original use levels
  - with the original attitudes

In order for monitoring to be an effective management decision-making tool, it is necessary to establish monitoring or recovery endpoints. Towards this end it may be necessary to define an achievable or "acceptable" level of recovery that may be less than complete recovery.

Because baseline or pre-spill information was not available in most cases, clearly defining the original conditions for either resources or services may not be feasible. Thus, we must identify other criteria for evaluating recovery. Pragmatically, recovery will be evaluated by investigation of only a sample of the species, habitats, and services affected by the oil spill and over a limited geographic area. Thus, it is necessary to identify key taxa and representative services that will adequately assess a spectrum of the injured resources and damaged services.

3.2 RECOVERY MONITORING

Recovery monitoring is the monitoring of natural, unassisted recovery of injured resources and damaged services. The primary focus of this plan is on recovery monitoring. The goal is to monitor the rate of natural recovery.

3.3 RESTORATION MONITORING

Restoration activities are those that involve anthropogenic manipulation to assist in the recovery of resources and services. Several of the restoration activities have a monitoring component to them to determine if the activity has been effective in restoring resources and services. This monitoring plan does not cover these existing monitoring programs; however, it will be coordinated with these efforts. Restoration monitoring covered by this plan will be limited to evaluating the effectiveness of select restoration activities. The decision on which restoration activities need to be monitored will be based on the Trustee Council's review of restoration studies. Those selected for potential monitoring can then be reviewed in light of the objectives and strategies described in Section 4.0.

3.4 LONG-TERM MONITORING

Long-term monitoring is defined here as monitoring that occurs over a five-year period, or longer, with the collection of data that enables trend analyses to be performed. The goal of long-term monitoring is to provide information on existing spatial and temporal conditions and natural variation such that changes due to perturbations can be detected. Long-term monitoring utilizes indicator measurements to detect change (e.g., sensitive physical, chemical, and/or biological parameters in which a change would be indicative of perturbation).

4.0 NEEDS, OBJECTIVES, AND STRATEGIES

Monitoring is essential to understanding if settlement activities have been successful at restoring, enhancing or replacing resources/services. The overall goal or need (used interchangeably) of the monitoring is stated in the draft Restoration Work Plan (1993). The over goals are to develop a comprehensive and integrated monitoring program that will:

- follow the progress of natural recovery,
- evaluate the effectiveness of restoration activities, and
- establish an ecological baseline from which future disturbances can be evaluated.

This goal has been further broken down into specific objectives in Section 4.0 of the draft conceptual monitoring plan. The objectives have been reviewed by many individuals including Restoration Team and Restoration Team Work Group members, peer reviewers and principal investigators. The objectives in this draft of the plan reflect the input received from this review process. The objectives, as stated in this draft are ambitious and need to be further refined in Phase 2 of the monitoring program, when the bounds of the monitoring program are set.

The following list and sequence of needs, objectives, and strategies of the conceptual monitoring plan reflects the general consensus derived from the interviews. Section 4.1 outlines needs, objectives, and strategies that pertain to all types of monitoring (e.g., recovery, restoration, and baseline), while sections 4.2 through 4.4 present issues that are specific to the elements of the overall goal.

4.1 GENERAL MONITORING PLAN

1. Need

Scientifically credible and publicly acceptable monitoring program.

Summary of Need

The monitoring program will be scientifically and publicly credible only if the individual projects are well-thought out, planned, and executed. Variability and uncertainty can be dealt with and minimized by the use of preliminary studies or historical data, reliable sampling, and analytical methodologies. The plans for the individual projects need to be subject to peer-review prior to project initiation and periodically throughout the project. All projects should also meet specified quality assurance and quality control (QA/QC) guidelines.

Objective

- Ensure a credible monitoring program, that whenever possible, limits the monitoring to testing hypotheses and sets limits on sample variability and account for natural variability for program elements.

Strategies

- Utilize a timely peer-review system to review proposals and reports for scientific credibility and technical feasibility, including their ability to detect change.

- Review recovery monitoring proposals (assess methods) and reports to ensure that, whenever possible, testable hypotheses are stated and uncertainties are addressed (sample and natural variation).

- Where needed, develop, or request development of, methods for monitoring.
- Develop a framework for QA/QC.

- Take public opinion and perception into account in developing the monitoring plan.

- Establish forums (e.g., scientific, community and agency participants) to evaluate effectiveness of recovery monitoring studies.

2. Need

- Secure funding for multiple years of monitoring.

Summary of Need

Long-term monitoring will be necessary to ensure that injured resources and damaged services have recovered. Recovery of several of the resources may not be detectable within a ten year period due to a variety of factors (e.g., time to reproductive maturity and fecundity). Due to this constraint, determining a long-term funding source should occur prior to implementation of some monitoring programs.

Additionally, even for resources where recovery can be measured in less than 10 years, the programs will likely involve multiple year studies, and/or periodic monitoring. To ensure that funding will be available to complete studies requiring periodic monitoring over several years, it will be necessary to establish a long-term funding mechanism.

Objective

- Fund long-term monitoring.

Strategy

- Establish an endowment to be used for long-term monitoring (i.e., greater than 10 years)

3. Need

- An accessible and/or integrated database.

Summary of Need

Accessibility of the data is critical for the monitoring to be of any value to resource managers, scientists, and the public. In order to be an effective tool for decision-makers and investigators, a catalog of the monitoring data, as well as other spill related data, should be centrally located and accessible by the various user groups. A centralized cataloging system will allow for the past, ongoing, and future data to be accessed to maximize the information gained from the spill and to allow for comparisons between and within resources/services.

Objectives

- To have knowledge of and access to existing *Exxon Valdez* monitoring, damage assessment, and restoration data.

- To have knowledge of existing monitoring and resource management data that may be useful in understanding recovery of resources/services injured by the oil spill.

- Ensure accessibility of monitoring data to the various user groups.

Strategies

- Develop a centralized, computerized catalog or library of databases that should include, but not be limited to, contact name/agency, parameters measured, resource/service studied, and when possible, the summary statistics calculated.

- Code existing and future *Exxon Valdez* oil spill databases with a common link for location/site and resource/service so that information on a resource/service is retrievable by a unique identifier, as is information on a location/site.

- Provide guidelines to principal investigators for standardizing components such as resource or location/site codes and reporting units, for ease in adding and retrieving data.

- Utilize a system that is user-friendly and provide step-by-step instructions on how to access and retrieve information from the catalog of databases.

- Design a flexible cataloging system to accommodate additional fields as new information becomes available.

- Identify an individual person (or agency) to oversee the centralized catalog, including acquisition of databases.

4. Need

- Consistency and timeliness in data reporting.

Summary of Need

To maximize the usefulness and compatibility of the data obtained through monitoring, standardization of reporting requirements and ensuring the timely submittal of results is necessary.

The guidelines developed will not dictate what methods investigators must employ to study their resource/service, rather the more general aspects to follow, such as reporting data in metrics, utilizing one of five possible software packages as a database software, etc.

Objectives

- Provide proposal/reporting guidelines (covering components such as publishing requirements, standardization of units, use of convertible software, status reports, QA/QC requirements, ideas on statistical methods to employ, etc.).

- Establish a method for ensuring timely submittal of deliverables.

Strategies

- Require periodic one page progress reports and project end reports with date of deliverables dependent on the resource/service-specific studies.

- Develop guidelines (covering components such as publishing requirements, standardizing units, convertible software, status reports, QA/QC requirements, ideas on statistical methods to employ, etc.) for principal investigators to follow.

- Develop recommendations for Request for Proposal and contract language that sets specifics for reporting and schedule commitments and penalties.

5. Need

- Information for long-term management of resources/services.

Summary of Need

Monitoring results provide a tool for decision-makers to determine which resources and services are recovering on their own and whether or not the rate of recovery is acceptable, which may never recover, and which may recover with human assistance.

Objective

- Provide information useful to decision-makers.

Strategy

- Collect long-term data documenting recovery of injured resources/services.

- Ensure accessibility of monitoring data to resource agency managers and other decision-makers, investigators, and the public.

4.2 RECOVERY MONITORING

1. Need

- Prioritization of resources/services to monitor, and elements thereof.

Summary of Need

Given that monitoring funding resources are finite, a series of decisions will determine how comprehensive and integrated the monitoring program will be. The recovery or restoration of some important resources and services may not be monitorable due to the physical properties of the system, biological properties of organisms, or logistical constraints in the area.

Objective

- Develop a method for prioritizing resources/services and the monitoring activities.

Strategies

- Develop selection criteria to prioritize resources and services to monitor.
- Establish priorities for recovery monitoring of selected resources and services by evaluating how well injured resources and damaged services meet criteria.
- Evaluate prioritization of monitoring programs in light of public opinion/perception. (Phases 1 and 2).
- Develop criteria to identify resource/service-specific monitoring activities (e.g., the life stage, behavior attribute, or population dynamic) that are likely to document the success or failure of recovery. (Phases 1 and 2).
- Obtain cost estimates for conducting specific monitoring activities. (Phase 2)
- Evaluate prioritization of monitoring activities in light of their cost-effectiveness to ascertain the quantity and quality of information to be gained versus the costs to be incurred. As necessary, reprioritize monitoring activities accordingly. (Phase 2).
- Evaluate the cost-effectiveness of recovery monitoring options. (Phase 2).

2. Need

- Establish linkages between resources in order to understand recovery.

Summary of Need

Although the tendency of monitoring is to focus on individual taxa, the *Exxon Valdez* oil spill had an impact on a large geographic area consisting of many different communities and trophic levels. By the very nature of the impacted areas, interactive and interdependent processes were disrupted, altered, or destroyed. Ecosystems are more than the sum of their parts, and the effects of perturbations such as the oil spill, can be experienced on the ecosystem level. The complexity of ecosystems, however, tends to render them difficult, if not impossible, to study as units. The study of recovery of such a relatively large association of altered communities could be not only difficult, but cost prohibitive. However, with the judicious choice of resources to be monitored, key components of the ecosystem's recovery can be addressed, and the recovery of the system as a whole can be inferred.

Objective

- Base the recovery monitoring plan on linkages between injured resources and damaged services that incorporates any knowledge of trophic levels interactions, and spatial and temporal variability.

Strategies

- Determine links, wherever possible, between resources and services by evaluating available information.

- Select resources and services for monitoring that are linked via trophic levels or that can be used to draw inferences about similar resources and services.

3. Need

- A mechanism to document recovery of injured resources and damaged services.

Summary of Need

To monitor long-term recovery, some monitoring projects should be designed in serially repeating phases. These projects could continue as long as deemed necessary to determine if recovery has occurred, providing satisfactory work was completed. Satisfactory work would be defined independently of the results obtained. Some of the resources near oil spills in cooler temperate climates show significant effects of the spills at least ten years after the event (references to be added). Provisions should be made for selective projects to continue for many years. Long-term monitoring could also occur by monitoring at periodic intervals of several years duration.

The preliminary assessment of damages has already occurred and will be used as a basis for defining recovery monitoring projects. It should be recognized that additional unsampled and presently undiagnosed long-term damage effects may be discovered, and they may need to be included in the monitoring plan at a later date. Numerous monitoring alternatives need to be examined for each project. These include, but are not necessarily limited to, timing of sampling, types of sampling, geographical area to be examined, specific parameters to sample, and logistical effort necessary to accomplish the project.

Additionally, the monitoring program should be flexible enough to alter and add projects as new data becomes available. Although monitoring of some resources will serve as indicators for a large number of other resources, those indicators may not necessarily be determinable prior to the initiation of the sampling program. Initially, many resources may need to be monitored in a given area, with the number of resources or services being reduced as data are analyzed to allow a sharper focus on fewer resources.

Objectives

- Establish a monitoring program to document the recovery of resources and services.
- Design a flexible monitoring program to accommodate rededication of efforts as new information becomes available.

Strategies

- Based on input from resource experts, establish what are acceptable rates of recovery for each resource and service, or elements thereof.
- Identify appropriate intervals (monitoring frequency) for determining recovery of a resource and service over time and space.
- Determine the influence of other perturbations (natural or anthropogenic) on recovery (e.g., winter kill).
- Utilize existing data for assessment of baseline conditions (pre-spill, and/or damage assessment and restoration control site data).
- Utilize existing data (from the spill and from other programs) for developing recovery monitoring methodologies.
- Implement a periodic review system that allows for rededication of efforts as new information becomes available.
- Involve scientific experts and resource and service specialists during development of the monitoring program.
- Develop a monitoring scope that encompasses the strategies above.

4. Need

- Knowledge that recovery is occurring, and the rate of recovery.

Summary of Need

In order for recovery monitoring to be an effective tool there must be measurable endpoints -- measures of the rate and acceptability of recovery for each monitored resource and service.

For any particular resource or service, the pre-spill or control condition will be defined as best possible by resource experts and/or existing data. The information that will be used to define the endpoint should, whenever possible, include some quantitative measure of central tendency, such as a mean, median, or mode, and some indication of variance. For some resources or services, such quantitative measures may not be available or possible to define. In these cases, the information available will be used to describe the pre-spill condition, service, or resource, and this shall serve as the indication of the condition.

It may not be feasible to monitor recovery of some resources to a level comparable to the pre-spill conditions. Some resources or services may have been on the decline prior to the spill, and some may be so severely impacted that recovery is not possible within a reasonable time period. As an example, numerous biological resources at other spills in temperate locations have been shown to remain altered as long as 10 or more years after the initial event (references to be added). Given the climatic regime of the *Exxon Valdez* spill area, some biological resources may not reach pre-spill levels for decades.

Still other resources or services may not become comparable to pre-spill conditions because of ancillary or unrelated changes resulting in an altered and non-comparable situation after the spill. For instance, a resource may not achieve pre-spill abundance and distribution if other resources have increased to fill the carrying capacity based on a common food source. The carrying capacity having been reached, the injured resource would not be able to achieve pre-spill levels.

Objectives

- Define recovery endpoints for injured resources and damaged services.

- Evaluate whether resources and services are recovering at an acceptable recovery rate, as defined for each resource/service.

Strategies

- Based on input from resource and service experts, define recovery endpoints for each injured resource and damaged service.

- Establish what constitutes acceptable rates of recovery for each resource and service based on what is known about the resource and service.

- Compare the resource or service-specific acceptable rate to the monitoring data obtained to reach a decision point: If rate of recovery is acceptable, evaluate need for continued or reduced monitoring frequency. If rate of recovery is unacceptable, evaluate restoration alternatives.

4.3 MONITORING THE EFFECTIVENESS OF RESTORATION ACTIVITIES

1. Need

- A mechanism to document effectiveness of select restoration activities at restoring resources and services.

Summary of Need

To monitor long-term effectiveness of restoration activities, some monitoring projects should be designed in serially repeating phases. These projects could continue as long as deemed necessary to determine if restoration has been effective, providing satisfactory work was completed. Satisfactory work would be defined independently of the results obtained. Some of the resources near many oil spills in cooler temperate climates show significant effects of the spills at least ten years after the event (reference). Provisions should be made for selective projects to continue for many years. Long-term monitoring could also occur by monitoring at periodic intervals of several years duration.

The preliminary assessment of damages has already occurred and will be used as a basis for defining monitoring projects. It should be recognized that additional unsampled and presently undiagnosed long-term damage effects may be discovered, and may need to be included in the monitoring plan at a later date. Numerous monitoring alternatives need to be examined for each project; these include but are not necessarily limited to, timing of sampling, types of sampling, geographical area to be examined, specific parameters to sample, and logistical effort necessary to accomplish the project.

Additionally, the evaluation of the effectiveness of restoration activities should be flexible enough to alter and add projects as new data become available. Although the monitoring of some resources and services may serve as indicators for a large number of other resources and services, those indicators may not necessarily be determinable prior to the initiation of the sampling program. Initially, many restoration projects may need to be monitored, with the number being reduced as data are analyzed allowing a sharper focus on fewer restoration programs.

<u>Objectives</u>

- Establish a monitoring program to document the effectiveness of restoration activities.

- Design a flexible monitoring program to accommodate rededication of efforts as new information becomes available.

<u>Strategies</u>

- Establish what is acceptable recovery for a resource during and/or alter restoration implementation.

- Identify appropriate intervals (monitoring frequency) for determining effectiveness of restoration.

- Determine the influence of other perturbations (natural or anthropogenic) on restoration activity.

- Utilize existing data for assessment of baseline conditions (pre-spill, and/or damage assessment and restoration control site data).

- Utilize existing data (from the spill and from other programs) for developing restoration monitoring methodologies.

- Implement a periodic review system that allows for rededication of efforts as new information becomes available.

- Involve scientific experts and resource and service specialists during development of the monitoring program.

- Develop a monitoring scope that encompasses the strategies above.

2. Need

- Knowledge that restoration activities are effective, and the resulting rate of recovery is acceptable.

Summary of Need

To determine whether restoration is effective, there must be measurable endpoints -- measures of the rate and acceptability of recovery for each monitored resource and service.

For any particular resource or service, the pre-spill or control condition will be defined as best possible by resource experts and/or existing data. The pre-spill conditions that will be used to define the endpoint should, whenever possible, include some quantitative measure of central tendency, such as a mean, median, or mode, and some indication of variance. For some resources or services, such quantitative measures will not be available or possible to define. In these cases, the information available will be used to describe the pre-spill condition of the service or resource, and this shall serve as the indication of the condition.

It may not be feasible to monitor the effectiveness of restoration of some resources and services to a level comparable to the pre-spill conditions. Some resources or services may have been on the decline prior to the spill, and some may be so severely impacted that restoration does not aid recovery within a reasonable time period.

Objectives

- Define restoration endpoints for injured resources and damaged services.

- Evaluate whether resources and services are being restored at an acceptable recovery rate, as defined for each resource and service.

Strategies

- Based on input from resource and service experts, define restoration endpoints for each injured resource and damaged service.

- Establish what constitutes acceptable rates of recovery after restoration for each resource and service based on what is known about the resources and services.

- Compare the resource-specific acceptable recovery rate to the monitoring data obtained to reach a decision point: If restoration results in an acceptable rate of recovery, evaluate the need for continued or reduced frequency monitoring. If restoration activities result in an unacceptable rate of recovery, evaluate continuing or selecting alternative restoration options.

3. Need

- Prioritization of resources and services to monitor for effectiveness of restoration and elements thereof.

Summary of Need

Given that monitoring resources are finite, a series of decisions will determine how comprehensive and integrated the monitoring program will be. The recovery or restoration of some important resources and services may not be monitorable due to the physical properties of the system, biological properties of organisms, or logistical problems in the area.

Objectives

- Develop a method for prioritizing resources/services and the monitoring activities.

Strategies

- Identify which restoration activities are questionable in aiding recovery of an injured resource/service.
- Develop selection criteria to prioritize restoration monitoring efforts.
- Establish priorities for restoration monitoring by evaluating how well injured resources/damaged services meet criteria.
- Evaluate prioritization of monitoring programs in light of public opinion/perception. (Phases 1 and 2).
- Develop criteria to identify resource/service-specific monitoring activities (e.g., the life stage, behavior attribute, or population dynamic) that are likely to document the success or failure of restoration activities. (Phases 1 and 2).
- Obtain cost estimates for conducting specific restoration monitoring activities. (Phase 2)

- Evaluate prioritization of monitoring activities in light of their cost-effectiveness to ascertain the quantity and quality of information to be gained versus the cost. As necessary, reprioritize monitoring activities accordingly. (Phase 2).

- Evaluate the cost-effectiveness of restoration monitoring options. (Phase 2).

4.4 ECOLOGICAL BASELINE MONITORING

1. Need

- Information on natural, temporal, and spatial variation of indicators to allow identification of a catastrophic event.

Summary of Need

In order to detect change that is outside the range of natural variation it is necessary to establish the bounds of natural variation. Long-term monitoring is required to define these bounds. Once established, monitoring should then be able to detect changes that extend beyond the bounds of natural variation.

Objectives

- Develop a monitoring program to detect spatial and temporal changes in biological and/or physical parameters that fall outside the range of natural variability.

- Follow long-term trends to provide baseline information for future perturbations.

Strategies

- Review past and present trend monitoring programs to identify matrices/parameters useful in detecting environmental change.

- Review past and present damage assessment and restoration data to identify resources with population effects attributable to the oil spill.

- Select physical, chemical, and/or biological indicator matrices/parameters for monitoring temporal and spatial changes in environmental quality based on the following:

  - parameters sensitive to perturbations (i.e., those that will show a change), and
  - parameters that are well understood (i.e., a solid basic knowledge of natural variation, and/or thorough knowledge of life history).

- Evaluate ease (i.e., cost-effectiveness, ability to dove-tail with other studies, frequency of sampling required) of monitoring these parameters.

- Design and implement a program that encompasses the above strategies.

2. Need

- Establish linkages between physical, biological, and/or chemical parameters.

Summary of Need

It is necessary to select indicator parameters for monitoring because it is not economically or logistically feasible to monitor all resources. Indicators should enable inferences to effects on other resources or parameters, but first one must establish the linkages between the parameters must be established.

Objective

- Enable inferences to be made concerning higher trophic level exposure/health.

Strategies

- Determine links, wherever possible, between parameters monitored by evaluating available data on interactions between physical, biological, and chemical features, including exposure mechanisms.

- Select parameters that are linked via trophic levels or that can be used to draw inferences about similar species.

- Evaluate selected parameters in relation to the geographic location and physical setting (e.g., enclosed embayment) to determine if they will be effective indicators.

5.0 RESOURCES AND SERVICES TO BE MONITORED

The settlement requires that use of restoration funds be linked to injured resources and damaged services resulting from the *Exxon Valdez* oil spill. The injuries summarized in the Restoration Framework (Exxon Valdez Oil Spill Trustees 1992) and recently completed damage assessments, along with input from the Restoration Team Work Group were used to prepare a list of injured resources and services. Injured resources are further divided into those effected at the population level (direct effects), and those indirectly effected. The injured resources and services are then to be prioritized for recovery monitoring.

5.1 RESOURCES

Resources injured at the population level are listed below in no particular order:

- Sea Otters
- Common Murre
- Pigeon Guillemot
- Black Oystercatcher
- Intertidal biota
- Harbor Seals
- Marbled Murrelet
- Harlequin Duck
- Sockeye Salmon
- Subtidal biota

Resources that were injured but did not appear to experience a population decline as a result of the spill include:

- Killer Whales
- Bald Eagle
- Dolly Varden
- Pacific Herring
- River Otter
- Cutthroat Trout
- Pink Salmon
- Rockfish

Other injured resources include:

- Archeological sites and artifacts
- Designated wilderness areas

Other resources may have been injured either directly or indirectly as a result of the oil spill but were not studied during the NRDA process. The list of injured resources may change as monitoring results become available.

5.2 SERVICES

Damaged services identified as important to monitor include:

- Commercial fishing
- Commercial tourism
- Passive uses (also called aesthetic, wilderness, intrinsic or non-use value)
- Recreation (i.e., sport fishing, sport hunting, boating, kayaking, camping)
- Subsistence

5.3 VALUE AND USE OF CRITERIA

The Trustee Council will be faced with deciding which resources and services to monitor and with choosing specific monitoring activities. How will the Trustee Council choose the resources and services to be monitored, the types of studies to be undertaken, and which of the studies to fund? Given the demands for settlement funds and the number of resources and services that could be monitored, it is important to develop a tool for evaluating the potential range of monitoring activities. A list of criteria have been developed to assist the Trustee Council in deciding which resources and services should be monitored and which studies of the monitorable resources/service will meet the goals of the monitoring plan (Section 5.4).

The criteria can be used as both a planning tool and a decision making tool. As a planning tool, the criteria can be used to the Trustee Council to:

- determine which of the injured resources and damaged services to monitor,
- develop specific requests for proposals for monitoring activities, and
- evaluate and rank proposals received in response to request for proposals to monitor specific resources and/or services.

The criteria could also be used by respondents to the request for proposal in preparing a monitoring proposal. Any proposed monitoring activity should consider each criterion in preparing a monitoring plan.

As a decision making tool, the criteria will be useful by the Trustee Council in deciding if a particular monitoring program is documenting recovery. The list of criteria should be used to evaluate the results of the monitoring activities (either on an interim basis or at the end of a monitoring element) to determine if recovery is occurring. If recovery is occurring or has occurred, the Trustee Council can make decisions to:

- continue funding the program,
- continue funding the program with reduced sampling effort and/or over a different time scale, or
- to discontinue funding.

If recovery is not occurring, the Trustee Council can use the criteria as a guide to:

- evaluate the need to invest in restoration alternatives for the resource/service,
- evaluate the need to continue recovery monitoring but with a different focus, and,
- decide if a feasibility study is necessary to determine why the resource/service is not recovering.

5.4 CRITERIA FOR SELECTING AND EVALUATING MONITORING ACTIVITIES

Criteria are proposed to assist the Trustee Council in prioritizing monitoring activities. The criteria are a series of questions that can be applied to each resource/service. Formulation of the criteria was based on verbal and written input from the Restoration Team Work Group and Restoration Team, peer reviewers, and principal investigators. An example of a decision tree developed from the criteria is illustrated in Figure 4. The question "is the resource/service monitorable?" can be further broken down, as provided below, to allow prioritization and/or weighting of those resources/services that are monitorable.

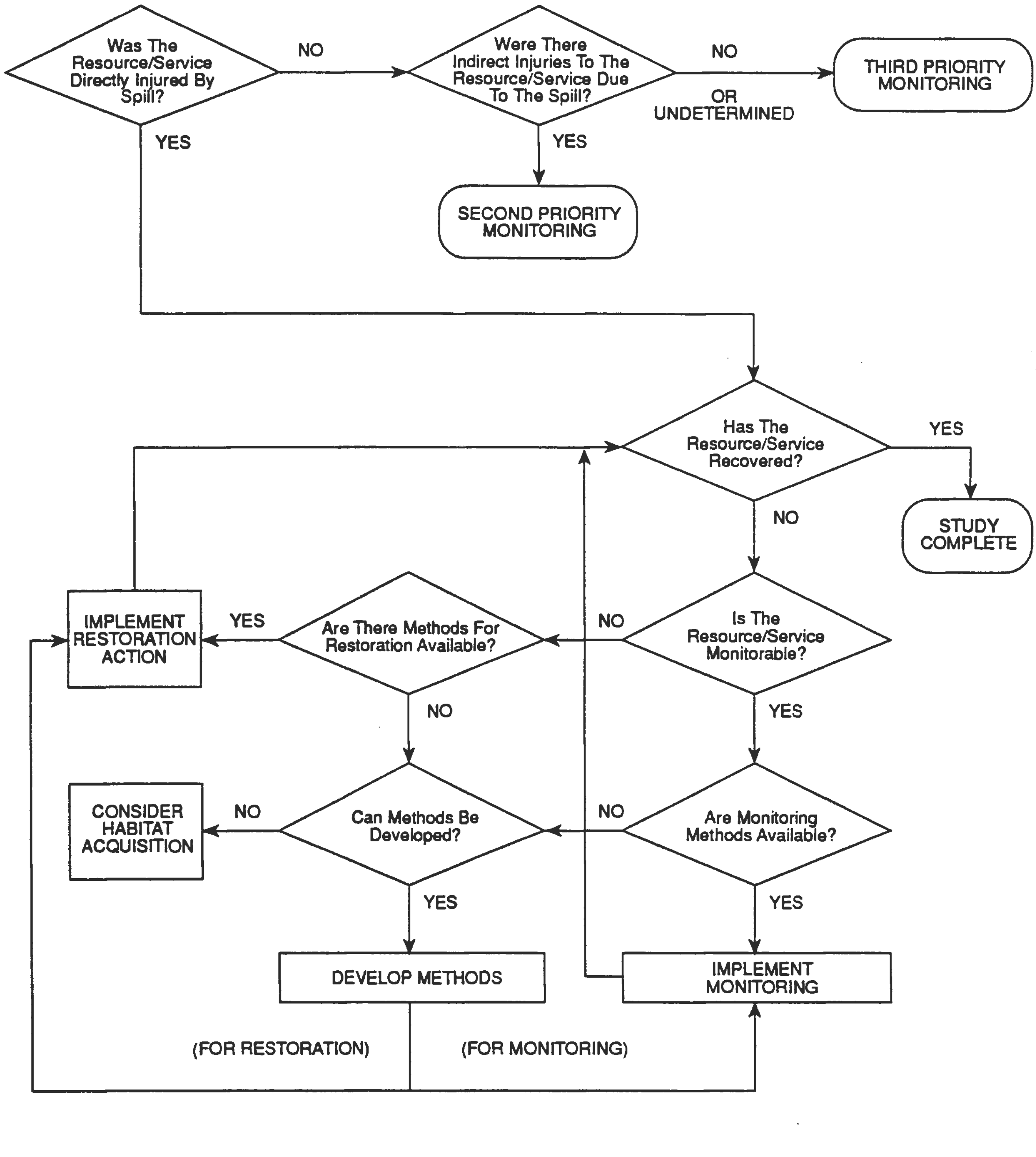

Figure 4.
Example of a Decision Tree Illustrating the Application of Selection Criteria to the Resources/Services

- Was the resource/service injured/damaged by the oil spill?
- Is the resource/service monitorable?
  - Are changes in resource/service detectable, quantifiable, and statistically testable?
    - Can confidence intervals be calculated around estimates of mean values?
    - Can recovery of resource/service be monitored using testable hypotheses?
    - Is baseline (pre-spill or control site data) available to determine spatial and temporal variation?
  - Is resource/service ecologically linked to a spill-injured resource/service?
    - Will monitoring resource/service contribute to an understanding of like resources/services?
  - Is there a recovery/restoration endpoint?
    - Will monitoring the resource/service allow a determination of the rate of recovery?
      - Are non-destructive techniques available to monitor the resource?
      - Are there regulatory restrictions to monitoring the resource/service?
      - Is the life history of the resource understood?
      - Do all life stages of a resource have to be monitored?
      - Are all life stages monitorable within the spill area?
      - Are potential sources of stress known, understood, and monitored?
      - Can the recovery monitoring provide information to aid restoration of the population?
      - Can monitoring of resource/service be integrated or coordinated with other monitoring programs?
      - Is the monitoring of a resource/service the responsibility of a resource management agency?
  - Will monitoring provide valuable baseline data for future perturbations?

Once criteria have been applied to the injured resources/services they can be applied to the specific monitoring elements of the resource/services. Figure 5 is an example of the decision tree applied at this level.

One of the top criteria (as indicated by the hierarchy and larger bullet above) can be addressed by Table 1, a matrix developed to assist in establishing linkages between resources and services.

TOPIC FOR THE WORKSHOP

Currently the criteria are presented as a list of questions that can be answered with "yes", "no", or "maybe/do not know". The majority of individuals who reviewed and commented on the proposed criteria indicated that there should be a more standardized and quantifiable basis for rating and evaluating recovery monitoring programs and activities. Each criteria could be applied to each resource/service and rated with a numerical rating (e.g., 1=low, 2=medium, and 3=high, or 1=no, 2=maybe/do not know, and 3=yes). The relative importance (rank) of a resource/service could then be assigned based on the results of the numerical rating.

For example, the criteria can first be applied to the resources and services listed in sections 5.1 and 5.2 above. Results of the numerical rating can be used to rank the resources and services into three categories (high, medium, and low). Once a prioritized list of resources and services to be monitored is developed, requests for monitoring proposals can be developed for each of the highest ranked resources and services. The Trustee Council may receive five proposals to monitor the recovery of one highly-ranked resource or service. The criteria would then be applied a second time to each of the proposals for that resource or service. The proposals would be ranked and categorized into three categories (high, medium, and low). After all proposals for each of the top-ranked resource or service are ranked, all of the highest ranked proposals would be evaluated to determine any overlap between studies, identify opportunities for coordination between studies, and to determine if there are linkages between the different proposed studies that will assist in understanding recovery through trophic linkages.

Resources and services that do not receive a high rank during the first application of the criteria will not necessarily be eliminated from consideration in the future. Similarly, proposed studies that do not fall within the highest ranking category will not automatically be eliminated from consideration for funding.

[TO BE COMPLETED]

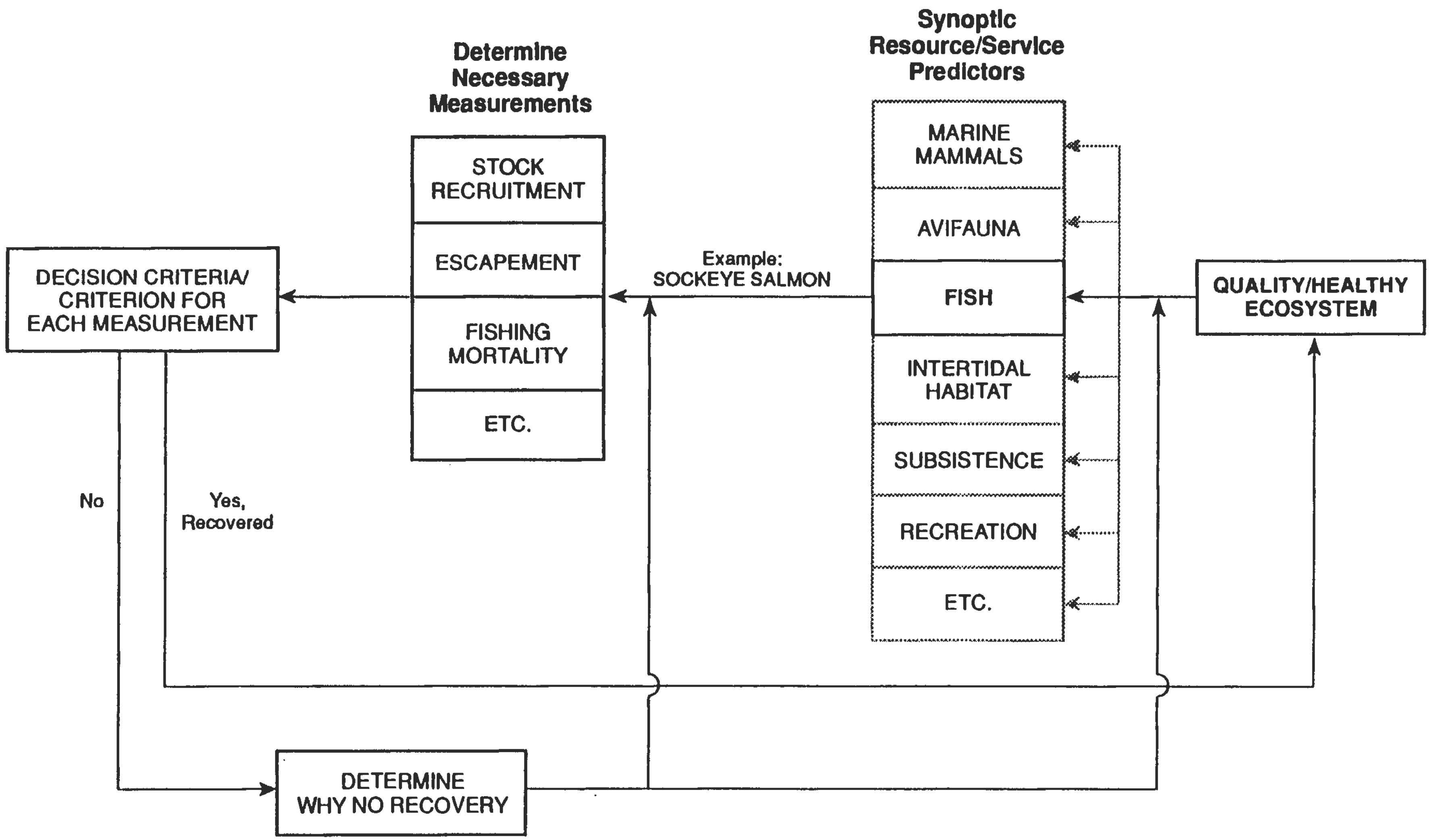

Figure 5.
Example of the Application of Selection Criteria to a Specific Resource Monitoring Element

Table 1.
Matrix Table of Linkages Between Resources

	KILLER WHALE	SEA OTTER	HARBOR SEAL	RIVER OTTER	PIGEON GUILLEMOT	BLACK OYSTERCATCHER	COMMON MURRE	MARBLED MURRELET	HARLEQUIN DUCK	CUTTHROAT TROUT	DOLLY VARDEN	SOCKEYE SALMON	PACIFIC HERRING	ROCKFISH	PINK SALMON	ARCHEOLOGICAL SITES/ARTIFACTS	INTERTIDAL COMMUNITIES	SUBTIDAL COMMUNITIES	BALD EAGLE	DESIGNATED WILDERNESS AREA	COMMERCIAL FISHING	COMERCIAL TOURISM	PASSIVE USERS	SUBSISTENCE	RECREATION	MUSSELS	FORAGE FISH
Killer Whale		●	●									●	●		●							●	●				●
Sea Otter	●																●	●				●	●	●		●	●
Harbor Seal	●											●	●		●		●						●	●			●
River Otter														●			●	●					●		●	●	●
Pigeon Guillemot													●				●	●					●				●
Black Oystercatcher																	●						●			●	
Common Murre																	●		●		●	●	●				●
Marbled Murrelet																					●		●				●
Harlequin Duck												● (Roe)			● (Roe)		●	●	●				●			●	
Cutthroat Trout																			●						●		●
Dolly Varden												●	●		●			●	●						●		●
Sockeye Salmon	●		●					●	● (Roe)		●						●	●	●		●				●		●
Pacific Herring	●		●		●						●						●	●			●			●			● (Larvae)
Rockfish				●														●			●						●
Pink Salmon	●		●						● (Roe)		●						●	●	●		●						●
Archeological Sites/Artifacts																						●	●		●		
Intertidal Habitat		●	●	●	●	●	●		●			●	●		●											●	●
Subtidal Habitat		●		●	●				●		●	●	●	●	●												●
Bald Eagle						●	●	●	●	●	●	●			●					●			●				●
Designated Wilderness Areas																			●			●	●		●		
Commercial Fishing							●	●				●	●	●	●												●
Commercial Tourism	●	●														●				●			●		●		
Passive Uses	●	●	●	●	●	●	●	●	●							●			●	●		●			●		
Subsistence		●	●										●													●	●
Recreation										●	●	●				●			●	●		●	●			●	
Mussels		●		●		●			●								●							●	●		
Forage Fish	●	●	●	●	●		●	●		●	●	●	● (Larvae)	●	●		●	●	●		●			●			

6.0 GUIDANCE ON SAMPLING DESIGN

6.1 GENERAL GUIDANCE ON SAMPLING AND STATISTICAL ANALYSES

Statistics is playing an increasingly important role in environmental monitoring and research. This has been prompted by a need for valid and repeatable evaluations of elements of environmental (physical, biological, cultural) systems with known levels of confidence and uncertainty. Statistical sample design and analytical techniques can be employed to obtain rigorous descriptions of environmental conditions.

The purpose of this section is to summarize the statistical issues relating to any monitoring program designed to evaluate recovery from the *Exxon Valdez* oil spill. It is important that all monitoring programs use comparable techniques to design the programs, collect and analyze the data, and interpret the results.

Green's Ten Principles (Green 1970) outline considerations to the design of a defensible program.

1. Be able to state concisely to someone else what question you are asking. Your results will be as coherent and as comprehensible as your initial conception of the problem.

2. Take replicate samples within each combination of time, location, and any other controlled variable. Differences among can only be demonstrated by comparison to differences within.

3. Take an equal number of randomly allocated replicate samples for each combination of controlled variables. Putting samples in "representative" or "typical" places is *not* random sampling.

4. To test whether a condition has an effect, collect samples both where the condition is present and where the condition is absent but all else is the same. An effect can only be demonstrated by comparison with a control.

5. Carry out some preliminary sampling to provide a basis for evaluation of sampling design and statistical analysis options. Those who skip this step because they do not have enough time usually end up losing time.

6. Verify that your sampling device or method is sampling the population you think you are sampling, and with equal and adequate efficiency over the entire range of sampling conditions to be encountered. Variation in efficiency of sampling from area to area biases among-area comparisons.

7. If the area to be sampled has a large-scale environmental pattern, break the area up into relatively homogeneous subareas and allocate samples to each in proportion to the size of the subarea. If it is an estimate of total abundance over the entire area that is desired, make the allocation proportional to the number of organisms in the subarea.

8. Verify that your sample unit size is appropriate to the sizes, densities, and spatial distributions of the organisms you are sampling. Then estimate the number of replicate samples required to obtain the precision you want.

9. Test your data to determine whether the error variation is homogeneous, normally distributed, and independent of the mean. If it is not, as will be the case for most field data, then (a) appropriately transform the data, (b) use a distribution-free (nonparametric) procedure, (c) use an appropriate sequential sampling design, or (d) test against simulated H_0 data.

10. Having chosen the best statistical methods to test your hypothesis, stick with the result. An unexpected or undesired result is not a valid reason for rejecting the method and hunting for a "better" one.

Although evaluating testable hypotheses is a goal for monitoring, this may not always be possible. In such cases, the methods used should be established and thoroughly documented methods.

The first step in designing a specific monitoring programs is to define the purpose of the program (e.g., to determine if the population of bald eagles is recovering in Prince William Sound after the *Exxon Valdez* oil spill). This purpose will be used to develop the rest of the monitoring program, including what specific element(s) will be monitored to meet the purpose. Statistical theory and methods, in addition to knowledge of the characteristics of and influences on the resource or service to be evaluated, are used to guide the development and execution of the following components of a monitoring program:

- formulation of testable hypotheses
- statistical sample design issues
- what to sample
- where to sample
- how to sample
- when to sample
- statistical analyses
- interpretation of results

These elements are addressed in more detail in the following subsections. Following Section 6.1, statistical considerations are discussed in terms of resources categories.

6.1.1 Formulation of Test Hypotheses

Based on the purpose defined for a specific monitoring program, a statement (the null hypothesis, H_0) is formulated that addresses the purpose in simple, concrete terms. This null hypothesis identifies the state of the element that is to be tested (e.g., if the purpose of a monitoring program is to determine if the population of bald eagles is recovering in Prince William Sound after the *Exxon Valdez* oil spill, the null hypothesis could be stated as "There is no statistically significant change in the number of bald eagles residing in Prince William Sound during the breeding season between 1993 and 1994.") Testing this null hypothesis via statistical methods will be the objective of the sampling and analysis process.

It is possible that the data will indicate that the null hypothesis is not likely true. An alternative statement (alternative hypothesis, H_A) is formulated that defines a different state of the resource or service. Should a statistical test indicate that the null hypothesis is false, the data can be evaluated in terms of the alternative hypothesis. For example, the alternative hypothesis for the null hypothesis in the previous paragraph could be stated as "There is a statistically significant increase in the number of bald eagles residing in Prince William Sound during the breeding season from 1993 to 1994".

When testing a hypothesis, two types of error exist (Figure 6). Type I error (α), the probability of rejecting the null hypothesis when it is actually true, is commonly set at 5%. Type I error is also known as the significance level of a test. Type II error (β) is the probability of accepting the null hypothesis when it is actually false. Decreasing one of these two errors will increase the other.

		H_0 Accept	H_0 Reject
H_0	True	Correct Decision (1-α)	Type I Error (α)
	False	Type II Error (β)	Correct Decision (1-β)

Figure 6. Possible outcomes of hypothesis testing.

6.1.2 Statistical Sample Design Issues

In order to develop an optimal sampling design for a monitoring program that tests the specified null hypothesis, some statistical issues must be addressed, including the significance level (α), power level (β), sources and magnitudes of variation, and minimum detectable change (MDC).

As noted in the previous section, two types of error are present in hypothesis testing, Type I (α) and Type II (β). These errors need to be balanced, since decreasing one increases the other. The only way to reduce one error level without increasing the other is to improve the sampling design, (e.g., increasing sample size). A sampling design must adequately and realistically address both types of error.

In environmental monitoring and sampling, many sources of variation exist in addition to those commonly addressed in experimental designs (e.g., within-sample, between-sample, analytical, and random). These additional sources also exist at very different magnitudes and dimensions. An optimal sampling design must also address temporal variation, spatial variation, and natural system variations. Replication is one sampling technique that can be used to quantify many of these sources of variation. Data from previous studies may also be useful in evaluating potential sources of variation.

When testing a hypothesis, a level of change exists below which the null hypothesis is not rejected. This minimum detectable difference (MDC) of a statistical test is affected by several other test parameters:

- inherent variation (natural variation, within- and among-sample variation, and analytical variation)
- sample size (n)
- significance level (α)
- power ($1-\beta$)
- temporal (r) and spatial (ρ) autocorrelation

The minimum detectable difference should be small enough to meet the needs of the monitoring program but not so small as to require a prohibitively large sample size or reduce the power of the test below an acceptable level. Due to the amount of variability usually found in environmental data and limited sampling budgets, a balance between the MDC, sample size, significance level, and power has to be reached.

The following function can be used to study the balance between these quantities or evaluate the level of power associated with statistical tests under consideration for a single hypothesis.

$$MDC = (Z_{\alpha/2}+Z_{\beta})*s*[2(1-r)/n]^{1/2}*[1+\rho(n-1)]^{1/2}, \text{ where}$$

$Z_{\alpha/2}$ and Z_{β} are the normal "Z" values for various levels of α and β;
s is a quantity that estimates the inherent variation (commonly a standard deviation);
r is the temporal autocorrelation, or a quantity that estimates it;
n is the sample size; and
ρ is the spatial autocorrelation, or a quantity that estimates it.

A preliminary sampling effort should be made, if possible, to evaluate the design, evaluate the sampling and analytical procedures, and identify and quantify sources of variation. If the sampling design does not require extensive modifications, the preliminary data could be included in the monitoring program analyses. Another approach to address the variation issues would be to over-sample and use extensive replication the first two to three years to ensure adequate sample sizes and obtain estimates of variation components and then reduce the sample scope for the remainder of the monitoring program.

Any sampling design that is developed for a monitoring program should be flexible. It is quite possible that changes will have to be made after the first or second year to address inadequacies in the design or possibly budget constraints, especially if the amounts and primary sources of variation cannot be adequately assessed during the design phase.

6.1.3 What to Sample

After defining the purpose of a monitoring program, the resource or service to be sampled is decided so that the null and alternative hypotheses can be formulated. The state of the resource or service that is being tested should be relevant to the defined purpose of the monitoring program. There will probably be more than one characteristic of the resource or service that can be measured to test the hypotheses (e.g., the total number of bald eagles or the number of juvenile bald eagles). There should be little or no difficulty collecting data on the resource or service, and the data should have the capacity to evaluate the null hypothesis via statistical testing.

Choosing a characteristic of a resource or service that has been sampled in previous studies or monitoring programs may be advantageous. Data from these previous sampling efforts can be used to extrapolate properties associated with the measured characteristic prior to designing the sampling effort. Variability components could be estimated from the previous data to determine adequate sample size. Sampling and analysis problems encountered in prior work could be avoided or accounted for in the current study.

6.1.4 Where to Sample

Where to sample encompasses two issues, the study area and actual sample locations. The study area should encompass the entire area of interest for the monitoring program, and the sample locations will be sited within this study area. Previous studies can provide insight into appropriate methods and possible pitfalls.

The process of choosing the actual sample locations has implications to the statistical tests and their interpretation. How the sample locations are chosen influences the relationship between sample locations, variability estimates, and the inference basis for the statistical tests. Conventional statistical analysis methods were developed for data collected as random samples. Random, or probability, samples are considered independent and representative of the population from which they are sampled, and estimates of parameters such as means and variances computed from such samples are unbiased for those populations. By removing the randomness from the sample locations, as in judgment sampling, bias can influence the parameter estimates and restrict the interpretation of statistical tests.

There are three main sampling approaches that generate random samples: random, stratified random, and systematic random. In the random approach, samples are randomly located within the entire study area. In the stratified random approach, if the population of the resource or service under study is known or suspected to be unevenly distributed within the study area, homogeneous subgroups can be formed within the study area and random samples taken from each subgroup. The systematic random approach makes use of a two-dimensional grid that is randomly placed in the study area, and the sample locations are taken as either the intersections of the grid lines or at the same location in each grid area (e.g., the center). Other sampling schemes that produce random samples have been or can be developed from these basic approaches.

6.1.5 How to Sample

The methods used to collect data in the various monitoring programs should follow standardized protocols. These standardized protocols will ensure the data is consistent, accurate, and comparable between sampling events, monitoring years, and monitoring studies. Standardized protocols are also important to ensure that the data that will be analyzed is of acceptable quality for statistical analysis.

It is likely that many of the standardized protocols exist as a result of previous environmental studies. These will simply need to be assembled into a cohesive set. However, others may need to be developed from scratch, but previous research may provide useful insights into possible methods and difficulties.

Sampling methods should be documented in detail, specifying the exact steps to be taken from locating sample sites to shipping the collected samples to the analytical laboratory.

- locating sample sites
- collection of field observations (e.g., temperature)
- collection of sample(s)
- preparation of field spikes, duplicates
- preserving, packaging, labeling of samples
- storage, transportation of samples
- documentation of samples (e.g., chain-of-custody forms)

Laboratory analytical methods should also be documented in detail; however, some of these will incorporate state and/or federal protocols.

- receipt of samples from the field
- preparation of samples
- preparation of laboratory spikes, blanks
- analytical procedures
- reporting formats, including units and qualifiers
- documentation of samples (e.g., chain-of-custody forms)

Compositing samples can be used to reduce the costs of analyzing large numbers of samples, increase the amount of sample material available for analysis, and reduce the between-sample variability caused by heterogeneous sample material. However, the consequences of compositing include the loss of ability to estimate between-sample variability, and this would need to be addressed in the sampling and analysis design.

6.1.6 When to Sample

Deciding when to sample is influenced by many factors. Natural factors, such as weather conditions and time of the year, influence the variability of the data that is collected. Organizational factors, such as sampling and analytical costs, can limit the amount and frequency of sample collection.

Historical data, if available, and knowledge of resource or service characteristics can be very useful in determining the best time(s) for sampling. Sampling times should be consistent from year to year (e.g., such as counting bald eagles present during the breeding season only). The severity of the weather in the oil spill area and constraints of monitoring (e.g., seasonal migrations of certain species) may also restrict sampling times.

6.1.7 Statistical Analyses

There are many statistical analysis methods available for evaluating environmental conditions. The primary focus of monitoring programs is to detect change over time, and several standard analysis methods can be used to this end, such as analysis of variance (ANOVA), trend analysis, and time series analysis. Regression, correlation analysis, and other multivariate techniques can be used to evaluate hypothesized relationships between different variables measured in the monitoring program. Unless special circumstances require, standard analysis methods such as these should be used for the sake of clarity, comparability, and repeatability.

The analysis method used to test the null hypothesis is chosen prior to sampling, and it should be appropriate and rigorous for the stated null hypothesis and sampling methods used. Typically, the significance level (α) for a hypothesis test is set at 5%. For the chosen analysis method, the assumptions associated with the method must be addressed, since violations of an assumption can compromise the validity of and confidence in the analysis results.

Since spatial variability will most likely influence all monitoring data collected to some degree, statistical spatial analysis techniques may need to be considered. Any spatial methods used should be thoroughly researched and carefully applied. Geographic information systems (GIS) may be useful in such analyses.

6.1.8 Interpretation of Results

Interpretation of results of analysis must take into consideration the tested hypotheses, sampling methods, and analysis methods and associated assumptions. The conclusions reached and interpretations made must be supported by the data and take into consideration the sampling methods used and the assumptions and restrictions associated with the analysis methods.

When drawing conclusions regarding environmental data, caution must be used. While a significant change may have been detected, can it be attributed to a recovery process or is it a result of some natural event (e.g., a decrease in predator population)? Because so many factors are not measured, conclusions regarding relationships between elements should be viewed as associations and not necessarily cause-and-effect relationships. Establishing cause-and-effect relationships in the environment requires controlling all factors not measured.

6.2 GENERAL GUIDANCE ON SAMPLING RESOURCES

Below are general guidelines for monitoring specific resource categories. Two general categories are specified: (1) avifauna and mammals, and (2) intertidal and subtidal communities.

6.2.1 Avifauna and Mammals

Throughout this general discussion of what, where, and how to measure, birds have been used to illustrate the point.

6.2.1.1 What to Measure

Some preliminary work is needed in order to determine what to monitor/measure. A primary consideration is to determine what questions need to be asked with regard to bird or mammal populations in the oil spill area. Development of these questions should rely heavily on previous work. If possible, new monitoring programs could be designed to be compatible with previous data collection such that meaningful comparisons are possible. When the appropriate questions (clear and unambiguous) have been asked, a monitoring program must be developed that will provide answers to these questions.

The monitoring program must be specific to the questions. For example, general questions on avian ecology may require censusing of large numbers of bird species and communities. For questions related to a specific food resource or foraging technique, a single representative species may be studied. For questions on reproductive success, surveys of breeding colonies and fledgling rates would be important. Toxicological questions requiring specimen analysis will have additional constraints on the selection of study species. Some species, such as bald eagles or marbled murrelets, may be important for study because of their status or public interest.

Further guidance on what to measure is available through use of the criteria presented in Section 5.0. The list of criteria should be reviewed to verify that they are appropriate for selection of suitable species. The criteria should also be relevant to the questions being asked. Applying these criteria to the list of species injured should lead to the selection of appropriate study species.

6.2.1.2 Where to Measure

Deciding where to measure a particular resource also depends on the questions that have been asked. If the objective is a long-term comparison with an existing pre-spill data set, it would be important to monitor in the same location and in the same manner as the previous work. If the question is how current conditions in the spill area compare with undisturbed areas, there would obviously be a set of parallel locations that differ only in their exposure to oil.

Where to measure would also depend on the species selected for study. For instance, some bird species can be studied most effectively in their breeding areas, while others are more easily studied in foraging areas. For some seabird species that nest far from the marine environment, such as marbled murrelet or harlequin duck, parallel studies might be needed in nesting and foraging or wintering areas.

When species and general locations have been selected, specific sampling sites should be chosen with a randomized procedure to ensure the statistical validity of the results. Likewise, an adequate number of replicate sites should be sampled to accurately estimate the means and variances of the variables to be studied. Simple random sampling could be appropriate for species that are known to be distributed homogeneously throughout the study area. For example, simple random sampling might be used for bald eagle nesting territories along the coastline or for area-wide censusing of foraging seabirds. Stratified random sampling would be more suitable for species with heterogeneous overall distributions that can be separated into homogeneous subgroups. For example, many seabirds nest in large colonies, each of which could be considered a homogeneous subgroup of the whole population. Specific colonies could be selected based on prior information, but sampling within a colony should be randomized.

6.2.1.3 How to Measure

Many measurement techniques are available, and the selection again depends on the questions being asked and the species being studied. If the objective is to duplicate previous data collection, identical techniques should be used, if possible.

Aerial surveys may be appropriate for broad-scale censusing, for instance of seabird concentrations, but these would not be suitable for species that are small, difficult to identify, or with few, widely scattered populations. Aerial surveys could provide very accurate estimates of bald eagle breeding territories and nesting success. For some species, different sampling methods would be necessary for different aspects of their life cycle. For example, boat surveys are needed to census foraging marbled murrelets, but ground surveys are necessary to locate possible nesting areas. Populations of some seabird colonies might be accurately determined from boat surveys, but in other colonies, sampling from the ground might provide better results.

Demographic studies of populations could require trapping and marking many individuals with distinct color bands. For studies on home ranges, it might be necessary to use radio transmitters on individuals. In small home ranges, a portable radio receiver on the ground would be most useful. In large home ranges or for long distance movements, it might be necessary to use helicopter or airplane-mounted receivers.

For physiological studies or for toxicological analyses, it would also be necessary to capture individuals. Some tests might be possible with samples collected in the field from animals that could be released. For example, blood and urine samples can be quickly and easily collected with no harm to a bird. Other tests, such as trace element analysis of organ tissues or electron microscopy of subcellular structures, would require the sacrifice of some individuals. Special review and approval should be required for studies of this sort.

6.2.1.4 When to Measure

Timing of sampling will often follow directly from the primary questions being asked and the species selected for sampling. Surveys of seabird breeding colonies would obviously have to be conducted during the breeding season. However, the breeding seasons are unlikely to overlap completely for all possible species of interest. Thus, prior information must be used to choose the optimum time for sampling.

Multiple sampling periods during a breeding season may be appropriate for some species. For example, a preliminary aerial survey could locate active bald eagle nests, and a survey later in the breeding season could determine the success rate of active nests. For other species, separate surveys in separate locations may be necessary for breeding and wintering populations. For example, harlequin ducks will breed on interior rivers and spend the winter in the near-shore marine environment.

The sampling season would also be important for some toxicological or physiological studies. Studies on hormonal changes related to breeding would have to be conducted over a time period spanning the breeding season. A study on trace element concentration in fat deposits would require sampling when body fat would be at a maximum.

6.2.1.5 Data Organization

As is the case with other aspects of a sampling design, the data organization is also dependent on the questions being asked. Data organization will also depend on the analytical procedures to be used, and the analysis should be considered in the initial phases of sampling design. Data organization must be carefully planned to be compatible with prior data sets and with the overall restoration database. However, it is possible that trying to make a data set conform to a general database might make it very difficult to use those data to answer the specific questions of a specific study. In those cases it would be most efficient to organize the data in a manner that will provide the most usable results for the specific study.

In general, it is anticipated that data would be organized in a matrix format. The simplest format would be a two-dimensional matrix with columns representing independent and dependent variables and rows representing individual measurements. Typical independent variables would be time, location or condition descriptors, and environmental factors. The dependent variables could be any factor that could be measured and would contribute to answering the initial questions. A three-dimensional matrix might be appropriate if similar, simultaneous studies are to be conducted on several species. Computational techniques allow the use of multi-dimensional data matrices, if that degree of complexity is appropriate to the initial questions.

6.2.1.6 How to Analyze

Data analysis is a key element of a sampling program, and it is essential to consider analysis when designing field data collection procedures. The analytical procedure must be focussed toward answering the basic questions of the study, and the data collection must be appropriate for the analytical procedures. If the proper data were not collected initially, the best analytical techniques will never produce meaningful results.

The analysis should consider the nature of the data in determining the appropriate statistical methods. Nominal variables are purely qualitative and cannot be assigned numerical values, and thus they may only be suitable for signs-based nonparametric or categorical statistical methods. Ordinal, or ranked, variables can be assigned numerical values, but the differences among ranks are not necessarily proportional. These variables must be analyzed with non-parametric statistics. Many ecological measurements will be discrete variables, which can have only integer values. Examples of discrete variables would be the number of bald eagle nests on an island or the number of seabirds on a transect. Discrete variables may be treated with parametric statistics, provided they satisfy the assumptions of those methods or can be transformed via a monotonic mathematical function to do so; categorical test methods may also be appropriate. Other environmental measurements are continuous, with no limits on possible values. Many physiological or toxicological measurements, such as metabolic rates or chemical concentrations, are continuous and can be analyzed with parametric statistics.

Depending on the statistical methods used for analysis, appropriate numbers of replicate samples must be taken to obtain accurate estimates of means and variances. If it is not possible to collect enough replicates, it may be necessary to use non-parametric statistics. However, without sufficient replication, statistical power may be compromised. Data available from previous studies can be used to evaluate variability and determine an acceptable level of replication.

A number of basic assumptions must be satisfied in order to use parametric statistics. It is essential that samples be taken at random. Errors in measurements must be independent and normally distributed, and variances of samples must be equal. Most parametric statistics assume normally distributed data; however, they are usually quite robust in the presence of non-normal data and/or heterogeneous variances. Unequal sample sizes and non-independent samples cause the most trouble. Statistical tests of these assumptions should be made to ensure the validity of the results. If the data do not fit a normal distribution, they may be normalized by a transformation. For example, most data on species populations are not normally distributed, but they can be normalized with a log transformation. If a suitable transformation for the data cannot be found, non-parametric statistics should again be used.

The particular statistical procedure to be used will obviously depend on the nature of the questions being studied. If the objective is to make comparisons between areas that were directly affected by the oil spill with other areas that were not affected, then a *t*-test or analysis of variance might be appropriate. To demonstrate functional relationships simple or multiple regression is possible. Trend analysis could be particularly useful in demonstrating recovery of a resource over time.

6.2.1.7 How to Interpret

If the objectives of the monitoring program were clearly stated, specific questions were addressed, and a comprehensive monitoring procedure was implemented, then interpretation should be straightforward. The results of the analyses should directly answer the questions that were asked, and reasonable conclusions should be drawn from the results.

Interpretation of the results of any particular study must be firmly based on reliable data that have been analyzed by statistically valid and relevant procedures. Any interpretation is only as rigorous as the weakest element in the entire data collection and analysis sequence. Care should be taken to avoid extrapolating to any interpretation beyond that which is justified by the available evidence and analysis.

6.2.2 Intertidal/Subtidal

The marine benthic environment exists at the interface of the bottom sediment and the overlying water column. The organisms found in this habitat consist of those that live exclusively in the sediment, those that live exclusively in the water, and those that can make the transition between the water and the sediment. Furthermore, these organisms can range in size from the minute to large. It is important to remember that all of these types of organisms constitute the entire biological assemblage found in a given area.

In sampling marine benthos, it is often logistically impossible or scientifically undesirable to sample for one specific taxon. Rather, the emphasis is on examining the component of the whole assemblage, community, or ecosystem. The sampling and analytical methodology described here is consistent with sampling protocols in use elsewhere (Tetra Tech 1987). Intertidal and subtidal monitoring programs should concentrate on the numerically dominant or ecologically important taxa and be able to sample them so that predictions and analyses have sufficient statistical power to be meaningful. Nevertheless, it is recognized that indicator or key taxa may not be abundant and that the sampling effort may be altered to address testable hypotheses concerning these rarer taxa.

Preliminary sampling/studies addressing the variations seen in the study area are desirable. These preliminary sampling periods often expose flaws in the sampling or analytical design that are easy to correct prior to the actual study implementation, but which can be difficult to change after the project is fully geared up and functional. Support should be allotted for such preliminary or exploratory sampling.

The optimal sampling design is dependent upon which aspect of the benthos is ng studied, and upon the habitat being examined. It needs to be emphasized that these components of the benthos are defined by the sampling methods, not by function, ecological interactions, or taxon. Similar sampling methods utilized in different habitats will sample different taxa. As an example, the infauna of embayments open to the ocean will likely be very different from nearby areas of comparable sediments in fjords with a sill across the entrance (Shimek 1990). Regardless of the component of the benthos being examined, after the data are collected,much of the analysis is similar.

No area to be sampled for the benthic assemblages can be assumed to be either spatially or bathymetrically homogeneous. Because of this potential variation, sampling should be at defined stations in the area. Based on preliminary analyses, the data from these stations may be shown to be statistically indistinguishable from station to station. If that is the case, those data may be pooled for subsequent analyses.

The data collected from such sampling would be analyzed with expressed intent of defining and describing the populations of the numerically dominant taxa. The total number and abundance of all collected taxa would be determined, of course; however, the abundant and, presumably, important or target taxa would be the focus of the analyses. Interpretation of the variations seen in these taxa will vary from project to project, depending upon the project design. Nonetheless, the proximate questions generally will involve addressing temporal or spatial changes in population abundances of either particular taxa or groups of taxa.

6.2.2.1 What to Measure

Generally the target taxa will be determined by the questions being asked in the particular monitoring plan. Generally, however, the investigator should concentrate on the numerically dominant and ecologically important taxa. These taxa will need to be sampled so that predictions and analyses have sufficient statistical power to be meaningful. Several descriptive and derived quantitative ecological indices can be used to describe diversity and dominance of the faunal array from each station.

Descriptive Measures

For any station, these population descriptive measures will be calculated from the unmodified species-abundance data pooled over all station replicates. The descriptive measures follow:

Numerically Dominant Taxa

For each station, the abundance data will be pooled and the taxa ranked in descending abundances until those taxa which constitute an arbitrary percent (generally 75 percent or 90 percent) of the cumulative sampled benthic faunal abundance are listed. This will result in a relatively small list of taxa. This list, defined as the numerically dominant taxa, generally includes most of the ecologically important taxa.

Number of Numerically Dominant Taxa

This is the number of taxa present in the array of the numerically dominant taxa, as defined above.

Number of the Non-Numerically Dominant Taxa

This is the number of taxa present in the array of the numerically dominant taxa subtracted from the total number of taxa.

Total Number of Taxa

This is the total number of taxa sampled for any station and is the simplest measure of species richness.

Total Numerical Abundance at a Station

This is the number of individuals of all taxa, summed over all taxa.

The Mean Taxon Abundance, ± 1 Sample Standard Deviation

This is the most common measure of the population abundance central tendency coupled with the most common estimate of sample variance. Other common measures such as median abundance or standard error may also be used.

Other descriptive measures, such as the number of any particular feeding group, or the relative abundances of pollution-intolerant taxa may also be useful in specific situations.

Derived Quantitative Indices

To be formally correct, all of the following indices should be calculated using individuals identified to the same level of taxonomic precision. These indices should be calculated only for those taxa identified to species.

The Shannon-Wiener Diversity Index (H')

The Shannon-Wiener index is derived from the mathematical discipline of "Information Theory". It ranges upward from zero, and gives a quantitative measure of the relative amount of new information contained in each individual specimen collected. Where the sample is dominated by a few taxa, or has few individuals, the amount of new information likely to be gained by enumerating any given specimen is small. Where the sample is diverse or large, new information is more likely to be gained because each new specimen might be a representative of a previously unsampled taxon.

The following is the formula for the Shannon-Wiener Index:

$$H' = -\sum_{i=1}^{s} p_i \ln p_i$$ (1) The Shannon-Wiener Index

Where: H' = The Shannon-Wiener or information theory index of diversity calculated with natural logarithms ($\log_e = \ln$);
p_i = The proportion of taxon "i" in the sample; and
s = The number of all taxa in the sample.

The Evenness Index (J)

The Evenness Index (J) ranges from 0 to 1 and measures the dispersion of all taxa in a sample as a proportion to the maximum possible dispersion. Samples with J values near zero are dominated by few species, while samples with values near one have approximately equal numbers of individuals in each species.

The following is the formula for the Evenness (J) Index:

$$J = \frac{H'}{\ln s}$$ (4) The Evenness Index

Where: J = The Evenness Index.
H' = The Shannon-Wiener Index.
s = The total number of taxa in the sample.

Simpson's Dominance Index (C)

Simpson's Index (C) ranges from 0 to 1, and measures the probability of randomly drawing two individuals of any given species from the total sample. This index is a measure of the degree the sample is numerically dominated by one or a few taxa. Values near 0 indicate a diverse array (the probability of drawing two individuals of the same taxon is small), while those near 1 reflect dominance by one taxon (the probability of drawing two individuals of the same species is large). The formulation for the Simpson's Index given here is the standard unmodified form (Poole 1974).

$$C = \sum_{i=1}^{s} \frac{n_i(n_i - 1)}{N(N-1)}$$ (3) Simpson's Index

Where: n_i = the number of individuals in species "i", and
N = the total number of individuals in the sample.

Standing Crop

The standing crop should be estimated. Perhaps the easiest method is by the determination of wet-weight biomass. The wet-weight biomass for major organism groups (taxa) from each sample will be determined by pouring the animals from the sorted sample through a preweighed 0.25-mm screen. This screen will be placed on absorbent paper and either blotted dry from underneath, or allowed to remain on the paper until no more fluid is absorbed. The sample will be dried for no more than 30 seconds because small animals, polychaetes in particular, quickly dry out and rupture or tear. The screen and sample will be weighed. Finally, the animals will be washed back into the vial with 70 percent ethanol, and the wet-weight biomass determined by subtracting the container weight.

6.2.2.2 Where to Measure

Once the questions that address the benthos have been asked, the decision of where to measure will become obvious. If the questions involve comparisons of data collected over long-term sampling periods, then sampling should be in the same location as previous work. If the questions involve comparisons between current conditions in a spill and reference areas, then paired locations need to be chosen.

6.2.2.3 How to Measure

Assemblages are often measured to discern changes from either the assemblages present at a reference or control area, or differences between the abundances in a sampled area and some pre-defined level of abundance that indicates recovery or restoration. In assessing assemblages of organisms, two measurable factors define many of the observed variations. These parameters are:

- the diversity of the various taxa, and
- the abundance of those taxa.

It must be recognized that changes in these measured factors are themselves due to changes in other more important, but often unknown, parameters that determine the survival and growth of the separate individuals that constitute the populations of each individual taxon.

These ultimate factors may or may not be measurable, but without well-defined experimental projects, their effects will remain uncertain, and will be defined primarily by correlative techniques. All investigators, regulators, and interested readers should be well aware of, and be frequently reminded of, the statistical dictum: "Correlation does not imply causation." The examination of marine benthos by the descriptive-correlative approach, also referred to as the mensurative approach, depends upon the accumulation of a body of observations to support or reject appropriately constructed hypotheses.

These hypotheses must be clearly and precisely phrased to have any validity, and they must be tested with data collected in a manner that insures that confounding hidden factors are minimized. The data need be gathered, analyzed, and interpreted with sufficient awareness of the limitations of the sampling, statistical, and analytical procedures. Clearly, investigators must be cautious about interpreting correlative data without experimental confirmation. However, data collection and sampling plans also must reflect the potential limitations of the procedures involved. Many of those limitations can be addressed by the use of appropriate sampling and statistical methodology.

There are two underlying assumptions to sampling that should be explicitly stated and addressed.

- The first assumption is that the sampling device is physically adequate to sample the populations in question with a minimum of bias. This assumption can be addressed by with a precise definition of the sampled assemblage. Once the assemblage has been defined, an adequate sampling device can be chosen.

- The second assumption is that the sampling plan provides sufficient samples to represent the assemblage. This assumption has to be operationally tested by calculating three related factors:
  - the species-area relationships in the study area,
  - the statistical power of the data resulting from the sampling plan, and
  - the adequacy of the replication.

6.2.2.4 When to Measure

The period in which sampling will take place will follow directly from the primary questions being asked, although it may be modified by logistical and seasonal concerns. Annual monitoring will probably suffice for long-term monitoring programs. Nevertheless, seasonal, monthly, or even more frequent sampling periods may be necessary, for example, when questions of reproductive fitness are addressed, the need may exist for sampling gonadal indices over a longer period.

No particular season is likely to provide better data on the organisms than any other. Most Northeastern Pacific benthic infauna show relatively long periods of recruitment and spawning (Strathmann 1987). Consistency from year-to-year, however, will be important in establishing trends.

6.2.2.5 Data Organization

The organization of the sampling and data is dependent upon the questions being asked, the sampling methodology, and the analytical methods. In general, benthic infaunal data consists of a series of station by taxon by abundance matrices. Other data, such as sediment parameters may be also be included in these matrices. Several types of data management systems involving benthic infaunal data presently exist, and most are relatively readily interchangeable.

In general, in addition to the data collected by the sampling effort, some additional data coding will be necessary for the data to be interchangeable between sampling plans or monitoring programs. The most likely candidate code is the National Oceanic Data Center (NODC) coding for individual taxa. In general, two dimensional matrices of variables and observations can be easily manipulated by most spreadsheet software to conform broadly with any other data set, albeit no two monitoring projects will likely collect identical sets of data.

Little attention should be given to requiring the collection of extraneous data. If the data being collected are not pertinent to the project at hand their collection and manipulation can add significantly to the project costs, with no immediate return. This will lead directly to the poor data collection and management and foster questions about the reliability of the required data.

6.2.2.6 How to Analyze

Benthic data are generally voluminous and complex. Analytical methods will vary with the questions being asked. The constraints, assumptions, and inherent strengths and weaknesses of the various analytical methods should be carefully weighed before they are used. The number of valid analytical techniques is too large to simply address here. Individual researchers must be certain of the applicability and adequacy of the techniques they propose.

The utility of inexpensive statistical software programs for microcomputers has put a significant amount of computational power literally at the fingertips of investigators.

Unfortunately, many of the statistical subroutines are used without an adequate understanding of the properties of the given statistical test. While is recognized that innovative methods often result in particularly insightful conclusions, investigators are encouraged to be conservative in their utilization of statistical techniques wherever possible. Analysis of variance and t-tests are often the best choices for analyses using parametric statistics, while Friedman Rank-Sum tests and Mann-Whitney U tests would be the corresponding choices as non-parametric tests.

Investigators should make as few assumptions as possible about their data. Particularly, they should not assume that their data are normally distributed; the assumptions inherent in normality can and should be tested. Many valid data transformations exist to allow the data to be analyzed with parametric tests, if those tests are sufficiently more powerful to be desirable in any given situation.

6.2.2.7 How to Interpret

Reference stations are chosen for two reasons. First, to provide indications of overall basin or bay wide changes in the fauna. Numerous bay wide or basin wide changes have been documented in large areas such as Puget Sound in the last twenty years (Nichols 1988), and the occurrence of such changes should be considered in the analyses of the recovery or restoration. Second, reference areas are often used as a benchmark to assess normality of a study area. The study stations and the reference stations are statistically compared, and the results of those comparisons are used in assessing whether or not the study area is "normal."

It should be recognized that it may be difficult or impossible to find true reference stations that are adequate for comparison to the *Exxon Valdez* restoration and recovery stations. Reference stations need to be chosen on the basis of sediment and hydrographic parameters to reflect "normal" or unstressed environments similar to the study areas. Previous work has indicated that shallow water unconsolidated reference areas may be difficult to find. This is primarily due to two factors:

- The hydrographic conditions of any given area are unique and result in a fauna that may not be particularly similar to other sites.

- The assumption must be made that much of the shallow-water benthic habitats in the spill area have been altered, to a greater or lesser degree. Few sites can be found that can be considered *a priori* to be undamaged and thus suitable for providing appropriate monitoring reference stations. Any site must be sampled and analyzed prior to its designation as a reference station. Reference sites are often chosen on the basis of a close match of physical factors and a subjective judgement of normality; the biota are assumed, therefore, to be normal for such an area. Such an assumption without sampling and validation runs the risk of becoming circular: reference areas are chosen because they are assumed to be normal, and then normality of these sites is presumed because they are designated as reference areas.

The first factor precludes use of distant reference stations. The hydrographic conditions in distant areas are likely significantly different from the monitoring stations, and the potential organism groups that may be found might be significantly different due the availability of recruits or unknown physical effects. The second factor means that no or few nearby sites will likely be useful as strict reference area as well.

Rather than try to find a strict reference area for each of the habitats to be sampled, nearby stations may be chosen to provide "background" information about the basic trends in the abundance and composition of the benthos. The background stations will be used to provide an indication of bay or basin-wide changes in the fauna. These background stations will be in located in habitats similar to those being monitored.

Attempts must be made to match the descriptive sediment parameters at the background stations to those of the monitoring stations; however, the background stations may not be chosen specifically to provide the closest possible sediment match to the monitoring stations. These background stations should approximate as many of various monitoring stations' physical parameters as is possible. Although the background stations may not precisely match any particular monitoring station, they should provide a general benchmark in the event of geographically widespread faunal changes.

6.3 GENERAL GUIDANCE ON SAMPLING SERVICES

Several services provided to the public were also damaged by the *Exxon Valdez* oil spill and, in conjunction with resources, should be considered for monitoring. Although services appropriate for recovery monitoring are identified in Section 5.0 above, there is an apparent need to make the service damage assessment data accessible in order to determine measurement parameters important to some services. Some damaged services are directly tied to an economic value (e.g., commercial fishing) of a resource and some are not (e.g., passive uses).

Service oriented monitoring programs should be integrated with monitoring studies on resource recovery. A service oriented study should be funded if a study is funded on recovery of an injured resource that supports one or more services. In general, probable uses of resources should be paralleled with recovery monitoring activities associated with services for two reasons:

- Services, especially consumptive services, may affect recovery of a particular resource or linkages within the ecosystem, and

- It is important to understand how alleviating or changing the management of a particular service may affect recovery of a resource.

6.3.1 Recreation

Recreational services include activities such as sport fishing, sport hunting, boating, kayaking, camping etc.. Some of these activities (i.e., sport fishing) have a direct link to some of the injured resources. Factors to consider in integrating recreational services into a recovery monitoring program are:

- Recovery monitoring of recreational services should focus on the overlap between the different user groups and the injured resources.

- Native Corporations need to be involved in recovery monitoring activities of recreational resources because they own significant amounts of land used by recreational user groups and they are also major developers of recreational and support activities.

- Recovery or restoration of recreational service is best determined by evaluating increases in use levels (e.g., angler days), and people's perceptions after the oil spill.

- Define recovery levels or end-points for recovery of recreational services need to be defined. One approach to defining a recovery level or an end-point for recreation is to evaluate existing data to determine what the natural range of variation is for recreation and to use the evaluation to identify the variation that will be considered an acceptable end-point.

6.3.2 Subsistence

A variety of subsistence resources are used by many residents of Prince William Sound, Kenai Peninsula, lower Cook Inlet, and Kodiak islands. Subsistence resources provide food, resources, and products that are used in daily life and in cultural practices and traditions, and are a means of providing a subsistence-cash economy. Many of the subsistence resources that support a healthy subsistence community are resources that were injured by the oil spill. Important factors to consider in planning and implementing a monitoring program for subsistence are:

- As with injured resources and other damaged services a reasonable definition or end-point of subsistence recovery needs to be defined. Subsistence recovery could be defined as "when the community is harvesting resources (not necessarily the same resources) at a range comparable to pre-spill harvest rates". One approach to defining recovery or an end-point for subsistence is to evaluate existing data to determine what the natural range of variation is for subsistence harvest, and use the evaluation to identify the variation that will be considered as an acceptable end-point.

- Involvement of subsistence communities in recovery monitoring. Involvement of subsistence communities in recovery monitoring allows for an opportunity for the communities to take ownership in recovery monitoring activities. For example, cooperative agreements could be established between subsistence communities, the Trustee Council, and an agency or university with expertise and experience in data collection and management. The communities could actually implement a recovery monitoring program with oversight by an agency/university.

- Decide which subsistence communities to monitor. A decision on which subsistence communities to monitor could be determined by evaluating where changes have occurred, and the extent of changes in subsistence harvest that have already been documented, identifying representative communities in the oil spill area, and selecting representative sites within the representative communities.

- Decide what to monitor. Suggestions from interviews with subsistence experts include monitoring levels of participation and shifts in harvest areas, contaminant levels in species that subsistence users depend on, village-wide consumption levels of injured resources that support subsistence, subsistence user perceptions, economic activity of the areas, and market assessments.

- Identify an appropriate method to implement monitoring programs. For example, household interviews are one method that can be used to assess the qualitative measurement of well-being. Interviews could occur initially at all representative sites within representative communities followed by a reduced sampling effort to a few representative sites in representative communities. Using harvest levels as a basis for comparison with pre-spill harvest measurements (taking into account shifts or a new emphasis on certain resources) is also an available method to evaluate levels of participation.

- Include subsistence and fishing mortality data (from state and federal catch data) in monitoring recovery activities of particular resources.

6.3.3 Commercial Tourism

Commercial tourism is related to passive use values discussed below in Section 6.3.5. Commercial tourism is dependent, in part, on undeveloped wilderness lands and developed lands within the oil spill area. Many areas and resources that support a healthy tourist industry were injured by the oil spill. Important factors to consider in planning and implementing a monitoring program for tourist services are:

- Recovery monitoring of tourism should focus on the overlap between the different user groups and the injured resources

- Consideration of the resources that, in part, draw tourists to Alaska

- As with other injured resources and damaged services, a reasonable definition or end-point for recovery of tourism needs to be defined. Recovery of tourism could be defined as when the levels of use are at a range comparable to pre-spill levels of use . One approach to defining a recovery level or an end-point for tourism is to evaluate existing data to determine what the natural range of variation is for tourism and use the evaluation to identify the variation that will be considered an acceptable end-point.

- Decide what to monitor. Suggestions from interviews with service experts include recording the amount of use of particular areas by tourists, monitoring levels of tour boat visitors, and comparing pre- and post-spill ferry passenger data.

6.3.4 Commercial Fishing

The commercial fishing industry is the second largest generator of revenue in the state (Exxon Valdez Oil Spill Trustees 1992). Several of the injured resources identified in Section 5.0 support important commercial fisheries. Recovery monitoring of commercial fishery activities will be important for documenting recovery of the commercial resource and for minimizing impacts to fisheries through fishery management practices. Important factors to consider in planning and implementing a monitoring program for commercial fishing services are:

- Focus monitoring recovery of commercial fishing on the overlap between the different user groups and the injured resources.

- As with other injured resources and damaged services, a reasonable definition or end-point of commercial fishing recovery needs to be defined. Commercial fishing recovery could be defined as when the commercial harvests are at a range comparable to pre-spill commercial harvest rates. One approach to defining a recovery level or an end-point for commercial fishing is to evaluate existing commercial catch data to determine what the natural range of variation is for commercial catch harvest and use this evaluation to identify the variation that will be considered an acceptable end-point.

- Decide which commercial fisheries to monitor. A decision on which commercial fisheries to monitor could be determined by evaluating where changes and the extent of changes in commercial fishing have already been documented, identifying representative fishing communities in the oil spill area, and selecting representative sites within the representative commercial fishing communities.

- Decide what to monitor. Suggestions from interviews with commercial fishing experts include monitoring fishing mortality (from commercial as well as subsistence catch), effects of hatchery production on the service, escapement, economic activity of commercial fishing areas, and market assessments.

- Identifying an appropriate method to implement monitoring programs. For example, the collection of data from fish tickets is one potential method that can provide an inventory and indicator of the health of the fishery.

6.3.5 Passive Uses

Passive uses are related to recreational services and tourism and are represented by values that people place on a resource or habitat. Passive use values include aesthetic, wilderness, intrinsic, and non-use values. People generally place a high value on knowing that large undeveloped lands provide habitat for fish and wildlife and opportunities for aesthetic enjoyment and appreciation. Important factors to consider in planning and implementing a monitoring program for passive use services are:

- Monitor recovery of passive uses based on the overlap between the different user groups and the injured resources.

- As with other injured resources and damaged services a reasonable definition of a level of or end-point of passive use recovery needs to be defined. Passive use recovery could be defined based on people's perceptions.

- Decide which passive uses to monitor. A decision on which passive uses to monitor could be determined by identifying the non-use values and attempting to quantify those values.

- Identify an appropriate method to implement monitoring programs. For example, well-prepared surveys and interviews could be conducted to determine perceptions of recovery.

6.4 RELATIONSHIP OF THE *EXXON VALDEZ* SPILL MONITORING PLAN TO OTHER MONITORING PROGRAMS

Several programs have been identified that may prove useful to coordinate and/or integrate with the spill monitoring program. Many of these are listed in Table 2, a matrix for identifying elements in common between monitoring programs. Listed below are the programs that may best serve the purpose of this program.

6.4.1 Alaskan Monitoring Programs

- Cook Inlet Regional Citizens Advisory Council monitoring program
- Prince William Sound Regional Citizens Advisory Council monitoring program
- Agency resource management programs such as Alaska Fish and Game programs

6.4.2 Federal Programs in Alaska

- Agency resource management programs such as U.S. Fish and Wildlife Service programs
- NOAA Status and Trends/Mussel Watch program

6.4.3 Future Programs With Which to Coordinate

- U.S. National Park Service coastal program
- EPA's Environmental Monitoring and Assessment Program (EMAP) - Near Coastal

6.4.4 Monitoring Programs to Learn From

- Puget Sound Ambient Monitoring Program
- Beaufort Sea
- Chesapeake Bay

(The pros and cons of coordinating/integrating with these programs will be discussed once the matrix table is completed -- a workshop activity.)

Table 2.
Matrix of Exxon Valdez Injured Resources and Elements Monitored by Other Programs

	Injured at Population Level (Direct Effects)										Injured but no Population Decline							
Monitoring Element	Sea Otter	Harbor Seal	Common Murre	Marbled Murrelet	Pigeon Guillemot	Harlequin Duck	Black Oystercatcher	Sockeye	Intertidal Communities	Subtidal Communities	Killer Whale	River Otter	Bald Eagle	Cutthroat Trout	Dolly Varden	Pacific Herring	Rockfish	Pink Salmon
Sediment Chemistry																		
Sediment Toxicity (bioassays)																		
Biological Sediment Mixing Depth																		
Water Chemistry																		
Water Column Toxicity (bioassays)																		
Tissue Chemistry (fish and shellfish)																		
Groundwater Chemistry																		
Submerged Aquatic Vegetation																		
Vegetation																		
Habitat Distribution/Condition																		
Benthic: abundance, biomass, species composition																		
Fish and/or Shellfish: gross pathology, abundance, species composition																		
Mussel Watch																		
Zooplankton																		
Phytoplankton																		
Bacteria																		
Birds: water-, land-based																		
Reptiles, Amphibians																		
Mammal(s): abundance, tissue																		

Table 2. (continued)
Matrix of Exxon Valdez Injured Resources and Elements Monitored by Other Programs

			Injured/Damaged Services					
Monitoring Element	Archeological Sites and Artifacts	Designated Wilderness Areas	Commercial Fishing	Commercial Tourism	Passive Use	Subsistence	Recreation	Monitoring Programs:
Sediment Chemistry								EMAP-Near Coastal, Chesapeake Bay Basin, PSAMP, NOAA S&T, Beaufort Sea, Cook Inlet RCAC, Great Lakes, Prince William Sound RCAC
Sediment Toxicity (bioassays)								EMAP-Near Coastal, PSAMP, Denali National Park & Preserve
Biological Sediment Mixing Depth								EMAP-Near Coastal
Water Chemistry								EMAP-Near Coastal, National Surface Water Survey, Chesapeake Bay Basin, PSAMP, Great Lakes, Denali National Park & Preserve
Water Column Toxicity (bioassays)								EMAP-Near Coastal
Tissue Chemistry (fish and shellfish)								EMAP-Near Coastal, Chesapeake Bay Basin, PSAMP, NOAA S&T, Cook Inlet RCAC, Great Lakes
Groundwater Chemistry								Chesapeake Bay Basin, Great Lakes
Submerged Aquatic Vegetation								EMAP-Near Coastal, Chesapeake Bay Basin, Beaufort Sea *(kelp)*
Vegetation								Chesapeake Bay Basin, Denali National Park & Preserve
Habitat Distribution/Condition								PSAMP, Denali National Park & Preserve
Benthic: abundance, biomass, species composition								EMAP-Near Coastal, Chesapeake Bay Basin, PSAMP, Cook Inlet RCAC, Great Lakes, Denali Park & Preserve
Fish and/or Shellfish: gross pathology, abundance, species composition								EMAP-Near Coastal, National Surface Water Survey, Chesapeake Bay Basin, NOAA S&T, Beaufort Sea (fishery catch data), Cook Inlet RCAC, Great Lakes, NOAA (fisheries), AKF&G(fisheries)
Mussel Watch								EMAP-Near Coastal, NOAA S&T, Beaufort Sea, Cook Inlet RCAC, Prince William Sound RCAC
Zooplankton								National Surface Water Survey, Chesapeake Bay Basin, Great Lakes
Phytoplankton								Chesapeake Bay Basin, Great Lakes
Bacteria								Chesapeake Bay Basin
Birds: water-, land-based								Chesapeake Bay Basin, PSAMP, Beaufort Sea (oldsquaw, common eider), Great Lakes (comorants), Denali National Park & Preserve
Reptiles, Amphibians								Chesapeake Bay Basin
Mammal(s): abundance, tissue								PSAMP, Beaufort sea (bowhead whale, ringed seal), Denali National Park & Preserve (small land mammals), FWS(seabirds, sea otter, boat bird surveys), NMFS (harbor seal, sea lion)

7.0 IMPLEMENTATION AND MANAGEMENT OF MONITORING PROGRAM

MORE MATERIAL WILL BE ADDED TO SECTION 7.0

7.1 IMPLEMENTATION AND MANAGEMENT

Implementation of the monitoring program will occur during Phase 3. The overall monitoring effort should be managed by a single contractor, including coordination of the centralized data library. The format for the individual monitoring elements can either be contracted under one contract or under individual contracts for specific resource/services, or monitoring type. However, all contractors must agree to comply with a set of guidelines in order to be awarded the contract. Whether or not the monitoring is contracted as a single project or as multiple projects also depends on the funding available for monitoring. It may be most cost efficient to have a single contractor perform the work since the QA/QC overhead expenses, etc., would all be covered by a single element. Technically this could be better as well, because the sampling effort and techniques would remain consistent throughout the program. (However, information may be lost if the resource expertise is not applied.)

Management of the program consists of coordinating not only the implementation of the program but the reevaluation phases of the program, including peer review. The most certain way to ensure that the data are collected, analyzed, and presented in a scientific and meaningful manner is to remove bias from the selection and funding processes, while directing the projects with a set of requirements and guidelines. A competitive bid process is recommended, utilizing peer reviewers from the proposal to the final award stage. To achieve the goals and objectives established in the conceptual monitoring plan, monitoring activities will require effective, well coordinated management. There are many competing interests for the settlement funds. There are also numerous competing objectives and goals outside those generally agreed to in the conceptual monitoring plan. Thus, effective, well coordinated management is essential to leading monitoring activities in a manner that will attain the overall monitoring goal.

What are the elements of effective management? Although this question has many different answers, there are several components of management that will help to ensure that it is effectiveness:

- make decisions in a logical manner,
- direct activities toward established goals,
- involve interested parties,
- make decisions on a timely basis, and
- communicate decisions immediately to the involved parties

To accomplish effective management it often requires authorities to delegate more responsibility than they wish. It requires all interested parties to take substantial risks that their interest may not receive the highest priority.

At a more elemental level, a process or mechanism, such as a detailed schedule with trigger dates, can be developed for each of the project funding areas, such as monitoring, damage assessment, and restoration. These can then be overlaid to provide the Trustee Council with an overall restoration schedule from which to plan. Feedback from the public, in the form of review by the Public Advisory Group, or public comments on draft project documents, can be used by the Trustee Council to determine the priorities of the public.

Feedback from principal investigators and peer reviewers can be used by the Restoration Team and Trustee Council to determine priorities and trigger points for scientific concerns. For instance, in order for a restoration activity to be underway in summer 1993, the principal investigator may need six months' notice to allow for securing logistical support. Therefore their monitoring proposal must be reviewed and a funding decision made with at least six months' advance notice. Priority should be given to schedules that are calendar-driven or otherwise inflexible so as not to lose information. Using the example above, priority would be given to an advanced review of the proposed study so that a field season of data collection is not lost.

There may be trigger dates that overlap between project areas or time lines that are impossible for the Trustee Council to meet. These should be negotiated at the onset of planning. In situations such as these, the Trustee Council should utilize outside expertise to prioritize and/or reschedule activities and/or if possible, delegate some of their responsibilities.

The schedule described above, should be continually revised and updated, but should be relatively stable on a quarterly basis, to allow the Trustee Council to plan ahead. In addition to trigger dates, dates can be backed into the calendar to allow the Trustee Council advance warning of an upcoming event. Again, using the example above, if the Trustee Council needs a two-month period to send a proposal out for expert technical review prior to its own review of the proposal, this date can be backed in, as well as the review period necessary for the Trustee Council to complete its review.

7.1.1 <u>Peer Review Panel</u>

A panel of peer-reviewers could be selected to review and grade all proposed projects, following guidelines developed by the Trustee Council or utilizing a format similar to that used by a well accepted funding agency, such as the National Science Foundation (NSF). A similar peer-reivew process should be used for all project renewals, and for review of draft and final reports. Projects should be fully funded by the restoration funds with no subsidiary funding by a private or state agency in order to eliminate any potential conflicts of interest. This does not preclude agency scientist from bidding on monitoring elements; however , the contract award and work conducted must be discreet from other activities conducted by the agency. Matching or assisting funding might be allowed if the funding agency was independent such as the NSF or the National Institutes of Health.

A peer-reivew panel could be selected by using lists of NSF reviewers, or from the National Academy of Sciences. Additional reviews could be done through the mail, as per NSF. The panel should be relatively large, six to twenty members, and should reflect all types of biases. If the proposals are ranked on merit, the resulting reviews will be relatively unbiased as the extreme views will balance each other. Personal bias is unavoidable and unremovable, but the relatively unbiased selection of meritorious projects, coupled with adequate QA/QC procedures should foster the development of projects whose results will reflect the unbiased nature of the analyses.

Because it is expected that some of the monitoring activities may continue for several years, it may be useful to have a rotating review panel, with the rotations staggered like that of many board of directors. For instance, each reviewer can serve on the panel for only two years at a time. The first terms would be of a staggered length (one to three years) such that all the peer reviewers would not be lost at the same time.

7.1.2 Data Dissemination

All of the monitoring data should be kept in a central repository or library, accessible by a computerized system linking the available resources. How and who can utilize this system will be a decision of the Trustee Council, but oversight of the repository will be the responsibility of a contractor or agency.

It is important that the monitoring results be made known to the public. This may take the form of summary factsheets, summaries of activities in the Restoration Work Plan, or in another form selected by the Trustee Council.

The method for data dissemination is not an objective of the monitoring program, rather that data be accessible and in a format that can be readily utilized by scientists, resource managers, investigators and other interested parties.

7.1.3 Avoiding Duplication of Effort

As discussed above, integration and/or coordination with other programs is essential to avoid duplication of effort. In order to avoid this duplication it is essential to coordinate monitoring efforts between studies, both those that are funded with settlement funds, and those that are not.

To facilitate the coordination between programs it would be useful to do the following:

- Develop a computerized and hard copy table that identifies ongoing routine agency monitoring activities for resources and services affected by the oil spill or that occur within the oil spill area. (Incorporating GIS for maps may be useful for this purpose).
- Communicate with state and federal resource agencies to follow changes in routine agency monitoring activities.

TO BE COMPLETED

7.2 REQUEST FOR PROPOSAL AND CONTRACT LANGUAGE

Developing contracts to ensure timely performance requires incentives for completing tasks on schedule and disincentives for tardiness. Incentives for completing tasks on time could include financial bonuses or some type of preferred status in selection for future rounds of project work. Disincentives could include the loss of money through financial penalties, or exclusion from consideration for any further project work. For example, Standard contract language could require the contractor to inform the client within a certain number of days if there was a problem that would affect schedule. However, early notification to the client must be tied to some type of penalty for being late with deliverables without notification.

If the funding for a project comes from public sources, an additional means of encouraging timely performance might be a regular (monthly/quarterly) public review of project contracts. This could take the form of a public meeting where contractors had to explain where they were in the project and why they are not meeting the schedule. An alternative would be a regular display ad in local newspapers indicating who is responsible for different projects and whether they were on schedule.

Examples of a few contract scenarios that use the basic ideas of incentives and penalties are described below:

Assumptions:

- Lump sum type contract
- $100,000 total contract value
- $20,000 mobilization costs

7.2.1 Payment Tied to Deliverables/Schedule

This form of contract would incorporate a number of deliverables such as reports or milestones in the monitoring process and payment would come after satisfactory completion of those identified tasks.

Example: If the project had 4 equal cost deliverables or milestones, the contractor would receive $20,000 up front to get started and would receive the next $20,000 payment upon completion of Task 1. Upon completion of Task 2, another $20,000 would be paid.

The contractor does not receive money for the next phase until they produce the deliverable.

Pros/cons: This approach would work well with sequential tasks. The contractor would not have funds to work on Task 2 until they successfully completed Task 1. I would assume both public and private groups would want a relatively steady stream of funding to avoid extreme staffing fluctuations.

This idea would not work as well if the tasks have to occur concurrently, or if the contractor has to be paid up-front. If the tasks occur simultaneously, then staffing levels would be committed and there would not be an incentive to complete tasks in a timely manner.

The same problem would exist if the contractor is a public agency and needs total funding up-front before beginning work. In this case, they have already received the money and there would not be a clear incentive for them to meet the schedule without including a penalty.

7.2.2 Percent Reduction in Contract Total Value

This concept would include a penalty clause that reduces the total value of the contract based on lack of timeliness. A % of the total value of the contract is established as a penalty and is withheld for each day/week the product is late.

Example: If the project has 4 equal cost deliverables, the contractor receives $20,000 up front to get started. If they are one week late delivering a report and the penalty is calculated at 1%/week, then the penalty is $1,000 ($100,000*.1). This penalty could be taken out of each specific task (i.e. payment of $19,000 for completion of Task 1), or it could be withheld from the last payment for completing Task 4. If the contractor was one week late on Task 1, met the schedule in Tasks 2 and 3, and was one week late in Task 4; then the final payment would be $18,000.

This approach could also be used to create an incentive. If the product or milestone is reached early, a reward could be established that would provide additional money or some other benefit to the contractor.

7.2.3 Incentive for Continuing Project Involvement

This approach assumes there will be a built in incentive for public/private groups to continue their involvement with the project. For example, an agency or private entity will want to be associated with doing the long-term monitoring of northern fur seals for the Exxon Valdez project.

Example: If a contractor successfully meets their deadlines/milestones in Year 1, they are automatically given first opportunity to do similar work in Year 2. If they miss a deadline, then the work in Year 2 is out for competitive bid and the contractor runs the risk of losing the work. If the contractor decides to bid on Year 2 after losing automatic "rehire" rights, they will have to explain to the satisfaction of the Trustees their lack of performance in Year 1.

Depending upon the details of the specific monitoring project, a combination of these concepts could be developed to include both incentives and penalties in the same contract.

7.2.4 Schedule for Deliverables, Performance Criteria and Proposal Ranking

- Schedule for deliverables so if problems arise client must be informed in advance and presented with solution.

- Performance criteria for meeting QA/QC requirements, standard protocols, compatible data

Again, an attachment could be used in a contract document to identify what specific protocols will be used, what criteria will be used to assess compliance with QA/QC requirements, and the data format to be used for compatibility.

- Factors for proposal review and requirements for ranking/rating proposals

This isn't a contractual issue and should be handled on a technical basis. A copy of the contract with payment provisions could be included in the Request for Proposal (RFP). The contractor could be asked to comment on any potential problems they saw with the contract format and that could be helpful in determining the successful candidate.

8.0 RECOMMENDATIONS

TO BE COMPLETED

9.0 REFERENCES CITED

TO BE COMPLETED

APPENDIX A

LIST OF INTERVIEW PARTICIPANTS

APPENDIX B

LIST OF WORKSHOP PARTICIPANTS

APPENDIX C

TEN STEPS TO A STRONG MONITORING PROGRAM[1]

CLEAR GUIDANCE: as to how data are to be used and type of decisions to be made.

GOALS: establish scientifically, technologically, logistically and financially achievable goals.

DECISION INTEGRATION: decision points and feedback loops should be clearly established and integrated into decision process prior to data collection.

AUTHORITY & CONTROL: explain, then define where they reside and provide local controls compatible with program controls and objectives.

PARTICIPANT COMMUNICATION: identify communication channels among participants and ensure they are both functional and interconnected.

INTEGRATE NEEDS: for regulation, data acquisition and management of local, state and federal agencies to optimize use of available resources.

PUBLIC & SCIENTIFIC INVOLVEMENT: establish mechanisms to ensure these groups are participants early and often.

COMMUNICATION: establish mechanisms to ensure conclusions are communicated to both decision makers and the public in terms they can understand and act upon.

REVIEW: include mechanisms for periodic review and re-direction when results or information justifies a change.

MANAGEMENT: identify in advance, actions to be taken in response to expected and unexpected results.

---

[1] Restated from Natural Research Council. 1990. Managing Troubled Waters, National Academy Press, Washington D.C.