International Council for
the Exploration of the Sea

ICES C.M. 1980/D:16
Statistics Committee

A MANAGEMENT INFORMATION SYSTEM FOR THE NORTHEAST MONITORING PROGRAM

by

Kathryn Paine, Eugene G. Heyerdahl
National Marine Fisheries Service
Northeast Fisheries Center
Woods Hole, Massachusetts 02543 USA

ABSTRACT

This paper proposes a method for developing an information system to manage data collected by complex, interdisciplinary, ocean monitoring programs.

The objectives met by this method are described and the rationale for each step explained, with examples. The proposed method is being considered for use in the design of NEMPIS, the Northeast Monitoring Program Information System.

TABLE OF CONTENTS

Page

ABSTRACT 0

1. INTRODUCTION 1
   1.1 An overview of the monitoring program 1
   1.2 The development of the information system 2

2. THE SYSTEM DEVELOPMENT PROCESS 2
   2.1 The Problem 2
   2.2 System Design Objectives 2
   2.3 System Characteristics 3
      2.3.1 Modularity 3
      2.3.2 Top down design 3
      2.3.3 Structured analysis, design, and programming . . 3
      2.3.4 Immediate, on-line documentation, with a status flag 4
      2.3.5 Computerized communications among participants . 4
      2.3.6 Immediate hands-on access to the system by all participants 4
      2.3.7 Immediately available first pass programs 4
      2.3.8 Data integrity 5

3. NEMPIS DESCRIPTION, A SYSTEM WITHIN A SYSTEM 5

4. SYSTEM DESCRIPTION COMPONENTS 6
   4.1 Overview of NEMPIS and NEMP 6
   4.2 The participants 6
   4.3 The data sets 6
   4.4 The platforms 7
   4.5 The types of analysis 7
   4.6 The analysis tools 7
   4.7 The products 7
   4.8 The population sampled, when and where 8
   4.9 The linkages 8
   4.10 Events in NEMP, including NEMPIS 8
   4.11 The NEMP schedule 8
   4.12 The history of NEMP, including NEMPIS 8
   4.13 Reference documentation 9
   4.14 Hotline, key status 9
   4.15 Glossary 9
   4.16 Appendices 9

5. IMPLEMENTATION OF NEMPIS 9

6. CONCLUSION 9

This paper not to be cited without prior reference to the authors.

International Council for the Exploration of the Sea

ICES C.M. 1980/D:16
Statistics Committee

A MANAGEMENT INFORMATION SYSTEM FOR THE NORTHEAST MONITORING PROGRAM

by

Kathryn Paine, Eugene G. Heyerdahl
National Marine Fisheries Service
Northeast Fisheries Center
Woods Hole, Massachusetts 02543 USA

1. INTRODUCTION

1.1 An overview of the monitoring program.

Currently under development is a U.S. program to monitor the condition of Atlantic coastal waters. It represents an integration of several previous programs sponsored by the National Oceanic and Atmospheric Administration (NOAA) such as Ocean Pulse, the MESA New York Bight Project, and the Ocean Dumping Program. The goal of the New England Monitoring Program (NEMP) is to provide a systematic time series of observations of components of the marine ecosystem, natural and introduced, so as to determine existing levels, trends, and natural variations in these components such that unnatural, and possibly destructive, variations may be detected. Data sets are to be collected from a variety of disciplines, including physiology, pathobiology, genetics, benthos, chemistry, oceanography, and fisheries. Participants in the program are associated with several agencies of NOAA, including the Northeast Fisheries Center (NEFC) of the National Marine Fisheries Service (NMFS), Research and Development (RD), and National Ocean Survey (NOS). They are also located in separate geographic locations ranging from Oxford, Maryland, to Gloucester, Massachusetts. An office to coordinate the entire program has been established, a management team designated, and functional responsibilities assigned to various NOAA components.

Monitoring activities of selected sites will be supported by a variety of platforms including research vessels, deployed instruments such as current meters, manned undersea observations, remote sensing, and coordinated laboratory experiments. Data from previously established research and monitoring programs will be used as appropriate to help establish baselines.

The NEMP program was initiated in 1980 and in addition to establishing a monitoring program for the health of the ocean in this area is being looked to for assessing the usefulness of some of the current and proposed techniques for monitoring pollution. The program is also expected to change as this information becomes known.

1.2 The development of the information system.

With data collections and analyses spread both geographically and over many disciplines, the ability to process and archive data on computers is basic to the success of a program such as NEMP in meeting its objectives.

There is a need for a more flexible and responsive approach to the task of analyzing the requirements of such a complex system than is traditionally employed. In itself, the need to manage and coordinate the development of the information system to handle the data, from collection to analysis, is a challenging requirement of the system.

A feedback mechanism which successfully collects and disseminates information about the system development in a timely and pertinent manner is needed to insure that no pieces are left unattended or unattached from the system as a whole. Active information exchange among all participants must delve into the key issues of the system development process at all stages, and the status of the system at any time must be clearly stated and known to all. The techniques employed in information exchange must be easy to use, so that investigators can actively participate without an inordinate investment of time in the communication process itself, and primary energy is invested in the problems at hand. Participants must be able to declare problems as they arise without fear of unduly disrupting the system, and the system must permit an active and thorough investigation of single threads without impeding progress on other components.

2. THE SYSTEM DEVELOPMENT PROCESS.

2.1 The Problem.

Many years of experience suggest that the time it takes to develop a working data processing system is probably close to an exponential function of the number of people involved. The common cries are "I can't get back at my data after it goes on the computers in time to meet research requirements", and "I can't tell what is happening in time to use alternative procedures." The computer people say "the user never knows what he really wants for output" and "I spend so much time putting together documentation and progress reports I have no time to do what's needed to make progress."

2.2 System Design Objectives.

The following set of objectives for the development of a data management system attempts to address these problems:

The system should have the capability to

- describe to any interested participant the current state of the system, in terms of its composition and developmental stage, with explicit definitions of and relationships among data elements and procedures.

- allow individual data sets to be immediately accessible to interested users as soon as the data enters the system, without the need to wait for anticipated higher level syntheses.

- provide sufficient control over the procedures for data collection, entry, and storage to assure specified levels of data integrity, i.e., outputs should be predictable based on a complete knowledge of the data base contents and the algorithm employed to portray the data. Such knowledge implies an accessible, time referenced record of any changes, additions, corrections, and deletions to the data or to collection or entry procedures.

- provide a means to accept proposals for system modification from participants and to expedite information dissemination, investigation, and the reaching of consensus so as to determine appropriate action.

- accommodate change in the requirements of the system and resulting system modification without undue disruption to the system as a whole.

2.3 <u>System Characteristics</u>.

A system development process which meets these objectives would have several characteristics.

2.3.1 <u>Modularity</u>.

One factor which allows for change and growth in a system is to build with simple basic blocks and then control the linkages between the blocks. This allows, for example, a first pass, simple report procedure to be replaced with a more elegant one at a later time, with minimal disruption to other system components.

2.3.2 <u>Top down design</u>.

Top down design of the entire system with successive refinement of the component parts allows a system to be up and functional at a minimal level in the shortest time possible. The technique is to lay out the entire structure of the system as viewed "from the top", and then develop individual strands of the system on a priority basis.

2.3.3 <u>Structured analysis, design, and programming</u>.

These techniques emphasize an accurate delineation of linkages between control modules in terms of the flow of data elements through the system together with immediate documentation of the definition of data elements in both user oriented and data processing terminology. There is an emphasis placed on readability and simplicity in the diagrams which display the linkages.

2.3.4 Immediate, on-line documentation, with a status flag.

In a more traditional approach nothing becomes documented and accessible for general review until it is actually part of the system. Thus too much time can be invested in a misunderstanding, or a good idea lost because it was never documented. The opportunity to control the growth and direction of a large system is dependent on the availability of good information, both for the project manager and the participants. The concept of a condition flag to describe the status or level of completion of each system component, allows the use of terms such as "identified", "proposed", "operational", and "excluded". See Table 2.1 for a list of possible status conditions.

2.3.5 Computerized communications among participants.

Traditional system design projects require a great deal of paper flow. The project manager is in the position of deciding who needs to know what, and when. If too much is distributed, it may not be read; if too little, a good idea may be lost. With automated documentation a participant can look at as much or as little as he himself thinks is appropriate at the time, with some confidence that the information is current.

2.3.6 Immediate hands-on access to the system by all participants.

Automated documentation beginning immediately in the analysis of requirements stage of the system design requires that participants become involved with the computer system per se, with the general purpose utilities and editors, at the start, to both enter and retrieve data. This will introduce the training experience that usually comes later on when people learn to use the actual data system, and may eliminate some of the frustration and confusion that usually accompanies this stage. Potential operational problems are likely to be addressed sooner, involving communications, hardware availability, operations staffing, etc. The introduction of early on hands on experience by all participants allows an opportunity to resolve the universal data processing problems that accompany any new system development so that users can concentrate on the system of interest when it is available.

2.3.7 Immediately available first pass programs.

No data file should exist in the system without a corresponding retrieval tool which will allow that data to be examined by an investigator to at least a minimal level. There should be a procedure to connect data with whatever common statistical packages and report generator tools are available on the computer system, and the procedure should be available to system participants without programmer intervention. Programs and procedures can then be developed as time allows to provide analyses and data displays of special interest and format, and to couple data among separated files. It is often difficult for the investigator to state in advance all of the permutations and algorithms and data portrayals of interest, and a great deal of time can be spent developing specialized software. A computer is much more efficient than a human for organizing and summarizing data in a variety of ways and the ability to do this is usually already available on a computer system without a special programming effort. The investigator can then perform higher level analysis and combinations of data manually as needed on an interim basis until the more sophisticated procedures are automated.

2.3.8 Data integrity.

The most difficult concept to integrate with the other characteristics is that of data integrity. The capability for flexibility in data base design and software presents a problem in recreating early results at a later time in system development. As new disciplines are integrated into the system there will be a desire to go back and redo some of the earlier analyses. If data is filtered into the system as it becomes available, rather than waiting for a complete data set, a second analysis may give different results. There will need to be reliance on the status code to identify levels of confidence in the data and results, on reference dates, on archiving versions of both data sets and documentation, and providing informative caveats on data products produced by the system.

3. NEMPIS DESCRIPTION, A SYSTEM WITHIN A SYSTEM.

The System Description provides the key control function for the development and implementation of NEMPIS.

Table 3.1 is the table of contents to the System Description. Employing the philosophy of this proposed approach, the table of contents would be entered into the computer immediately and given the status "proposed". Table 2.1 gives a list of status codes or conditions. This table would also be entered on the computer, as Appendix A, with a status of "proposed". The structure of the table of contents file, the status code file, and section format file would be entered in Appendix B, Section Formats. Three simple to use procedures would be developed to produce a report on these three components, and a reference to these entered in Appendix C, "Directory of System Status Reports." The programs to produce the reports would be interactive, offering instructions to the user at execution time.

The basic philosophy of the proposed system has thus been fulfilled:

A top down overview has been provided.

The modules are defined so that an upgrade of any can occur with minimal impact on another.

The status is clear.

Reports on the system as it stands are immediately obtainable by any system participant.

Table 3.1

A proposed Table of Contents for Online System Documentation.

1. Overview of the system.
2. The participants.
3. The data sets.
4. The platforms.
5. The types of analysis.
6. The analysis tools.
7. The products.
8. The population sampled, when and where.
9. The linkages.
10. Events in the information system.
11. The system schedule of events.
12. The system history.
13. References.
14. Hotline, key status.
15. Glossary.
16. Appendices.

The step wise refinements can work from discipline to major data sets to data files to data elements. For each level, there can be established linkages, as required, to other components in the system: the participants, the platforms, the reports, etc.

Again, nothing is entered that cannot be immediately reported using a simple procedure file available to any user.

A convenient way to store the more detailed information about elements and files would be to use an existing data description system such as that already developed by the Northeast Fisheries Center.

4.4 The platforms.

This section would contain a reference to each of the participating (or proposed) research vessels, ships of opportunity, remote sensing vehicles, and major instrumentation systems which are used to collect data for the monitoring program. A broader definition of the term allows reference also to the capabilities provided with the information system to analyze data, including the computer central processor and special hardware components such as digitizers, terminals, and plotters.

4.5 The types of analysis.

This section describes which statistics may be relevant, which graphical display techniques are desired or under consideration, etc. If implemented, it would point to how a participant accesses the tool. Included could be references to some of the data processing oriented procedures, such as quality control, data editing, archiving, and even systems analysis and design. The use of a computer to create a product is not necessarily implied, this section is essentially a menu of possibilities.

4.6 The analysis tools.

There is a direct relationship to section 4.5, analysis types, but this section is more descriptive of what NEMPIS actually offers to NEMP participants. Again, the opportunity to use a status code allows information to be captured without waiting until the process is complete.

4.7 The products.

Given the data sets and analysis tools, a wide variety of reports are possible. The end products as required to support the monitoring program will vary in complexity and difficulty in obtaining. Stepwise refinement and status indication play an important role in this section. "Final" products can be defined at any time, but interim products which will minimally satisfy requirements in defined time frames can also be defined. It is in developing the information in this section that users will have to have a

4. SYSTEM DESCRIPTION COMPONENTS.

4.1 Overview of NEMPIS and NEMP.

The automated documentation provided here would be mainly narrative description, of a nature similar to that provided in Section 1 of this paper.

4.2 The participants.

As in all other sections, this would contain an introduction and a status of the section as a whole. At first the information may be entered in free form as "comments", but in time a more useful format might include the following data elements:

name, status
organization
commercial telephone number, FTS number
work address
which major data sets are of interest, and for each,
  the level(s) of involvement, i.e.
    provides data, uses, manages
and for each level of involvement, when or how often,
  and other factors as seem convenient.

As usual, no data structure is put into the system without also at the same time developing the ability to create a report about the data, the procedure directly available to any system participant. Also the status of each person is clearly stated as well as the list as a whole, so that an analyst can capture and document information without waiting for a final writeoff from the system manager.

It would be appropriate to add to the glossary a reference to the organizations, with some narrative description which spells out an acronym, shows the relationship to a parent or child organization, gives the major mission, and/or other pertinent data.

Finally, a data flow chart would incorporate the information in this section, linking individuals within and between organizations based on data of interest. Until graphic tools are available, the chart would remain as part of the manual complement to the automated sections of the information system.

4.3 The data sets.

It is in this section that the top-down/status techniques are really useful. The individual investigators of NEMP have varying levels of expectancy from the data system, and it is defining the data to be included that will be most complex and variable. The structure of this section can allow at first simply the designation of major disciplines as proposed, and then each discipline can be refined on a schedule as determined by the system manager.

ability to archive compared to the cost to recreate (or not be able to recreate) reports will depend on the availability of journaling capabilities in data base management systems software and the cost to both archive and retrieve the data.

4.13 Reference documentation.

This section allows the participants to be directed to reports for reference.

4.14 Hotline, key status.

This section should be developed to provide "bottom line" status reports of NEMP and NEMPIS. It would contain information culled from the rest of the NEMPIS Description System by the system manager as being of immediate interest. It should also accept comments from participants. Reference to "red flag" reports would be made. Major breakthroughs or breakdowns in NEMP or NEMPIS would be noted.

4.15 Glossary.

This will provide a definition of key terms, especially any jargon employed in the system. When a question about the meaning of a term is asked, the term should be added to the glossary.

4.16 Appendices.

Included would be a table of status codes with the meaning of each; the formal computer file structure of each section in NEMPIS Description; a reference to each program which generates reports on NEMPIS Description; the data element dictionary; the data file directory; a source document directory; and other tables as needed.

5. IMPLEMENTATION OF NEMPIS.

The philosophy described in this proposal suggests a staffing pattern that is different from that of a typical system development project. Implied is the availability of a computer system from inception, rather than a separate stage for analysis, design, implementation, operation, and revision. There is therefore a need to provide staff time for the function of data entry, quality control and editing, programming, analysis, design, user training and assitance, almost from the first day.

6. CONCLUSION.

There is an overhead associated with the recommended approach, the redoing of sections of the system as each component is refined. However, the payoff seems to justify the cost:

clear and specific understanding of what must be available on call, as opposed to what is preferred. The relationship of what the automated part of the data system can do to supplement manual report generation must be defined. If a report or portion thereof is to come directly from the computer, an opportunity to enter manually derived parameters should be available, in the event that anticipated computer supplied data is not available.

The concept of the "red flag" report would be highlighted as an available report.

4.8 The population sampled, when and where.

The second most variable aspect of NEMP is the definition of the population described. There is a set of proposed NEMP stations, but not all data elements are sampled equally, or over the same time span. This section allows participants to identify and propose without actually committing the system.

4.9 The linkages.

The most complicated (and interesting) component of the system is a definition of the linkages between each of the components defined above. As NEMPIS evolves, this section will provide the description of what is actually hooked together, what is proposed, what has been eliminated from consideration, etc. This section provides the description of the system design at a given point in time.

4.10 Events in NEMP, including NEMPIS.

This section can be used as a control function for NEMP events to whatever degree seems reasonable to the NEMP manager. Cruises, major reports, acquisition of gear, entry of new disciplines, hiring of personnel, any of these could be noted. The usual NEMPIS standard is employed, nothing should be entered that can't be reported easily.

4.11 The NEMP schedule.

As above, this provides a mangement tool which can be used to the level of detail that seems most useful. The concept of project control for data processing activities has been implemented on a small scale in NERFIS and could be used to supplement or provide the complete capability for NEMPIS.

4.12 The history of NEMP, including NEMPIS.

This section provides an opportunity to retrace steps as needed, by recording the change history of each component of the NERFIS System Description. Complementing this section would be an archived version of the section to correspond to the revision date. The corresponding archival of revised data sets will depend on the nature of this change. A compromise between the

Table 2.1

A proposed set of conditions for the status - of - the - system.

identified

proposed

planned

excluded, considered but never entered system

in development

quasi-operational, test stage

operational

not operational, being changed

operational, but may be excluded

temporarily disconnected

permanently dropped, but was in system once

- the accessibility of data as soon as it is available, in at least some form.
- the ability to create a product immediately, to some level of interest.
- the availability of a description of the current version of NEMPIS, allowing progress reports to be generated on call.
- the availability to all participants of the status of the system, at any time.
- the implied opportunity to control growth and set priorities based on this availability of information.
- the opportunity for participants to identify and propose as a need becomes apparent without actually affecting the system immediately.

Figure 6.1 System Objectives.

- Status of the system development project known to participants at all times.
- Data accessible to system participants immediately after entering system.
- Immediate feedback on design proposals.
- Multi-way communication among system participants.
- Controlled growth of system with flexibility to accomodate changing requirements.
- Documentation about system available as system evolves.
- Linkages and relationships between system components defined.
- Control of the system "front end" for data integrity.

Figure 6.2 System Characteristics

. Modular, top down design and implementation with successive refinement.

. Use of a status flag to describe the level of implementation for each component.

. Use of structured techniques for analysis, design, and programming.

. Online documentation for all levels of implementation.

. Use of computer system for multi-directional communication among participants.

. Programs available to participants to retrieve and format data files and simple reports, immediately available when data enters system.

. First pass, then successively higher level report and analysis products.

. Easy access to "canned" statistical and graphics software.

. Immediate hands on access to computer system for all participants.

. Use of journaling techniques, archiving of versions of system.

. A controlled "front end", a flexible "Back end."