NBSIR 83-2806

Scientific Littoral Data Management Study Summary Report

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Washington, DC 20234

September 1983

Issued January 1984

Sponsored by:
U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Washington, DC 20234

and

Coast Engineering Research Center
U.S. Army Corps of Engineers
/aterways Experiment Station
.O. Box 631
icksburg, MS 39180

NBSIR 83-2806

SCIENTIFIC LITTORAL DATA MANAGEMENT STUDY SUMMARY REPORT

Ralph E. Schofer and Frank F. Goodyear*

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Washington, DC 20234

September 1983

Issued January 1984

Sponsored by:
U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
National Engineering Laboratory
Center for Applied Mathematics
Washington, DC 20234

and

Coast Engineering Research Center
U.S. Army Corps of Engineers
Waterways Experiment Station
P.O. Box 631
Vicksburg, MS 39180

*Fisk Systems
4701 Broad Brook Drive
Bethesda, MD 20814

U.S. DEPARTMENT OF COMMERCE, Malcolm Baldrige, *Secretary*
NATIONAL BUREAU OF STANDARDS, Ernest Ambler, *Director*

TABLE OF CONTENTS

Scientific Littoral Data Management Study

COASTAL ENGINEERING RESEARCH CENTER
U.S. ARMY CORPS OF ENGINEERS
SCIENTIFIC LITTORAL DATA MANAGEMENT STUDY

I. Introduction

In the summer of 1981, the Coastal Engineering Research Center (CERC), U.S. Army Corps of Engineers, asked the Operations Research Division, National Bureau of Standards to review existing littoral data management practices at CERC and provide recommendations and criteria for development of an improved data management system. The first assignment was to focus upon "current" data defined as automated data in a format comparable to that being currently collected in 1981. Subsequently a second phase was initiated which focused upon "old" data defined as automated data no longer collected in that particular mode or format. A final phase which focused upon the indexing and cataloging of unautomated data was begun in the Spring of 1983 and was concluded in September 1983. These data included analog littoral data, sand and core boring samples, and selected maps and charts.

The three bound reports coming out of this two year study comprise the collected reports of the individual tasks of the three study phases. During the course of this study the CERC has been relocated from Ft. Belvoir, VA to Vicksburg MS. The NBS Operations Research Division has taken special care during the study to provide for the needs of relocating a data base at the same time indexing and cataloging were being accomplished.

II. Study Objectives

To develop a viable, accessible data base supportive of CERC's research and data dissemination requirements.

III. Study Methodology

Review and assess CERC's data requirements, and current (and previous) procedures for collecting, editing, processing, indexing, labeling, storing, and retrieving these data. Evaluate these procedures considering observed requirements and the principles of good data management practice. Prepare recommendations for improvement and an implementation plan for installing these improvements.

IV. "Current" Data Study

A. Findings

A detailed description of findings has been provided in the "current" data report. These findings can be summarized as follows:

- Data activities are extremely fragmented at CERC; no one person has overall responsibility for overseeing data functions and researchers are called upon to retrieve data for other users.

- Organization of computer resources is not efficient and is not suitable for effective coordination or performance of the data management functions.

- Generally data are not properly labeled, indexed and organized to facilitate retrieval and use.

- Procedures are not developed for condensing data files and for maintaining or reevaluating or discarding aging data.

B. Recommendations:

Recommendations are indicated below in two groups: recommendations selected for accelerated implementation; and recommendations selected for routine implementation.

Recommendations Selected for Accelerated Implementation

- Establish centralized data management position.

- Discontinue computer operations at Kirtland AFB. (completed)

- Adopt a distributed data processing mode of operation. (This recommendation was adopted, but has been delayed due to the impending relocation of CERC.)

- Develop a system to compress and provide indices to the raw WAVE data. (completed.)

Recommendations Selected for Routine Implementation

- Use procedures presented in Subtask G of Task III of the NBS "Current" data report to CERC to develop a location file index (one record for each data collection location). Location of data collection is one of the most useful parameters in selecting data.

- Use procedures presented in Subtask G of Task III of the NBS "Current" data to CERC to develop a data indexing and cataloging system for quickly determining whether existing data are available to support a particular research need. This system will relate selected data parameters (wave height, date, gage numbers) to a specific record (or records) e.g. magnetic tape number of disk name or number and will enable potential users to search either in a manual or an interactive mode. The latter capability provides for conversation with a computer file using a terminal.

- Utilize the S2000 DBMS*, available to Department of Army Users under existing arrangements, for interactive retrieval of catalogued indices of data sets for determining existence and location of desired data. Data can easily be loaded into this sytem which provides for interactive search and retrieval on combinations of specified criteria. This system is available for Corps of Engineers use.

- Develop a manual data index retrieval system. Use the procedures described in Subtask G of Task III of the NBS "Current" data report to CERC to develop a manual retrieval system. After this system is developed, it should be maintained and be available in the Coastal Engineering Information Center (CEIC).

- Use standardized procedures as recommended in Subtask F of Task III of the NBS "current" data report to CERC for logging and labeling of CERC "current" data files. This standardization is particularly important at CERC because of the decentralized mode of data file creation. Without strict adherence to standardized procedures, it is necessary to rely on organizational stability and long term availability of key employees. This is unacceptable in today's mobile environment.

- Relocate the Rapid Sand Analysis (RSA) Program from MERADCOM to the inhouse MODCOMP IV Computer. This will reduce ADP and personnel costs.

  Document the RSA Computer programs and keep documentation up-to-date in a secure location.
  Principal investigators should be required to enter location of sample (i.e. field site) in the data files.

- Urge MERADCOM to change their operating system to take advantage of the safeguards which tape labels normally provide.

  As described in Subtask E of Task III of the NBS "Current" data report, current operating procedures at MERADCOM do not afford the normal safeguards which tape labels usually provide. Users must work without these safeguards at MERADCOM until their operational procedures are improved. It is recommended that MERADCOM be encouraged to improve the operating system as soon as possible.

- The location of BPAS data collection sites should be placed in specific fields rather than be randomly located. Columns 79-80 should be used to indicate the initials of the state in which the data are collected.

- Develop data sharing procedures. Consideration should be given to joining with other Federal data center(s) to facilitate distribution of CERC data to a broader user clientele.

*Commercial equipment are identified in this paper to describe the procedures which were recommended or implemented to utilize equipment currently available to the sponsor. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the equipment identified are necessarily the best available for the purpose.

- Document data file computer programs. All programs which create data files must be well documented and maintained. Sufficient personnel time must be allowed to meet this requirement.

- Coastal Engineering Information Center (CEIC) should be the central focus for all data requests. This will facilitate central response to data requests and will remove the data dissemination functions from the research staff and provide a list of data users at one location.

- Select records to be kept permanently. There are currently about 1300 WV tapes (which contain digital wave data) plus an undetermined number of CERC tapes in storage at the Kingman Building. Some of these data are unique; but most of these data are likely to be routine observations, one set of which can be substituted for another set. With the current volume of incoming data, the amount of data in storage is growing at a high rate. It is recommended that CERC obtain the services of a qualified sampling statistician, familiar with spectral analysis, to develop multiple parameters for classifying three dimensional wave data. These parameters should be identified and summarized as part of the primary data analysis when new wave data first enters the system. After a 3-5 year holding period, these parameters could be used to select data to be retained and data to be discarded. Implementation of such a procedure would reduce data management problems and the volume of tapes to be stored.

- Improve ADP support of data management functions. The lack of effective ADP support underlies a number of data management problems within CERC. As one example, the LEO masterfile has not been updated for the years 1979-1981. In addition, LEO project personnel need to be trained in the usage of the new system so editing and correcting the 1982 data can proceed. It is recommended that the ADP function be relocated wihin CERC to provide closer supervision of this vital function.

C. Deliverables

In addition to the aforementioned listing of general recommendations, the following items were provided to CERC:

1. System design and detailed specifications for a computer program for compressing and providing indices to the raw waves data. This program was subsequently coded and tested by CERC personnel and is operational.

2. Prototype format for a location index for all major sites of CERC data collection activities. This system will provide information on location of data. It will not, as presently conceived, retrieve the actual data. The System 2000 with Procedural Language Interface (PLI) would be a logical retrieval system for CERC to use since the software has already been purchased by the Corps of Engineers. CERC plans to implement the system by entering the index data in the future.

3. Prototype uniform formats for labeling and logging data files.

4. A prototype procedure for selecting wave data tapes for retention. Extreme values should be determined and listed in index (labels) for subsequent use in making save/discard decisions.

More detailed information is available in "Report on the 'Current' Data Management Improvement Program", September, 1983.

V. "Old" Data Study

A. Findings

1. General

- None of the "old" data files have had extensive usage. This is probably the result of poor accessability. Only LEO data is centrally located and easily accessible.

2. WAVEDATA

- Indices for the "old" Waves data (1266 WV tapes for the 1966-1978 time period) does not exist.

- Many of the early spectral computer runs are of no value because the order of multiplexed gage readings does not agree with the order on the data tape dumps.

- The format of the "old" waves data changed eight times during the 1966-1978 period.

- Waves WV tapes began May 22, 1966 and went through September 22, 1978. Data Acquisition System (DAS) data came on line September 23, 1978 and continued until November 4, 1980.

- A systematic 2% sample of "old" data tapes were tested for readability. Only one of twenty tapes were not usable. On another tape in the sample the first record was unreadable. Extrapolating from this sample, it is estimated that 90% of old WV tapes are readable.

- Documentation was not prepared for the WV tapes.

- About 25% of the WV tapes are estimated to be error-free; about 55% of the tapes have trivial problems. The biggest problem category was the absence of one or more gages which was expected to be in the data set. About 80% of the WV tapes are available for analysis.

- The WV tapes are not in a useful form for easy user accessibility or for making informed decisions about retention.

3. BEACH PROFILE DATA

- Beach Profile data exist as separate data sets, not as a unified data base. Beach profile data was generated from 27 projects and is maintained in either the Engineering Development Division or the Research Division at CERC.

- The bulk of the Beach Profile data is stored on magnetic tape; much of the remainder is on punched cards. About 50% is in Beach Profile Analysis System (BPAS) format and the remainder can easily be converted to BPAS format.

- Beach Profile data usually must be accessed through the principal investigator.

4. Littoral Environmental Observation (LEO) DATA

- The LEO project was begun in 1968. The time period for "old" LEO data is 1968-1978.

- LEO data is maintained on four magnetic tapes, one for each coast and one for the Great Lakes.

B. Recommendations:

1. General

- All data items to be retained by CERC should be clearly labeled and numbered and stored serially for easy access.

- All data items to be retained should be listed in a catalog and some control should be exercised to safeguard these data items.

- Obsolete data items should be discarded promptly.

- A plan should be established for recopying magnetic data tapes when they reach 8-10 years old.

2. WAVES DATA

- Wave data should be put into an accessible format to facilitate search, retrieval, and the making of informed decisions or selective retention. It is strongly recommended that gage-specific data tapes be created (at a density of 1600 BPI or higher) with identification added at the beginning of each record. (A record thus defined contains readings from one gage for one continuous time period.) The advantages are many and significant and are fully described in Task III of the NBS Report on the Data Management Improvement Program for the "Old" Data.

- CERC should develop criteria for establishing a selective retention program for waves data.

3. BEACH PROFILE DATA

- Beach Profile data should be relabeled and stored in a central relocation at CERC.
- An index of the Beach Profile data should be developed and maintained in the CERC Technical Information Division Information Center.
- Beach Profile header data should be expanded to include state abbreviation in columns 79 and 80.
- Beach Profile data should be put on magnetic tape in "BPAS" format as and if requests for these data are received and fulfilled.

4. LITTORAL ENVIRONMENTAL OBSERVATIONS (LEO) DATA

- Switch main four LEO tapes with the four backup tapes occasionally in order to exercise both tape sets.
- "Remarks" from LEO observer data collection forms should be classified and coded. This will facilitate access to all the LEO data and obviate the need to retain these forms.

C. Deliverables

In addition to the aforementioned listing of recommendations, the following items were provided to CERC:

1. Indices from the "old" Waves data (1266 WV tapes for the 1966-1978 time period including the characteristics of each tape in machine-readable form. NBS performed the computer programming to produce the WV tape.
2. Updated LEO tapes (1968-1978) wherein remarks of observers have been encoded in the machine-readable records.
3. Assistance in the process of selection for discarding WV wave data tapes where wave height values (discernable from the log books) were not in the highest 10% for a gaging location. CERC discarded approximately 144 WV data tapes.

More detailed information is available in "Report on the 'Old' Data Management Improvement Program," September, 1983.

D. Data Catalog Files Created to Describe "Old" Data

NEWNAME - WV Tape characteristics in data collection format

FFGWVDB - WV Tape Index

E. Computer Programs Developed to Build and Manipulate Current Data Catalog Files

WAVEDB - COBOL program to edit and build the WV tape index

WAVEPR - COBOL program to List the WV Tape index

WAVSPEC - COBOL program to Print summary statistics

VI. Unautomated Data Study

A. Background

The objective of this final study was to classify the unautomated data into two groups; data worth retaining and data to be discarded. The criteria for retention was "unique, clean data which is readily identifiable". Non-automated data which duplicates automated data is not unique. CERC personnel participated in the classification decisions. The data selected for retention was indexed and cataloged.

The unautomated data consisted of paper tape rolls of analog wave data stored at Ft. Belvoir and at the Federal Records Center in Suitland, Maryland; Sand samples stored at Ft. Belvoir; Core Boring samples stored at the U.S. Geological Survey in Reston, Virginia; and Flat data (maps and charts) stored at Ft. Belvoir and at the Federal Records Center (F.R.C.) in Suitland, Maryland.

To facilitate field operations, data stored at Ft. Belvoir and Reston, Virginia was indexed and cataloged in Phase I and data stored at the F.R.C. was indexed and cataloged in Phase II.

B. Deliverables

The deliverables in the unautomated data study include all the data catalog files created to describe the unautomated data and all the utility computer programs written to build these files.

More detailed information is available in "Report on the Indexing and Cataloging of Unautomated Scientific Littoral Data", September, 1983.

C. Data Catalog Files Created to Describe Unautomated Data

These data files exist on disk at Boeing Computer Corp. and will be transferred to the Control Data Corporation facility in Rockville, MD.

NAME	DESCRIPTION
ALLCORE	USGS core borings index - 1239 records
CORINDX	Automated index to the USGS CERC core boring samples
FDIN	Input data file (1) - flat data index
FDINDX	Flat data holdings automated index
FDIN2	Input data file (2) - flat data index
PTDAT	Edited paper tape index from on-site holdings (102 characters)

NAME	DESCRIPTION
PTDAT1	Edited paper tape index from on-site holdings (129 characters)
PTDAT2	Edited paper tape index from on-site holdings (129 characters)
PTDAT3	Edited paper tape index from on-site holdings (129 characters)
PTIN	Paper tape input data for on-site paper tapes
PTIN2	Paper tape input data for FRC stored rolls - 1st 1/2
PTIN3	Paper tape input data for FRC stored rolls - 2nd 1/2
PTINDX	Index to the combined paper tapes rolls 129 character record
SANDIN	Automated data 1st 1/2 sand samples index input
SANDIN2	Automated data 2nd 1/2 sand samples index input
SNDDAT	Interim sand data base 1st 1/2
SNDDAT2	Interim sand data base 2nd 1/2
SNDINDX	Sand samples index - sorted on location
S2000 Databases	
CORINDK	S2000 core borings index
FLTINDK	S2000 flat data holdings index

PTINDK	S2000 paper tape rolls index
SNDINDK	S2000 sand samples index

D. Computer Programs Developed to Build and Manipulate Unautomated Data Catalog Files

NAME	DESCRIPTION
CONCOR	Program to convert the CORES index (CORINDX) to S2000
CONCORB	Binary version of above
CONFLT	Program to convert the Flat Data index (FLTINDX) to S2000
CONFLTB	Binary version of above

NAME	DESCRIPTION
CONSND	Program to convert the Sand Samples index (SNDINDX) to S2000
CONSNDB	Binary version of above
CORLIST	Program to build the core index (CORINDX) and manual listing
CORSORT	Utility program to sort USGS core files to ALLCORE
FFGSORT	Utility program to sort the paper tape index
FFGSRTF	Utility program to sort the flat data index
FFGSRTS	Utility program to sort the sand index
FGSRTSB	Utility program to sort the sand index in box # order
FLATPGM	Program to edit the flat data and build the index
FLTLIST	Program to list the flat data index
LISTNUM	Program to print box numbers for the sand data storage
PROCFIL	S2000 procedure file from Bob Little of BCS
PTLIST	Program to list the manual index to the paper tape roll
PTPGM23	Program to edit & build the 129 char paper tape index
RTCORE	Route to batch - S2K conversion program for core index
RTCRLST	Route to batch - Cores index listing program

RTFD	Route to batch - FLATPGM - to build flat data index
RTFDLST	Route to batch - FLTLIST - to list the flat data index
RTFLAT	Route to batch - S2K conversion program for flat data index
RTLN	Route to batch - Program to print box #'s for sand samples
RTPTL	Route to batch - PTPGM to list the paper tape index
RTPT23	Route to batch - PTPGM23 to list the 129 char paper tape index
RTSAND	Route to batch - S2K conversion program for the sand index
RTSL	Route to batch - SNDLIST to list the sand index

NAME	DESCRIPTION
RTSND	Route to batch - SANDPGM to edit and build the sand index 1st 1/2
RTSAND2	Route to batch - SANDPGM to edit and build the index 2nd 1/2
RTS2K	Route to batch - S2K conversion program for the paper tape index
SANDPGM	Program to Edit sand data and build database
SNDLIST	Program to list the sand samples index
SNDLSTB	Binary verison of above
S2KCORE	Define program for cores S2K index
S2KFLAT	Define program for Flat data S2K index
S2KIN	Program to convert the paper tape index to S2K
S2KINB	Binary version of above
S2KPT	Define program for paper tape S2K index
S2KSAND	Define program for sand samples S2K index

NBS-114A (REV. 2-80)

U.S. DEPT. OF COMM. **BIBLIOGRAPHIC DATA SHEET** *(See instructions)*	1. PUBLICATION OR REPORT NO. NBSIR 83-2806	2. Performing Organ. Report No.	3. Publication Date December 1983

4. TITLE AND SUBTITLE

Scientific Littoral Data Management Study - Summary Report

5. AUTHOR(S)

Ralph E. Schofer and Franklin F. Goodyear

6. PERFORMING ORGANIZATION *(If joint or other than NBS, see instructions)*

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

7. Contract/Grant No.

8. Type of Report & Period Covered

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS *(Street, City, State, ZIP)*

Coastal Engineering Research Center
U.S. Army Corps of Engineers
Ft. Belvoir, VA

10. SUPPLEMENTARY NOTES

☐ Document describes a computer program; SF-185, FIPS Software Summary, is attached.

11. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here)*

This summary report describes a review of management procedures for scientific littoral data, and lists recommendations for their improvement. Recommendations include both good general management practice and specific detailed actions. Indexing and cataloging of data are recommended in order to obtain broader data utilization and to remove the dependence of institutional memory on specific individuals. Development of a methodology for identifying and releasing obsolete data is detailed.

A list is provided of data catalog files, which were created to describe unautomated data, and of the computer programs which were developed to build and manipulate these catalog files.

12. KEY WORDS *(Six to twelve entries; alphabetical order; capitalize only proper names; and separate key words by semicolons)*

data bank management system; data management; data management system; littoral data

13. AVAILABILITY

☒ Unlimited
☐ For Official Distribution. Do Not Release to NTIS
☐ Order From Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402.
☒ Order From National Technical Information Service (NTIS), Springfield, VA. 22161

14. NO. OF PRINTED PAGES 14

15. Price $7.00

USCOMM-DC 6043-P80