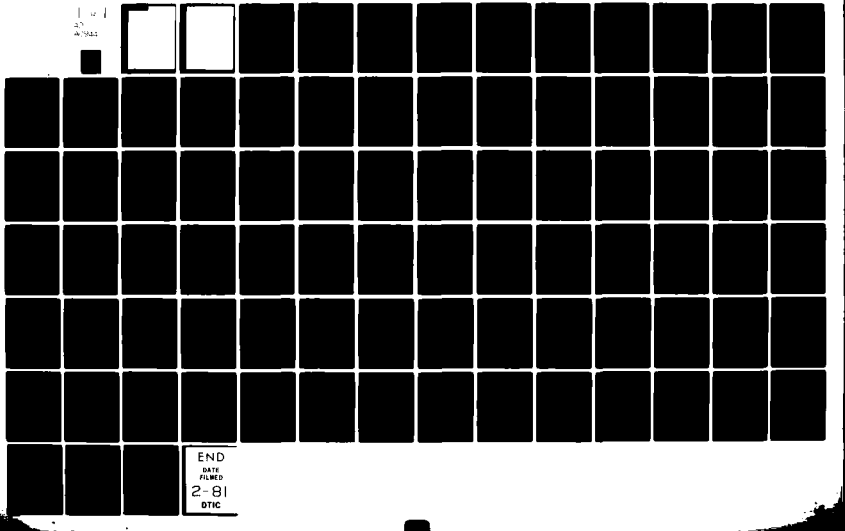
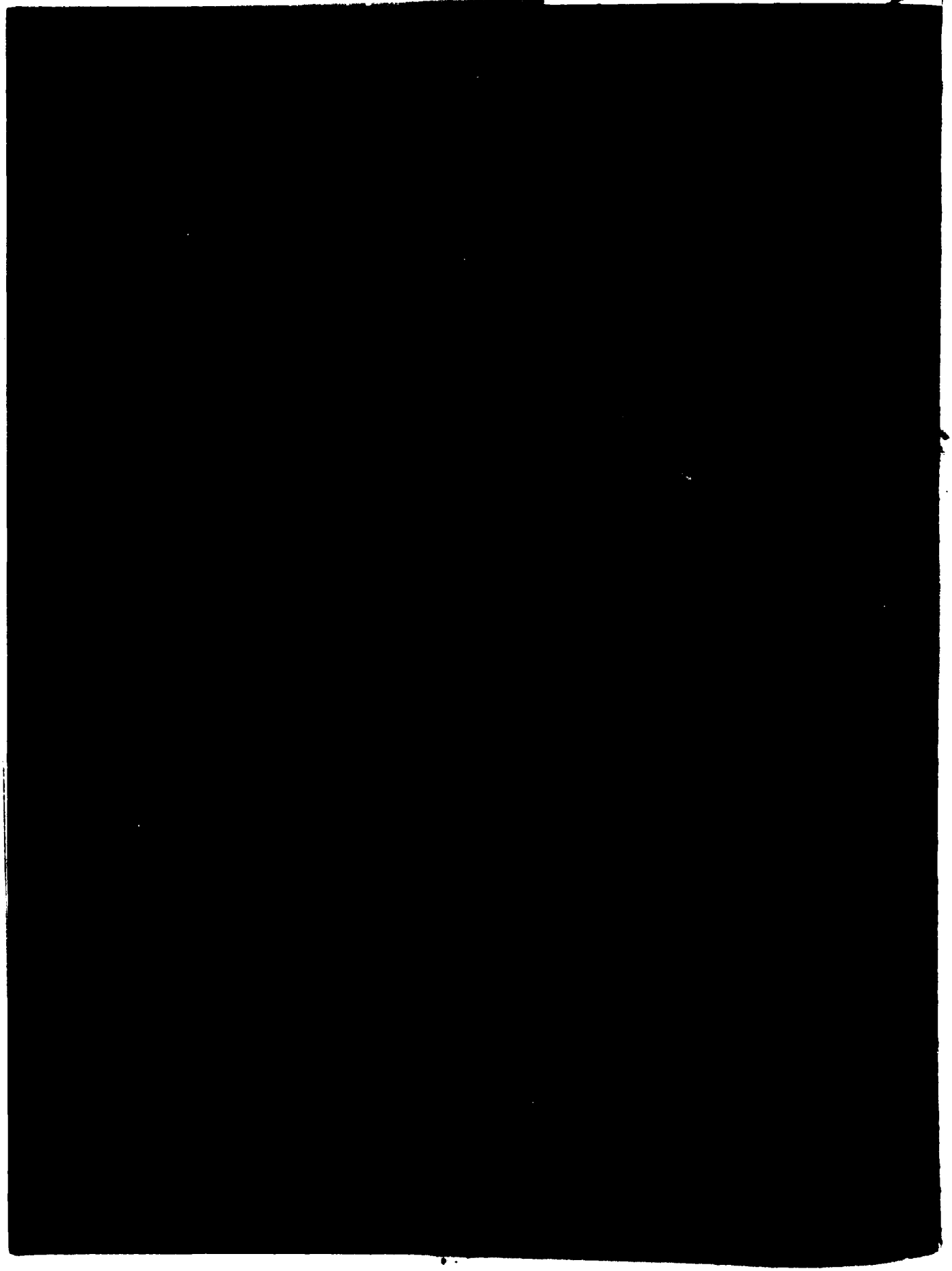
AD-A094 421 ARMY ENGINEER WATERWAYS EXPERIMENT STATION VICKSBURG--ETC F/G 5/2
MANAGEMENT OF WATER QUALITY DATA WITHIN THE CORPS OF ENGINEERS,--ETC(U)
DEC 80 M H SMITH, P L DOIRON
UNCLASSIFIED WES/MP/E-80-2 NL





Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Miscellaneous Paper E-80-2	2. GOVT ACCESSION NO.: AD-A094 421
3. RECIPIENT'S CATALOG NUMBER:	
4. TITLE (and Subtitle): MANAGEMENT OF WATER QUALITY DATA WITHIN THE CORPS OF ENGINEERS, 1979	5. TYPE OF REPORT & PERIOD COVERED: Final report
	6. PERFORMING ORG. REPORT NUMBER:
7. AUTHOR(s): Margaret H. Smith Phillip L. Doiron	8. CONTRACT OR GRANT NUMBER(s):
9. PERFORMING ORGANIZATION NAME AND ADDRESS: U. S. Army Engineer Waterways Experiment Station Environmental Laboratory P. O. Box 631, Vicksburg, Miss. 39180	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit No. 31608 (IVB), EWQOS Task IVB.1
11. CONTROLLING OFFICE NAME AND ADDRESS: Office, Chief of Engineers, U. S. Army Washington, D. C. 20314	12. REPORT DATE: December 1980
	13. NUMBER OF PAGES: 81
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Data collection systems
Water quality control
Water quality management

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report contains the results of a survey of current practices, needs, and recommendations in water quality data management by the Corps of Engineers Districts and Divisions. Reviews of the data management systems in use and others with potential for use by the Corps are included. Also included are recommendations for a system selection procedure for consideration in the event a Corps-wide data management system is implemented.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

PREFACE

This study (Work Unit No. 31608 (IVB)) was performed for the Environmental and Water Quality Operational Studies (EWQOS) Program, sponsored by the Office, Chief of Engineers (OCE). This study (EWQOS Task IVB.1) forms part of EWQOS Work Unit IVB, Data Management and Indices for Environmental Assessment.

The work was begun in September 1977 by Mr. P. L. Doiron, Environmental Laboratory (EL), U. S. Army Engineer Waterways Experiment Station (WES), under the direct supervision of Dr. T. D. Wright (EL), and the general supervision of Dr. J. L. Mahlock, Program Manager of EWQOS. The study was continued and this report produced in FY 79 by Mr. Doiron and Ms. M. H. Smith under the direct supervision of Mr. J. K. Stoll, Chief, Environmental Assessment Group, and the general supervision of Dr. C. J. Kirby, Chief, Environmental Resources Division. During the period of this study and report preparation, Dr. John Harrison was Chief of EL.

COL J. L. Cannon, CE, and COL N. P. Conover, CE, served as Commanders and Directors of WES and Mr. F. R. Brown was Technical Director during this time.

CONTENTS

MANAGEMENT OF WATER QUALITY DATA WITHIN THE CORPS OF ENGINEERS, 1979

PART I: INTRODUCTION

Background

1. The organization and management of data are critical elements of Corps of Engineers (CE) Districts and Divisions activities. These elements are particularly critical for water quality data collected by CE Districts and Divisions in compliance with public laws to ensure clean water in the Nation. For two decades, since 1950, Federal legislation, in accordance with public policy at those times, delegated the lead in prevention and control of water pollution to the states with the Federal role limited to assistance to the states. For example, the 1948 legislation, known as the Water Pollution Control Act - Public Law (PL) 845, assigned powers for enforcement to Governors of the states with Federal agencies authorized only to support research in water pollution projects in new technology and to assist the financing of treatment plants with limited loans. This law was subsequently amended to increase and broaden the authority of the Federal Government for research and technical assistance to the states.

2. In 1965 significant changes were made in the law when Congress approved a second set of major legislation for the water pollution control program. Among these decisions were the assignment of new responsibilities to the states, the continued use of a 1948 enforcement procedure, and the establishment of a new agency to administer the Federal portion of the program. Each state was required to develop standards for water quality within its boundaries which were to be submitted to the new Federal agency by 1 July 1967 for approval. The new agency was originally located in the Department of Health, Education, and Welfare; later it was transferred to the Executive Branch, first in May 1966 to the Secretary of the Interior and finally in December 1970 to the Administrator of the Environmental Protection Agency (EPA).

3. In 1972 the Federal Government was given more active

participation in water quality control when Congress passed PL 92-500, Federal Water Pollution Control Act Amendments of 1972. Under these amendments the basis of pollution prevention and elimination is the application of effluent limitations with water quality as the measure of program effectiveness and performance. Among the provisions of the law was the restatement of the permit system. The permit system, initiated in Section 13 of the 1899 Refuse Act, as restated in PL 92-500, prohibits the discharge of pollutants into navigable waters unless the permit procedure has been followed and a legitimate permit has been issued. Further, each state is required to adopt methods for control of pollution in freshwater lakes within the states and to restore their water quality. The enforcement of the provisions of the law is assigned to the EPA Administrator.

4. The most recent revision of the Federal Water Pollution Control Act was PL 95-217 which was passed in December 1977 and cited as the Clean Water Act of 1977. Among the revisions defined in Section 62 under the subtitle of Clean Lakes, the EPA Administrator is directed to issue information biennially on methods, procedures, and processes as may be appropriate to restore and enhance the quality of the Nation's publicly owned freshwater lakes.

5. The Office, Chief of Engineers (OCE), has responded to these regulatory requirements with Engineer Regulation ER 1130 and its revisions stating applicability, basic policies, reporting requirements, objectives, water quality management program reports, and data management. ER 1130-2-334, 16 December 1977, specifies in paragraph 4 (Appendix A) that, although Section 102b(2)(3) of the Federal Water Pollution Control Act Amendments of 1972 places the responsibility with EPA for water quality control in any reservoir project not in a construction status as of 18 October 1972, the responsibility for water quality management at all Corps projects clearly rests with the Corps. It further states that to meet this responsibility Division-wide Water Quality Management Programs must be established and to ensure success of this effort continual collection and evaluation of water quality data and reporting of water quality management activities are necessary. Field

offices of the CE Districts must evaluate water quality data management activities in accordance with the provisions of the Federal water quality laws and acceptable standards in their local and state areas and report them in accordance with ER 1130-2-334. Information on useful statistical analysis techniques is needed to allow the most accurate evaluation of data for preparing water quality monitoring reports and environmental assessments. The establishment or use of a current data base with reliable analyses, easy access, storage, and retrieval is needed to comply with current CE regulations. These regulations recommend the use of acceptable data management systems in conducting environmental inventories and assessments, in preparing environmental impact statements (EIS's), and in reporting water quality management activities.

6. The CE regulation ER 1130-2-334 (1 May 74) with an appendix related to water quality efforts specified the basic policies of the Corps with respect to its responsibility for water quality management, i.e., that the CE must evaluate and report water quality management activities annually on all its Civil Works projects. It defined the format of the report and listed specific physical-chemical, bacteriological, and biological parameters for inclusion in the report. The physical-chemical parameters were flow, turbidity, pH, temperature, conductivity, dissolved oxygen, and climatic conditions. Bacteriological parameters were total coliform and fecal coliform. Biological parameters were plankton population, benthos, and nektonic forms. Information on the current practices of collection and storage of these parameters is needed.

7. An update of ER 1130-2-334 was issued 16 Dec 77 (Appendix A) entitled "Reporting of Water-Quality Management Activities at Corps Civil Works Projects." It established reporting requirements and objectives for water quality programs at existing Corps Civil Works reservoir projects. It specifies the necessity for continual collection and evaluation of water quality data and the reporting of water quality management activities from Divisions to OCE and from OCE back to Divisions for Corps-wide water quality management activities. Paragraph 9 of the ER addresses information systems and states that the use of available

information systems should by pursued and that STORET (the storage and retrieval system developed by EPA) should be used unless extenuating circumstances dictate otherwise.

8. Presently, CE Divisions and Districts analyze, store, and retrieve their water quality data in a variety of ways. Some use STORET; some use the U. S. Geological Survey (USGS) system, WATSTORE; others have developed their own systems as the Ohio River Division (ORD) has done with AURAS; and still others are using combinations of existing systems or using only manual analysis and storage. A study was included in the Environmental and Water Quality Operational Studies (EWQOS) Program to evaluate existing data systems to find an acceptable system for CE use. A first step in the study (this work unit) was to determine present practices of the CE Divisions and Districts and to evaluate the systems currently in use.

Objective and Scope

9. The objective of this work unit was to survey all CE Divisions and Districts in the continental United States (except the Pacific Ocean Division) to determine their present practices in the use of information systems, to identify the data parameters collected, to evaluate the data management and storage systems used, and to identify needs as expressed by District and/or Division personnel. The survey was extended beyond the objective to include recommendations from CE personnel.

Approach

10. A review of the May 1974 and December 1977 ER's establishing guidelines for water quality management and requirements for reporting water quality management activities was made to determine the scope of storage and retrieval needs. A questionnaire survey by telephone of Districts and Divisions was conducted by personnel of the Environmental Laboratory (EL), U. S. Army Engineer Waterways Experiment Stations (WES). The responses were summarized as to practice, needs not met, and

recommendations. Several data management and storage systems were evaluated and the results are reported in appendixes to this report. Conclusions were made concerning the data management and storage systems currently in use, and recommendations were made based on the needs expressed by District and Division personnel.

PART II: CURRENT DATA MANAGEMENT PRACTICES OF THE CORPS OF ENGINEERS DIVISIONS AND DISTRICTS

Reservoirs Under CE Management

11. Some idea of the magnitude of the prospective CE data management activities can be envisioned when it is realized that in 1974 the CE operated over 500 existing Civil Works reservoirs.* A conservative estimate is that 20 percent more reservoir projects were under construction or had been approved for construction. The size of this potential data management problem must be kept in mind as the present data management practices are assessed.

Survey of Data Management

12. Personnel of EL conducted a telephone survey of CE Division and District water quality data management and storage practices in FY 78 and again in FY 79. For the 1979 survey, a list of contacts was compiled for securing information on personnel responsible for reporting water quality control activities in the Districts and Divisions and a questionnaire was developed to obtain the required information. A survey procedure was developed so that the elicited information from CE personnel would be as similar in format as possible. A copy of the questionnaire appears in Table 1.

13. Although this study included only water quality data, questions 11 and 12 were added to obtain additional environmental data currently collected by CE Divisions and Districts.

Summary of Survey Information

District Offices

14. Table 2 contains the responses of Districts by Divisions. Thirty-seven Districts (36 plus the New England Division (NED)) in ten

---

* Office, Chief of Engineers, Department of the Army. 1975. National Program of Inspection of Dams, Vol 1, Washington, D. C.

Divisions were contacted. Thirty-four Districts were familiar with the ER requirements and 33 were currently making annual water quality reports as required. Twenty-three indicated they used a computerized system for water quality storage and retrieval. These systems are included in Table 3. Seventeen Districts analyze their data with an automated procedure.

15. The responses to question 6 on types of data collected are summarized in Table 4. Parameters listed under physical-chemical, bacteriological, and biological of Table 4 are the principal ones specified in ER 1130-2-334, 1 May 1974. Those under additional data of Table 4 are those data collected that were not on the survey list but were given in response to the request for other water quality data collected. Data were collected on schedules differing from District to District and from parameter to parameter. The bacteriological data were generally collected during the summer months only, varying from weekly to seasonally. In a few instances, data other than bacteriological were collected hourly; generally other data were collected three to six times a year, seasonally, or for the duration of a requested study on a special interval dictated by the study.

16. Of the 23 Districts not using STORET, 21 indicated they were familiar with it and 11 of those said they would use it in the future. Twenty-six Districts felt their present storage practice met their needs. The remainder indicated needs were not being met by their system. These needs are listed in Table 5.

17. Table 6 is a compilation of recommendations made by persons interviewed. STORET was recommended more often than any other system; however, five Districts were not happy with STORET. AURAS and WATSTORE were recommended by some who were familiar with these systems and who were presently using them. A discussion of these and other systems can be found in PART III and in the appendices.

18. Data for 18 environmental parameters other than those listed in question 6 are being collected by CE Districts. They are listed in Table 7. No systems other than those mentioned for water quality data were used in analyzing, storing, and retrieving these data.

19. Materials were received from nine of the thirteen Districts offering to send information related to their reporting process--these were data sets or copies of annual reports. In reviewing these materials, it was noted that a few additional parameters were collected such as chlorophyll, DDT, silver, and selenium. Information was most often reported in manual tables. Other methods of reporting were computer-produced tables, manual and computer-produced plots, and in a few cases by narrative only.

Division Offices

20. The CE Division Offices were surveyed using those parts of the questionnaire not related to the actual data gathering. Nine Offices were surveyed (the NED response was included with the Districts since NED performs both District and Division activities). Five of the nine indicated needs not met with their present system. Some of these are:

a. There is a need for more flexibility in the output formats from STORET.

b. Periodic evaluations of current data management tools and opportunities for "hands on" experience with current data management systems need to be made available to those involved in the reporting process.

c. A unified data management system for the Corps could make data available to other Districts and Divisions in a consistent format and on a more timely basis.

d. One Division expressed an immediate need (not met) for a system to store elutriate and watershed data.

e. One Division expressed a Corps-wide need for familiarization with the methodology of storing and retrieving data, of data transforms and statistical options.

21. Only one Division was completely satisfied with its system, AURAS. Three Divisions recommended modified versions of STORET for CE use. One recommended WATSTORE.

PART III: REVIEW OF STORAGE AND RETRIEVAL SYSTEMS

Comparison of General Characteristics

22. Four major systems for storing and retrieving data are presently in use by the CE Districts: STORET, WATSTORE, AURAS and SIR (Scientific Information Retrieval). STORET and WATSTORE have resident data bases maintained by EPA and USGS, respectively. AURAS is the only system developed by a CE Division. It was developed by ORD and its data base is maintained by ORD. SIR is used by the Portland District. It does not maintain a resident data base but has the program capability of giving a user the option to build his own data base, which is what the Portland District has done. The New Orleans District has developed its own analysis program, HEIS. HEIS does have limited data storage but, to take advantage of a large existing data base and for ease of retrieval, the New Orleans District stores its water quality data in STORET.

23. Three other systems, though not used by CE Districts, seem worthy of including in the review: UPGRADE, NAWDEX, and SAS. UPGRADE does maintain a resident data base; NAWDEX and SAS do not. NAWDEX maintains another kind of base, a water quality information base for locating water quality information and data. SAS's proponents prefer to call it a data management system; like SIR, it gives the user the option of building and maintaining his own data base through SAS data management.

24. Table 8 gives six general characteristics of the seven systems for comparisons:

a. Name and series of computer on which system operates.

b. Input data format.

(1) Card image - 80 characters in columnar form input by card or keyboard.

(2) Interactive - Prompting questions from the terminal and response from the user.

c. Resident data base.

d. Analysis of data.

(1) Statistical.

(a) Limited - mean, standard deviation, maximum and minimum, time analyses.

(b) Moderate - includes linear regressions, distributions, variance.

(c) Extensive - multiple regressions, correlations, statistical inference, probabilities, multi-factor analysis, and other high-level analyses.

(2) Graphical interpretation.

e. Output format.

(1) Tables.

(2) Plots.

f. User training.

(1) Required to operate system.

(2) Supplied by contractors' user service.

Also included in Table 8 are the names and addresses of the contracting agency for each system.

System Reviews

25. The following paragraphs contain brief summaries of the reviews made of each system. The reviews can be found in appendices as indicated.

STORET

26. This system was developed and is operated by EPA. It maintains a data base on IBM equipment. There are few restrictions on its use and no control over data entered into the data base. There are no reliability estimates. Users inexperienced in the use of computers and terminals may have some difficulty in using the system and should plan to obtain training. Though the data stored have no standards for acceptance, the system does allow for exclusive use of a user's data if he chooses to protect it from other additions or from other users. The large data base contains much data unrelated to CE water quality management activities. See Appendix B for details.

WATSTORE

27. WATSTORE was developed and is operated by the USGS. It is

restrictive in data entered into the system. The user must qualify for a registration number and follow specific rules for entering data into the data base. Though it is more difficult to gain user acceptance and use than STORET, users express more confidence in the reliability of the data. User training is required. See Appendix C for details.

AURAS

28. This system was developed by ORD and is used by all ORD Districts. They are very satisfied with the system and two Districts in other Divisions have plans to use it in the future. The system resides on the Computer Science Corporation (CSC) INFONET system and, according to some users, requires considerable training to use. Users outside ORD may use the system through INFONET (Sciences Corporation Information Network). See Appendix D for details.

SIR

29. SIR was developed by SIR, Inc., as a Scientific Information Retrieval System. SIR resides on commercial computer systems, such as Boeing Computer Services. SIR was designed with the scientific researcher in mind and is, therefore, geared to his interests. SIR's capabilities include storing and retrieving scientific data as well as being able to perform simple mathematical and statistical analysis. SIR is designed so that the data retrieved can be interfaced with the Statistical Package for Social Sciences (SPSS) or the Biomedical Data Package (BMDP). To help the user in the operation of SIR, the language used by the SIR system is very similiar to SPSS language. With the above-mentioned capabilities, SIR is capable of building data bases and analyzing the data contained in them. SIR, as it stands, does not have a resident data base. See Appendix E for details.

UPGRADE

30. UPGRADE was developed and is maintained by the Council of Environmental Quality. The system resides on a commercial computer system and is accessible only by authorized users. This system contains a resident data base that is accessed interactively. The interactive process consists of a series of English language questions and answers. A user would need no computer training to operate this system. The analysis

procedures in UPGRADE are extensive as UPGRADE has access to SAS as well as its own internal analysis techniques. Because of the above-mentioned capabilities, UPGRADE appears to be a very powerful and versatile system for use by the Corps. See Appendix F for details.

NAWDEX

31. NAWDEX is a system developed and maintained by the USGS. The system resides on the computer system at the USGS National Center, Reston, Virginia. NAWDEX does not have a resident water quality data base. The systems' primary function is to assist users of water data to identify, locate, and acquire needed data. Therefore, NAWDEX is an information system and not a water quality data management system. See Appendix G for details.

SAS

32. SAS is a system developed by SAS Institute, Inc. The system resides on IBM or IBM-compatible equipment. SAS does not have a resident data base but contains the necessary software to build one. This data base storage and retrieval software is very powerful and allows the user a great deal of flexibility. Besides the data management aspect of SAS, the system also contains an extensive collection of statistical routines for analyzing data. In order to use the above-mentioned features of SAS, language was designed to be used with SAS that is easy to learn and use. See Appendix H for details.

PART IV: CONCLUSIONS AND RECOMMENDATIONS

Conclusions

Present storage
and retrieval practices

33. The survey showed an awareness of the responsibility of the CE for the quality of water in all its Civil Works projects and for reporting the findings. Thirty-three of the thirty-seven Districts are already participating in the annual reporting system with twenty-three Districts using some kind of automated data analysis and/or data storage and retrieval.

34. There is very little uniformity in data management within the CE. Table 3 lists 10 different systems in use that vary from all manual to almost all computerized.

35. STORET is the most preferred system with thirteen Districts and three Divisions using it. Even though it was the most popular, none were completely satisfied with it. Modification of STORET specifically for the CE, with its own data base and maintained by the CE, deserves careful consideration as a Corps-wide system. This was recommended by several Districts.

36. AURAS, the automated system developed by ORD is used by all ORD Districts and will soon be used by two other Districts which implies user satisfaction with the system.

37. The increase in water quality management activities on over 500 CE Civil Works reservoirs will be accompanied by a demand for an efficient data management system to enable compliance with CE regulations.

Other available systems

38. The users of SIR, UPGRADE, and SAS who were contacted expressed satisfaction with their choice of systems. UPGRADE appears to be the most versatile and does maintain a data base. SAS has the most powerful and the greatest number of analysis options. SIR, which is similar to SAS, is presently used by the Portland District but will soon be replaced by AURAS.

Considerations for the future

39. There appear to be six alternatives for the CE use of a water quality data management system that would include a data base.

a. STORET could be upgraded to CE requirements.

b. An existing system (such as UPGRADE, SAS, or SIR) could be adapted to CE use and maintained by contract.

c. The CE could lease or buy an existing system and adapt it to its requirements and maintain it within the CE.

d. The CE could develop its own system.

e. Each District within the Corps could select or develop its own system and maintain it within the District as ORD has done with AURAS.

f. The CE could implement a system similar to AURAS which provides easy interfaces to data analysis systems like SAS and automatic bulk data transfer of historical information to the STORET system.

Recommendations

40. Before any decision as to whether the Corps should or should not consider adopting (or developing) a unified data management for Corps-wide use, it is highly recommended that the following research be conducted.

a. Investigate in detail the existing system candidates, STORET, UPGRADE, SAS, AURAS, SIR, and any other promising system for Corps-wide data management use.

(1) Visit with experienced users of the systems and observe actual operations.

(2) Develop a set of critique guidelines and maintain critique forms and files for fair and unbiased comparisons of the systems.

b. Investigate the future data management needs and personnel capabilities within all Districts (projected for the next 10 years) to assess the degree of participation to expect from Districts.

c. Derive a set of typical data storage and retrieval situations for testing on candidate systems.

d. Secure a panel of experienced, objective data management personnel from within the CE to recommend an appropriate system and procedures for consideration.

Table 1

Telephone Survey Questionnaire with Instructions to the Surveyor

---

General Format

Introduction

Purpose of Call

Survey Questions

Request copy of any materials related to reporting procedures as required by ER 1130-2-334, 16 Dec 1977, or copies of reports.

Request copy of environmental data management used if other than STORET or WATSTORE.

Notes: Division:_____ District:_____

Date:_____ Name:_____

Telephone No.:_____

Interview

Locate Proper Contact Person

Ask for the person responsible for the annual Water Quality Management Reports from the District Engineer to the Division Engineer if a contact name is not available.

(Continued)

Table 1 (Continued)

Introduction

This is (name), Environmental Laboratory, Waterways Experiment Station, Vicksburg, Miss. We are interested in the state of the art of data management systems used in management of water quality data by the Corps of Engineers Districts. This is an effort to look at present systems, to establish needs, and to solicit recommendations.

Purpose of Call

To do this we are conducting a telephone survey of about 10 to 12 questions that can be answered briefly or with a yes or no. I believe we can complete it in about 8 to 10 minutes and would like to do so now if you can help me at this time.

Yes ( ) No ( )

If No, when would you be available?

Called back (date):

Survey Questions

1. Are you familiar with ER 1130-2-334, 16 Dec 1977, on reporting water quality management activities at Corps Civil Works projects?

Yes ( ) No ( )

Comment:

2. Are you currently sending annual reports on your District's water quality management activities to your Division Engineer?

Yes ( ) No ( )

Comment:

(Continued)

Table 1 (Continued)

3. Do you presently use a computerized system for water quality data storage and retrieval?

Yes ( )

a. System Name: _____
b. Did you develop it? Yes ( ) No ( )

No ( )

4. How do you store your data?

Comment:

5. How do you analyze your data?

Manually ( ), Automated system ( ), Combination ( )

6. Do you collect the following data and at what intervals?

Parameter	Yes	Collection Interval: Day	Week	Month	Other (Specify)
a. *Physical-Chemical:*					
Flow					
Turbidity					
pH					
Temperature					
Conductivity					
Dissolved oxygen					
Climate condition					
Other					
b. Bacteriological:					
Coliform					
Fecal coliform					
Other					
c. Biological					
Plankton population					
Benthos					
Nektonic forms					
Other					

d. What other data does your system include?
Comment:

(Continued)

Table 1 (Concluded)

7a. (If STORET is not used) Are you familiar with the STORET storage and retrieval system?

Yes ( ) No ( )

What are your reasons for not using it?

Comment:

7b. Will you use it in the future?

Yes ( ) No ( )

Comment:

8. Does your present practice of storing and retrieving water quality data meet your needs adequately?

Yes ( ) No ( )

If no, what needs are not met?

Comment:

9. Do you have materials you can send to us evaluating, summarizing, or describing your system (other than STORET or WATSTORE).

Yes ( ) No ( )

Will you send it to us at:

(Give mailing address)

10. What recommendations do you have for a storage and retrieval system of water quality data that would be useful Corps-wide?

Comment:

11. What other environmental data do you collect?

Comment:

12. What automated systems, if any, do you use in managing this environmental data?

Comment:

Table 2

Responses to Questionnaire by Division

Question Number	Division (No. of Districts in the Division)										
	Missouri River (2)	Ohio River (4)	New England (-)	North Atlantic (4)	South Atlantic (5)	North Pacific (4)	North Central (5)	Southwestern (5)	South Pacific (3)	Lower Mississippi Valley (4)	Totals (37)
1	2*	4	1	3	5	4	3	5	3	4	34
2	2	4	1	3	3	4	5	4	3	4	33
3	2	4	1	1	3	2	4	3	0	3	23
4	See Table 3 for systems used										
5**	0	2	1	0	2	2	5	3	0	2	17
6	See Table 4 for water quality parameters										
7a	-†	4	-	3	3	3	1	2	2	3	21 of 24
7b	-	2	-	1	3	2	0	1	1	1	11
8	1	4	1	2	4	1	4	5	2	2	26
	See Table 5 for needs not met										
9	0	1	0	0	0	3	4	5	3	1	17
Copies Received	-	0	-	-	-	1	2	4	2	1	10
10	See Table 6 for recommendations										
11	See Table 7 for other environmental data										
12	STORET, AURAS, in-house minicomputer										

* Number of Districts responding affirmatively to the question.
** Number of Districts using automated analysis of data.
† - indicates the question does not apply.

Table 3

Storage Presently in Use by Divisions

System	Missouri River (2)	Ohio River (4)	New England (-)	North Atlantic (4)	South Atlantic (5)	North Pacific (4)	North Central (5)	Southwestern (5)	South Pacific (3)	Lower Mississippi Valley (4)	Totals (37)
STORET	2*	0	1	1	2	1**	3(1**)	3	0	1**	14
AURAS	0	4	0	0	0	0	0	0	0	0	4
WATSTORE	0	0	0	0	0	1**	1	0	1**	0	3
Data sheets in file cabinet	0	0	0	3	1	1	0	1	1	3	10
Data cards in file cabinet	0	0	0	0	1	1	1	0	0	0	3
Combination manual and automated	0	0	0	0	0	2	0	0	1	2	5
Own automated system	0	0	0	0	1	0	0	0	0	1**	2
Printed on microfiche	0	0	0	0	0	0	0	1	0	0	1
SIR (Scientific Information Retrieval)	0	0	0	0	0	1	0	0	0	0	1
Automated, by contractor	0	0	0	0	0	0	1	0	1	1**	3

* Number of Districts in the Division using the system.
** Automated system used in combination storage system.

Table 4

Data Presently Collected by Districts

Parameter	Responding No. of Districts	Collection Interval			
		Day	Week	Month	Other
Physical-Chemical					
Flow	28	5	-	6	17
Turbidity	33	2	2	8	21
pH	35	3	2	10	20
Temperature	34	5	3	9	17
Conductivity	33	3	2	9	19
Dissolved oxygen	33	3	2	11	17
Climate condition	17	4	3	2	8
Other	20	-*	-	1	19
Bacteriological					
Coliform	24	-	1	6	17
Fecal coliform	25	-	1	6	18
Other	3	-	-	-	3
Biological					
Plankton population	21	-	-	3	18
Benthos	16	-	-	2	14
Nektonic forms	9	-	-	-	9
Other	4	-	-	-	4

Additional Water Quality Data Collected by Districts

Physical and Chemical

- Alkalinity
- Aluminum
- Arsenic
- Carbon dioxide
- Chemical oxygen demand
- Chromium
- Copper
- Cyanide
- Dissolved calcium
- Fluoride
- Inorganic and organic nutrients
- Heavy metals
- Iron
- Lead
- Magnesium
- Mercury
- Nickel
- Nitrogen (nitrates)
- Nitrogen (nitrites)
- Other hardness such as boron
- Pesticides
- Phosphorus
- Potassium
- Polychlornated biphenyl(s)
- Silica
- Sodium
- Sulfate
- Suspended sediment
- Zinc

Bacteriological

- Fecal streptococcus

Biological

- Presence of algae bloom
- Biological oxygen demand

* - indicates no collection at this interval.

Table 5

Expressed Needs Not Met in Present Practices

From users of STORET:

- Procedure for entering data is cumbersome and confusing (for small amounts of data it is not worth the effort).
- Computer-trained personnel are not available to use the system.
- Biological data are not available in the data base (it is understood that this will be available before the end of 1979).
- More flexibility in storage and retrieval is needed.
- Statistical and graphics packages are not adequate.

From AURAS user:

- Formats for printouts need to be changed; however, several new report writer programs have been written for the AURAS system in response to user request.

From a University Contract user:

- Time-share access to the computer is needed.

From those manually storing their data:

- Data are available only to those familiar with the files and to personnel in the office collecting the data.
- Speed in data handling and data manipulation is needed.
- Present practice is not organized. There is a need for an efficient storage system with easy access and retrieval.
- An automated system is needed.
- A more sophisticated system is needed.

Table 6

Recommendations for Storage and Retrieval from Districts

Twelve Districts recommended STORET for use Corps-wide. One thought an adaptation to CE needs would be better. Another recommended a special STORET system for CE use only.

Three recommended the use of AURAS Corps-wide with one suggesting a completely automated system from collection through storage and retrieval.

Two recommended WATSTORE or a system patterned after WATSTORE.

Five recommended an automated system other than STORET. Comments were: STORET has too much unrelated material; STORET data are not dependable; there are not enough choices in statistical analyses in STORET.

One recommended a Corps-wide system with a centralized data base (i.e., library or collection place) not only for data but also for printed reports containing data.

One recommended a Corps-wide system capable of summarizing data in several ways including maximums, minimums, durations, etc.

One recommended information on sampling procedure, whatever system was used.

It was recommended that a system be developed that would not require computer program training for the user.

It was recommended that Districts get feedback on the annual reports submitted and that more guidance be given in the ER on water quality data management and reporting of activities.

Table 7

Other Environmental Data Collected by Districts

Data for Environmental Impact Assessment, e.g., catalogue of permit applications and actions in compliance with PL 404 B, recreation usage of reservoirs.
Archeological
Socioeconomical
Vegetation
Habitat
Land use
Fisheries
Wildlife
Land resource
Cover
Soils
Slope
Elevation
Watershed boundaries
Political boundaries
Elutriate on dredging sites
Sewage disposal
Endangered species

Table 8

Characteristics of Storage and Retrieval Systems Examined in this Study

System	Computer	Input		Resident Data Base	Analysis		Output		User Training		Contract Agency
		Card Image	Inter-active*		Statistical	Graphical Interpretation	Tables	Plots	Required to Operate System	Supplied by Users Service	
STORET	IBM 370/168	Yes	IB	Yes	Light	--	Yes	Yes	Yes	Yes	Environmental Protection Agency Monitoring & Data Support Division 401 M Street, S.W. Washington, D. C. 20460
WATSTORE	IBM 370/155	Yes	IB	Yes	Light	Yes	Yes	Yes	Yes	Yes	Chief Hydrologist USGS, National Center Mail Stop 409 Reston, VA 22092
AURAS	Univac 1108	Yes	IB	Yes	Light	Yes	Yes	Yes	Yes	Yes	U. S. Army Engineer Division, Ohio River ATTN: ORDED-W P. O. Box 1159 (550 Main St.) Cincinnati, OH 45201
SIR	CDC 7600	Yes	IB	Yes	Extensive	--	Yes	--	Yes	--	SIR, Inc. P. O. Box 1404 Evanston, IL 60204
UPGRADE	IBM 360, 370	--	IA	Yes	Extensive	Yes	Yes	Yes	--	Yes	UPGRADE User Support Council on Environmental Quality Executive Office of the President 722 Jackson Place Washington, D. C. 20006
NAWDEX	IBM 370/155	Yes	--	--	--	--	Yes	--	Yes	Yes	Chief Hydrologist USGS, WRD 421 National Center Reston, VA 22092
SAS	IBM 360,370 4331, Amdahl Itel, CDC Omega, Riad	Yes	IB	Yes	Extensive	Yes	Yes	Yes	Yes	Yes	SAS Institute, Inc. P. O. Box 10066 Raleigh, NC 27605

* IA - English language questions and answers.
IB - Computer-prompted questions and answers according to system specifications.

APPENDIX A: ENGINEER REGULATION ER 1130-2-334: REPORTING OF WATER QUALITY MANAGEMENT ACTIVITIES AT CORPS OF ENGINEERS CIVIL WORKS PROJECTS

1. The Corps of Engineers regulation included in this appendix was issued 16 Dec 1977 and supersedes ER 1130-2-334, 1 May 1974. Pertinent words and phrases are underscored.

DEPARTMENT OF THE ARMY
Office of the Chief of Engineers
Washington, D.C. 20314

ER 1130-2-334

DAEN-CWE-HY
DAEN-CWO-M

Regulation
No. 1130-2-334

16 December 1977

Project Operations
REPORTING OF WATER QUALITY MANAGEMENT ACTIVITIES
AT CORPS CIVIL WORKS PROJECTS
(Reports Control Symbol DAEN-CWE-15)

1. Purpose. This regulation establishes reporting requirements and objectives for water quality programs at existing Corps Civil Works reservoir projects.

2. Applicability. This regulation is applicable to all districts and divisions having Civil Works responsibilities.

3. References.

a. ER 15-2-10.

b. ER 1105-2-8.

c. ER 1110-1-8100.

d. ER 1110-2-240.

e. ER 1110-2-1150.

f. ER 1130-2-415.

g. "The Federal Water Pollution Control Act Amendments of 1972," Public Law 92-500, 18 October 1972, Government Printing Office (GPO), Washington, D. C. 20402.

h. Executive Order 11752, "Prevention, Control, and Abatement of Environmental Pollution at Federal Facilities," 19 December 1973, GPO.

4. Basic Policies. It is a stated National Policy (ref 3h) that the Federal Government, in the design, construction, management, operation, and maintenance of its facilities, shall provide leadership in the nationwide effort to protect and enhance the quality of our air, water and land resources. Section 102b (2)(3) of the Federal Water Pollution Control Act Amendments of 1972 (ref 3g) places responsibility with EPA for determination of the need for, the value of, and the impact of storage for water quality control in any reservoir project not in a construction status as of 18 October 1972. The responsibility for water

This regulation supersedes ER 1130-2-334, 1 May 1974

quality management at Corps Projects, however, clearly rests with the Corps since it is an integral part of Corps water control management activities. To meet this responsibility Division-wide Water Quality Management Programs must be established, specific water quality management objectives for each reservoir project must be developed, and procedures must be implemented to meet these objectives. To insure success of this effort, continual collection and evaluation of water quality data and reporting of water quality management activities are necessary.

5. Reporting Requirements. The magnitude of the Corps of Engineers water quality management activities makes effective coordination and reporting a necessity. In order that adequate information on Corps water quality management activities will be available to the Chief of Engineers, Division and District Water Control Management Elements and other interested parties, annual reports on Division and Corps-wide water quality efforts are required. Division Engineers will submit two copies of annual Water Quality Reports to HQDA (DAEN-CWE-HY) WASH DC 20314 by 1 February. An annual report on Corps-wide water quality management activities will be prepared by DAEN-CWE-HY and sent to division engineers by 1 April.

6. Reporting Objectives. To monitor water quality management activities, data must be collected, evaluated and presented in a usable form. A major objective of the annual reports is to establish the success or failure of the design and operation of each project relative to established water quality management objectives, thus providing engineering feedback to improve project design. Additional objectives include input for reservoir regulation activities, summarization of available data, definition of problems and appropriate solutions in a timely manner and facilitating coordination with other agencies. Reports submitted by division engineers to OCE will provide data to serve as a basis for development of technical and policy guidance, formulation of Corps-wide programs and monitoring of division water quality management efforts.

7. Division Reports.

a. General. The annual division Water Quality Report should be prepared in two separate parts as discussed below. The first portion should address the Division Water Quality Management Program. The second part should present specific project information. The Water Quality Report may be submitted along with the annual Division Water Control Management Report providing that the required 1 February submittal date for the Water Quality Report is met.

b. Water Quality Management Program. The first report submittal following issuance of this ER should clearly establish the goals and

objectives of each division's overall Water Quality Management Program. Elements in the program should be discussed in terms of both present status and projected activities. Items to be specifically addressed include technical capabilities in the division and district offices, relationships between water quality and water control management activities, contracted workload, pertinent division regulations, laboratory facilities, data management systems, training, coordination with other agencies, research and development needs, special studies completed or required and scheduling for any required detailed project evaluations. Subsequent reports should emphasize progress made toward meeting division-wide water quality management goals.

c. Specific Project Information. Basic information on all pertinent factors affecting water quality should be developed for each reservoir project. This information should include such items as, watershed characteristics, physical project elements affecting water quality (e.g. selective withdrawal facilities), project water quality management objectives, project regulation/operation required to meet objectives, and a description of data collection programs. In order to keep report length to a minimum and recognizing that this information is relatively stable, it may be presented in the next scheduled revision to the reservoir regulation manual. Annual reports should present a project by project summary of water quality conditions, special regulation activities, new or modified data collection programs, problems encountered at each project, plans to address identified problems, possible Corps-wide applications of available data (e.g. R&D), and ongoing applied research.

8. OCE Report. DAEN-CWE-HY will be responsible for evaluation of Division efforts and preparation of an annual report summarizing Corps-wide efforts and accomplishments. This report will discuss division activities during the previous year. In addition, information will be included on policy and technical guidance, status of relevant R&D efforts, Corps-wide training, newly issued ER's, ETL's and EM's, coordination with other agencies and a summary of Committee on Water Quality (ref 3a) activities.

9. Information Systems. In order to made water quality data available to all interested parties in a timely manner, the use of available information systems should be pursued.

a. STORET. STORET (Storage and Retrieval), the National Water Quality Surveillance System, administered and operated by EPA, provides the Corps with a means for contributing to the national water quality network and a source for water quality data. All divisions and districts should use STORET unless extenuating circumstances dictate otherwise.

b. WATSTORE. WATSTORE (National Water Data Storage and Retrieval System) is provided and maintained by the USGS. Water quality data which is reported on a daily basis can be obtained by direct access from the Daily Values File in this system.

FOR THE CHIEF OF ENGINEERS:

JAMES N. ELLIS
Colonel, Corps of Engineers
Executive Director, Engineer Staff

APPENDIX B: STORAGE AND RETRIEVAL OF WATER QUALITY DATA WITH STORET*

1. With the passage of the Federal Water Pollution Control Act Amendments of 1972, the Environmental Protection Agency (EPA) was given the task of monitoring the quality of water within the United States. Due to the amount of data to be collected and analyzed, the need for an automated data management system was apparent. To meet this need, EPA developed its own storage and retrieval system, STORET.

System Identification

System Title: Storage and Retrieval of Data Relating to the Quality of Water

Code Name: STORET

Writer: EPA

Organization: Environmental Protection Agency
Monitoring and Data Support Division
401 M Street, S.W.
Washington, D.C. 20460
Telephone: (202) 426-7792
FTS 426-7792

Availability: STORET is available for use by Federal, State, and local agencies. It is not available for sale or lease.

Abstract

2. STORET is an acronym used to identify the system. The system is composed of a centralized data base and associated software for storing and retrieving data on water quality, water quality standards, point sources of pollution, pollution-caused fish kills, waste abatement needs, implememtation schedules, and many other water quality related items. The system is used by Federal, State, and local water quality agencies to obtain data used in the solution of such problems as defining the causes and effects of water pollution, measuring compliance with

---

* U. S. Environmental Protection Agency. "STORET Users Handbook," Washington, D. C.
_____. "STORET Handbook," Vols 1 and 2, Washington, D. C.

water quality standards, checking the status of waste treatment plants, and determining pollution trends.

3. An evaluation of the system is as follows:

a. Advantages:

(1) STORET provides a central depository of water quality and related data collected by EPA and other agencies. Data are made available to all users unless restricted by the user when entered into the system. The STORET stored data, therefore, provide a much larger data base for decisionmaking than can be gathered by any single agency.

(2) The availability of data entered into the system may be restricted to a single user. Thus, a user may store, retrieve, and analyze sensitive data with no fear of unauthorized retrieval.

(3) Water quality data entered in STORET may be retrieved in raw form, i.e., the form in which it was entered.

(4) Water quality data may be retrieved from STORET in statistically analyzed form, in time-dependent form, by comparison of water quality to Federal-State standards, by plots of data, or by maps.

(5) Data entered in STORET are available for retrieval immediately.

b. Disadvantages:

(1) Use of STORET requires specialized advance training. This disadvantage is offset somewhat by an efficient user training program. Date and location of the training can be obtained from the user service and training group at area code (202) 426-7792.

(2) There are no standards or screening techniques used on the data to maintain the reliability of data in the base. The accuracy of data contained in STORET is solely dependent on the user entering the data.

Technical Description

Major elements of the system

4. The major elements of the system are entering, analyzing, storing, and retrieving data. The system provides users with a collection of related computer programs and program elements that are activated or executed by commands (control cards) entered by users from

computer terminals. The original programs of the system were developed for the storage, retrieval, and analysis of water quality data only.

5. As the STORET system grew, additional programs were added for the storing and retrieving of other types of data. Other data files within the system include the Municipal Waste Inventory File, the Fish Kill File, and the Contract Awards File. The Water Quality File (WQF) remains the most widely used file in the STORET system.

6. STORET resides on an IBM 370/168 and an IBM 3032, which are operated by a commercial time-sharing service company.

Procedure to store data

7. Several computer languages are used in the various parts of the STORET system. However, the use of the system does not require a knowledge of them nor experience with them, although experience in the use of a remote terminal would be helpful. For a typical card deck setup to store a data set, see Figure B1, which is followed by Table B1 describing the information on each card. The tables, figures, flowcharts, and descriptions are excerpts or summaries from the STORET user's handbook. The cards are entered from a card reader (in a format specified in the contact procedure) at the user's remote terminal or from a keyboard where a line of typed entry contains the information described in Table B1 for a card. Figure B2 illustrates the setup for multiple data sets.

Output data procedures

8. A WQF retrieval request is composed of one or more instructions which define the information to be retrieved and how the information is to be presented. These instructions are entered to the STORET system as 80-column punched cards via a card reader or as lines of information via a keyboard terminal. An instruction is composed of a keyword--which generally identifies the function that the keyword activates--and, in most instances, functions and/or constants associated with the keyword function. Figure B3 describes the flow of steps with applicable keywords in a typical data retrieval and is followed by Table B2 with an explanation of each step. An example is included at the end of the table.

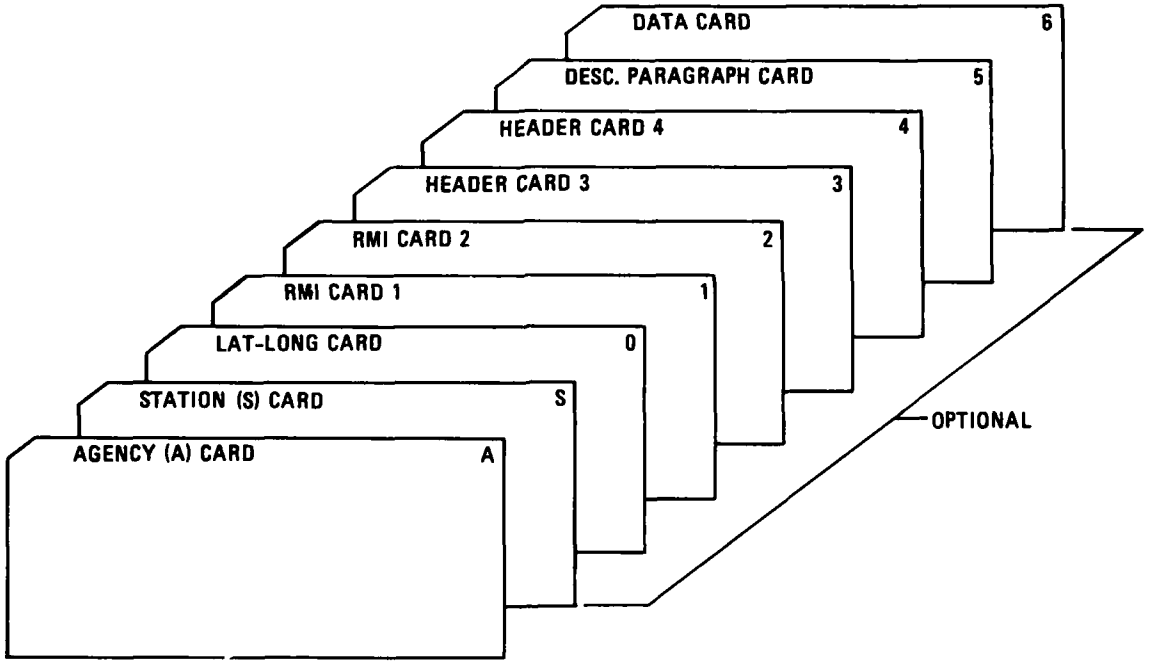


Figure B1. Storage deck setup

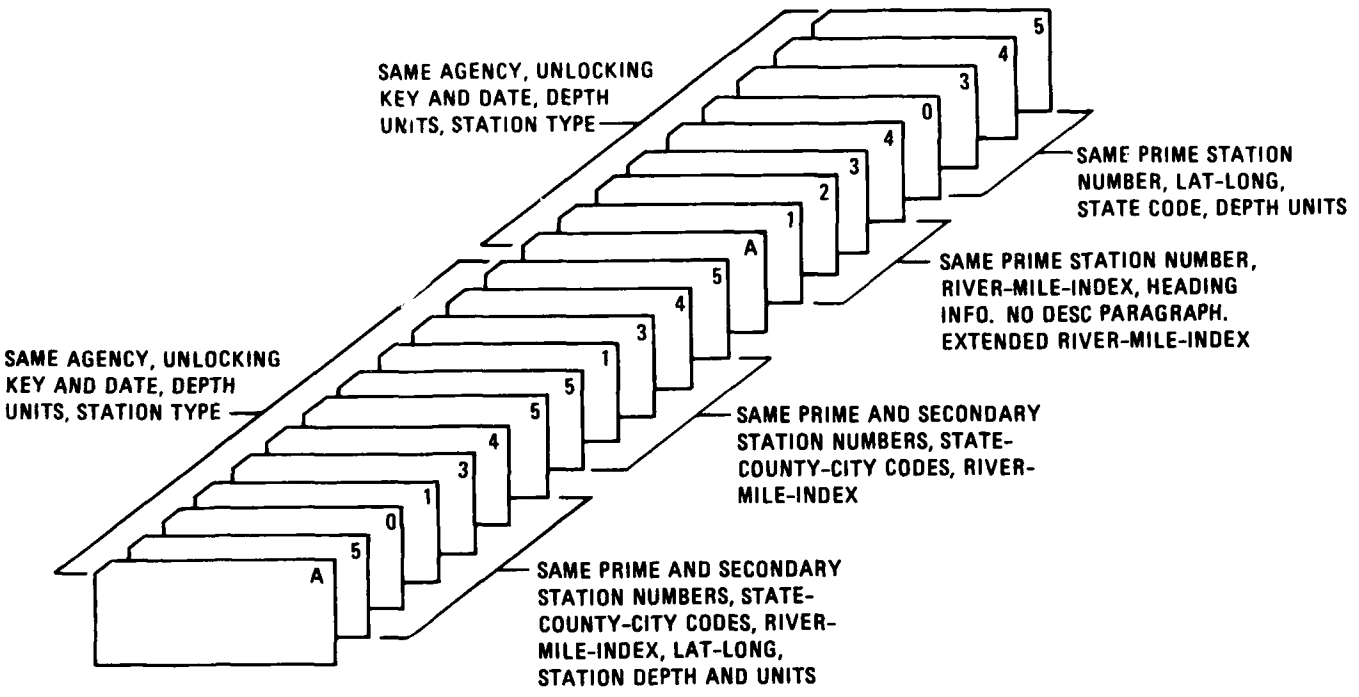


Figure B2. Arrangement of storage cards in sets

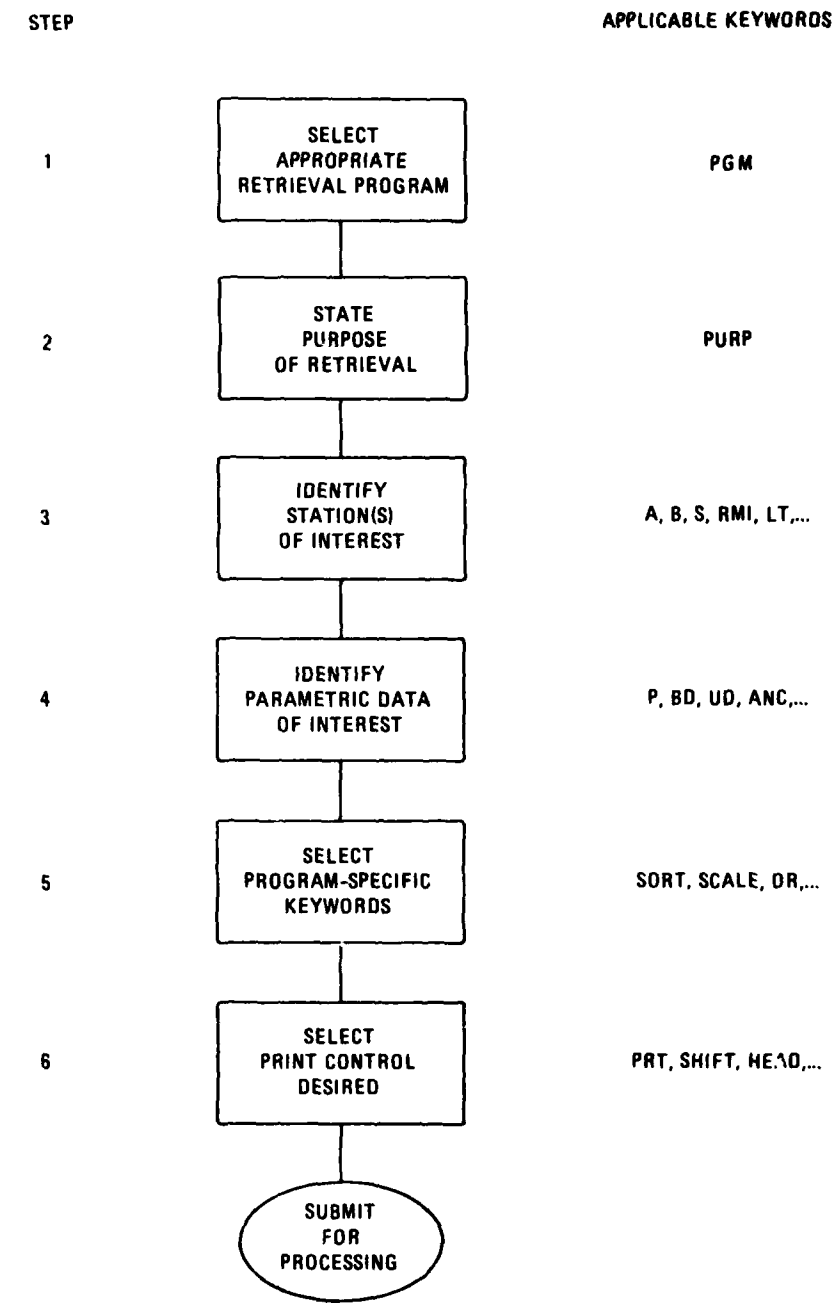


Figure B3. Flowchart of steps for data retrieval from STORET

Analyses available

9. The analysis of data on STORET is handled by the use of keywords during the retrieval phase as explained in Step 1 of Table B2. Analyses include summaries; tabulations of selected parameters; mean, standard deviation, and linear regression statistical analyses; and plots of statistical and time analyses.

Table B1

Explanation of Card Deck Setup for Storing Data

Card	Number of Cards	Information
Agency (A)	1	User code assigned by STORET operation. User name, location, and telephone number. Depth code, feet (F) or metres (M) at which sample is collected, store code, station type (required of all operations on STORET system).
Stations (S)	1	Specific information for location of each station.
Latitude-Longitude Header Card (0)	1	Station number, geographical coordinates of the station location, depth units, station depth, state code (optional).
River-Mile-Index Location Card (1,2)	1 or 2	Station number, hydrologic index, last level used. Hydrologic information up to 12 levels can be entered on the two cards.
Header Cards (3,4)	2	Heading information for the STORET printout.
Descriptive Paragraph (5)	1 to 11	Alpha-numeric characters (up to 792) are available to describe conditions at each collection station location.
Data Cards (6-)	Any number	Data recorded at station.

Table B2

Explanation of Steps in Data Retrieval

STEP 1: SELECT APPROPRIATE RETRIEVAL PROGRAM

The keyword PGM is used to select the appropriate WQF retrieval program. Some programs are appropriate to those whose interests are geographically oriented, while others are suited to those who are more interested in one or more water quality parameters, regardless of their specific geographical location.

The following list summarizes the primary functions of these programs and the type of information they provide:

Inventory of Parameters Sampled

INVENT	list summary information on parameters sampled (132-character line)
INV120	list summary information on parameters sampled (120-character line)

Listings of Actual Sample Values

RET	request tabular listings of selected parametric data
PUNCH	punch selected parametric data onto 80-column cards

Statistical Analysis of Raw Data

STAND	compare actual data against standards
MEAN	perform statistical calculations on selected data
REG	perform correlation and/or regression analyses on selected data
MSP	plot results of selected statistical analyses of data as functions of stations
PLOT	plot data as a function of time

Station Information Only

STA	list station codes for selected agencies
INDEX	list station header information
NOPAR	select stations meeting specified sampling criteria
LOC	plot a map showing station locations

The first instruction in a retrieval request must be the PGM keyword, whose value is one of the program names listed above, such as:

STEP 1	PGM=RET,	Request will provide tabular listings of selected parametric data.

(Continued)

Table B2 (Continued)

STEP 2: STATE PURPOSE OF RETRIEVAL

The PURP keyword, which identifies both the purpose and the ultimate recipient of the output of a retrieval, must be specified in each retrieval request and falls logically as the second step in building a retrieval. In most instances the purpose will be to satisfy some reporting requirement of Public Law (PL) 92-500 and the ultimate recipient will be a Federal agency, a State agency, an EPA regional office, or other user. The information provided by this keyword is used in various analyses of the uses made of the data stored in the WQF.

STEP 2	PMG=RET, PURP=305B/STA,	Request will satisfy requirements of Section 305B of PL 92-500.

STEP 3: IDENTIFY STATION(S) OF INTEREST

The WQF provides several station identification keywords used to define which of the more than 200,000 sites available within the WQF are of interest for a particular retrieval. The several basic ways that stations can be selected are listed below, along with the station identification keywords that are used to achieve the respective station selections.

	Station Retrieval Alternatives	Station Identification Keywords
1	all stations whose latitude/longitude coordinates are within a user-defined geographical area	LT, L
2	stations by agency code and/or station codes	A,S,B,
3	stations along a length of a waterway	RMI
4	stations within one or more states and/or counties	STC,CO

(Continued)

Table B2 (Continued)

STEP 3: (Continued)

	Station Retrieval Alternatives	Station Identification Keywords
5	stations by major/minor/subbasin	BS
6	stations within any designated state area-wide waste treatment management area as defined by Section 208 of PL 92-500	AREA
7	further screening of stations found within one of the above categories	U,ST, EXTRACT

STEP 3	PMG=RET, PURP=305B/STA, A=14AGNFS9, S=070009,	Request will select station 070009 owned and maintained by agency 14AGNFS9.

STEP 4: IDENTIFY PARAMETRIC DATA OF INTEREST

Data selection keywords enable a user to restrict the parametric data retrieved to those specific parameters, *specific sampling dates, sampling depths, and sampling* conditions of interest. There are over 1800 unique parameters defined within STORET.

STEP 4	PGM=RET, PURP=305B/STA, A=14AGNFS9, S=070009, BD=730101, P=11,P=300,	Request will select all values for temperature in degrees Fahrenheit (P=11) and dissolved oxygen in milligrams per litre (P=300) sampled since 1 January 1973 (BD=730101).

(Continued)

Table B2 (Continued)

STEP 5: SELECT PROGRAM-SPECIFIC KEYWORDS

Most of the WQF retrieval programs have one or more special keywords. For example, the PLOT program has a keyword that allows the user to choose symbols to identify plotted data values. This keyword would obviously not be applicable to programs whose function is to print station descriptions. DR indicates the number of decimals in the data as shown below.

STEP 5	PGM=RET, PURP=305B/STA, A=14AGNFS9, S=070009, BD=730101, P=11,P=300, DR=1,	Request will print values for parameter 300 as whole numbers.

STEP 6: SELECT PRINT CONTROL DESIRED

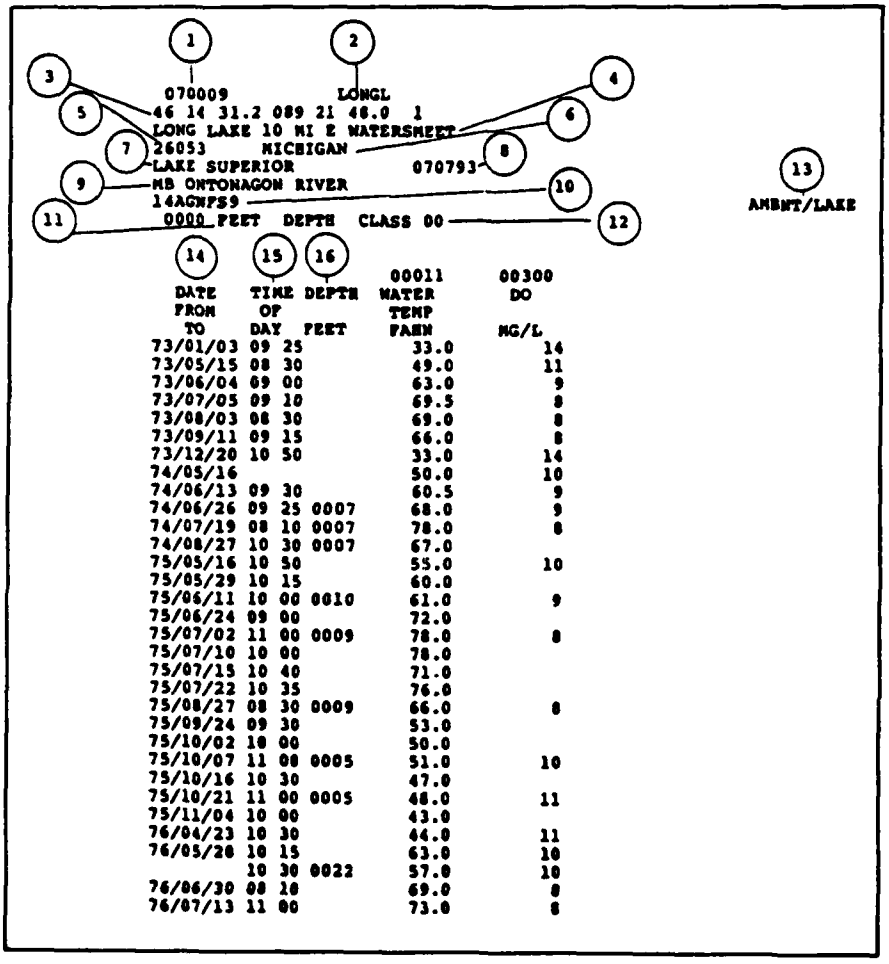
The final step in building a retrieval is to determine whether the standard print format for the chosen program is satisfactory and, if not, what changes should be made. The PRT, HEAD, SHIFT, and PRMI keywords allow the user some control over the content and format of the printed output.

STEP 6	PGM=RET, PURP=305B/STA, A=14AGNFS9, S=070009, BD=730101, P=11,P=300, DR=1, SHIFT,	SHIFT causes the station header information to be shifted from the upper right-hand corner to the upper left-hand corner of each page of the printed output.

(Continued)

Table B2 (Continued)

Once the retrieval has been set up and visually checked for omissions and coding errors, it is ready for submission to the STORET system for processing by the computer. The retrieval instructions chosen in the six basic steps can be keypunched onto cards and submitted via a card-reading terminal, or they can be keyed via a keyboard terminal. Computer output generated by the above retrieval is shown below.



DATE FROM TO	TIME OF DAY	DEPTH FEET	00011 WATER TEMP FAHN	00300 DO MG/L
73/01/03	09 25		33.0	14
73/05/15	08 30		49.0	11
73/06/04	09 00		63.0	9
73/07/05	09 10		69.5	8
73/08/03	08 30		69.0	8
73/09/11	09 15		66.0	8
73/12/20	10 50		33.0	14
74/05/16			50.0	10
74/06/13	09 30		60.5	9
74/06/26	09 25	0007	68.0	9
74/07/19	08 10	0007	78.0	8
74/08/27	10 30	0007	67.0	
75/05/16	10 50		55.0	10
75/05/29	10 15		60.0	
75/06/11	10 00	0010	61.0	9
75/06/24	09 00		72.0	
75/07/02	11 00	0009	78.0	8
75/07/10	10 00		78.0	
75/07/15	10 40		71.0	
75/07/22	10 35		76.0	
75/08/27	08 30	0009	66.0	8
75/09/24	09 30		53.0	
75/10/02	10 00		50.0	
75/10/07	11 00	0005	51.0	10
75/10/16	10 30		47.0	
75/10/21	11 00	0005	48.0	11
75/11/04	10 00		43.0	
76/04/23	10 30		44.0	11
76/05/28	10 15		63.0	10
	10 30	0022	57.0	10
76/06/30	08 10		69.0	8
76/07/13	11 00		73.0	8

(Continued)

Table B2 (Concluded)

1) Primary Station Code
2) Secondary Station Code
3) Latitude and Longitude
4) Station Location
5) State and County Codes
6) State
7) Major Basin
8) Basin Codes
9) Minor Basin
10) Agency Code
11) Depth Units
12) Archive Class (00=online)
13) Station Type
14) Date Sampled--required information stored for each sample. If no date appears on the printout, the date listed in the preceding line applies.
15) Time Sampled--optional information which may be stored for each sample.
16) Depth of Sample--optional information which may be stored for each sample.

APPENDIX C: NATIONAL WATER DATA STORAGE AND RETRIEVAL SYSTEM (WATSTORE)*

1. With the volume of data collected on the surface and subsurface water within the United States by the U. S. Geological Survey (USGS), it became apparent that in order to better organize and handle these data, a data management system was needed. The development of the National Water Data Storage and Retrieval System (WATSTORE) by USGS was the answer to this need.

System Identification

System Title: National Water Data Storage and Retrieval System

Code Name: WATSTORE

Writer: USGS

Organization: Chief Hydrologist
USGS, National Center
Mail Stop 409
Reston, Virginia 22092

Availability: WATSTORE is available for use by other Divisions of USGS as well as other Federal agencies who acquire and/or use water data. It is not available for sale or lease.

Abstract

2. The USGS Water Research Division investigates the occurrence, quantity, quality, distribution, and movement of the surface and subsurface waters that comprise the Nation's water resources. To support this effort, a large-scale computer system is used by the USGS to store and retrieve water data that have been acquired through its many activities. WATSTORE was implemented in November 1971. In addition to

---

* U. S. Geological Survey. 1976. "First Interim Report on Digital Geographic Data Handling Activities in the U. S. Geological Survey," Reston, Va.

its data processing, storage, and retrieval capability, WATSTORE has the capability of producing computer-printed tables, computer-printed graphs, statistical analyses of the water data, and plotter-produced maps. A flowchart of these program elements is presented in Figure C1.

3. An evaluation of the system is as follows:

   a. Advantages. WATSTORE provides a central depository of data dealing with the occurrence, quantity, quality, distribution, and movement of the surface and subsurface waters that comprise the Nation's water resources.

   b. Disadvantages. Training in the use of the system is required before attempting to use it. Experienced computer terminal-trained personnel would have to be used to operate the system.

Technical Description

Major elements of the system

4. The major elements of the system are entering, analyzing, storing, retrieving, and reporting the data. The data base of the system is comprised of several files--Peak Flow, Daily Values, Ground Water Site Inventory, and Station Header files.

Procedure to store data

5. Agency codes are assigned by the USGS and a request for a code assignment may be made by writing the address given in Organization under System Identification (page C1). Each user will be given a user's registration number and an account number for billing purposes. Once authorization is obtained, data on punched cards may be entered at the central computer in Reston, Virginia; at any USGS regional or district terminal; or at a selected user terminal, as long as the user terminal is capable of interfacing with an IBM 370/155 computer.

6. Before entering data into the system, an entry into the Station Header File must be made. The file contains identifiers that locate and describe the station from which data are to be entered. Since the identifiers are recorded in a separate file, a modification to the identifier will need to be made only once. The Station Header File also serves as an index to the data and the data's location. Through the Station Header

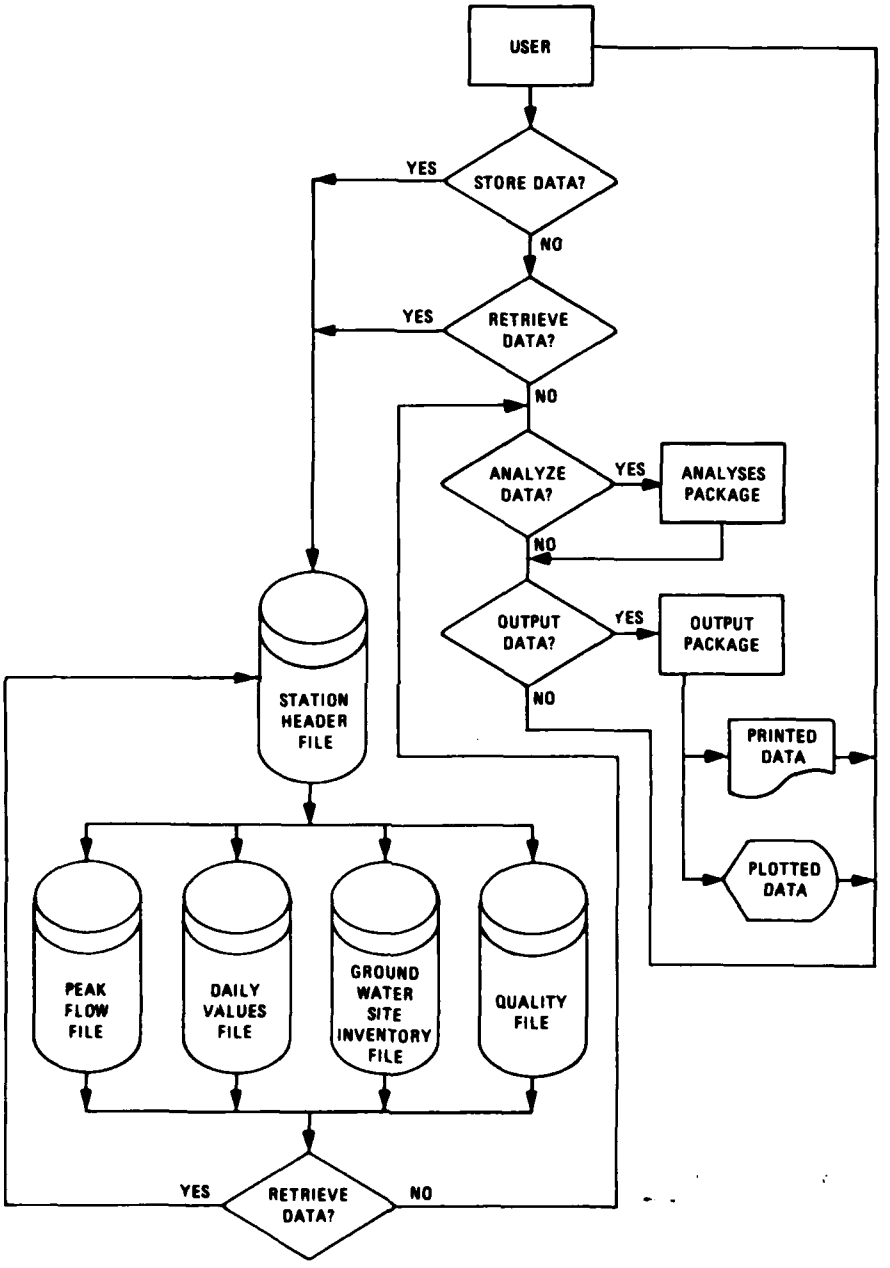


Figure C1. Flowchart of WATSTORE

File, the user has access to the other files within the system for storing or retrieving data.

Output data procedures

7. The retrieval procedures available permit: (a) listing of Station Header Files; (b) output of data on punch cards or magnetic tape; and (c) creation of temporary data sets for use with selected application programs such as line printer plotting programs, tabulation programs, and map plotting programs. The retrieval of data is handled by System 2000, a commercial data base management system.

8. Retrieval specifications may be either extremely simple or quite complex. The full list of retrieval categories is as follows:

a. Individual station.

b. Polygon of latitude-longitude.

c. State.

d. County.

e. Aquifer code (for groundwater sites).

f. Dates.

g. Individual parameters.

h. Boolean search procedures (greater than, equal to, less than).

Analysis available

9. Some statistical analysis software is available as well as graphic interpretation routines.

APPENDIX D: OHIO RIVER DIVISION'S DATA MANAGEMENT SYSTEM (AURAS)*

1. The system, developed within the Ohio River Division (ORD) for use by the Division in the handling of their water quality data, is known as the Automated Upward Reporting and Analysis System (AURAS).

System Identification

System Title: Automated Upward Reporting and Analysis System

Code Name: AURAS

Writer: Henry C. Jackson

Organization: U. S. Army Engineer Division, Ohio River
ATTN: ORDED-W
P. O. Box 1159 (550 Main Street)
Cincinnati, Ohio 45201

Availability: AURAS operates on the INFONET system and can be accessed through arrangements with ORD personnel.

Abstract

2. AURAS was developed as a means of systematically recording standard information about samples entering the laboratory, preparing bench sheets describing all samples logged into the system, recording test results, billing Districts for testing, preparing reports of completed and pending results, and providing management tools for efficient operation of the laboratory.

3. An evaluation of the system is as follows:

a. Advantages. This is the only operational data base management system developed by the Corps Offices and was developed specifically for the type of data collected and generated by Corps field tests. This makes it uniquely suited to the data management of other Divisions.

---

* System documentation in preparation.

b. Disadvantages:

(1) The large number of programs (Figures D1-D3) required to process the data make the system cumbersome to use.

(2) Computer user experience or training is required to operate this system.

Technical Description

4. The major elements of the system are data acquisition, data management, analysis and interpretation of the data, and application and upward reporting of the data. The system is basically divided into two parts: the lab part (Figure D1), and the INFONET part (Figure D2).

Data acquisition

5. Data acquisition consists of the recording of field measurements of water quality conditions; the collection of and subsequent laboratory analysis of physical, chemical, and biological samples; the recording of where and when measurements or samples were taken; and preliminary editing of recorded information concerning measurements and analyses.

Data management

6. Data management consists of the systematic editing and storage of water quality data; the retrieval of data for reports; plots and mathematical analyses; the transfer of data to other interested parties; the preparation of management information reports concerning the data; and the interfacing of the data with other systems such as SAS or STORET.

Analysis and interpretation

7. Analysis and interpretation consist of both automated and manual examinations of the data by such means as statistical summaries, plots, and raw data reports so as to form conclusions about the physical, chemical, and biological conditions represented by the data.

Application and upward reporting

8. Application and upward reporting consist of plan formulation for minimizing the effect of water quality problems identified through data interpretation; reservoir operations; and upward reporting of water quality information through reservoir regulation manuals, environmental impact statements, and annual reports.

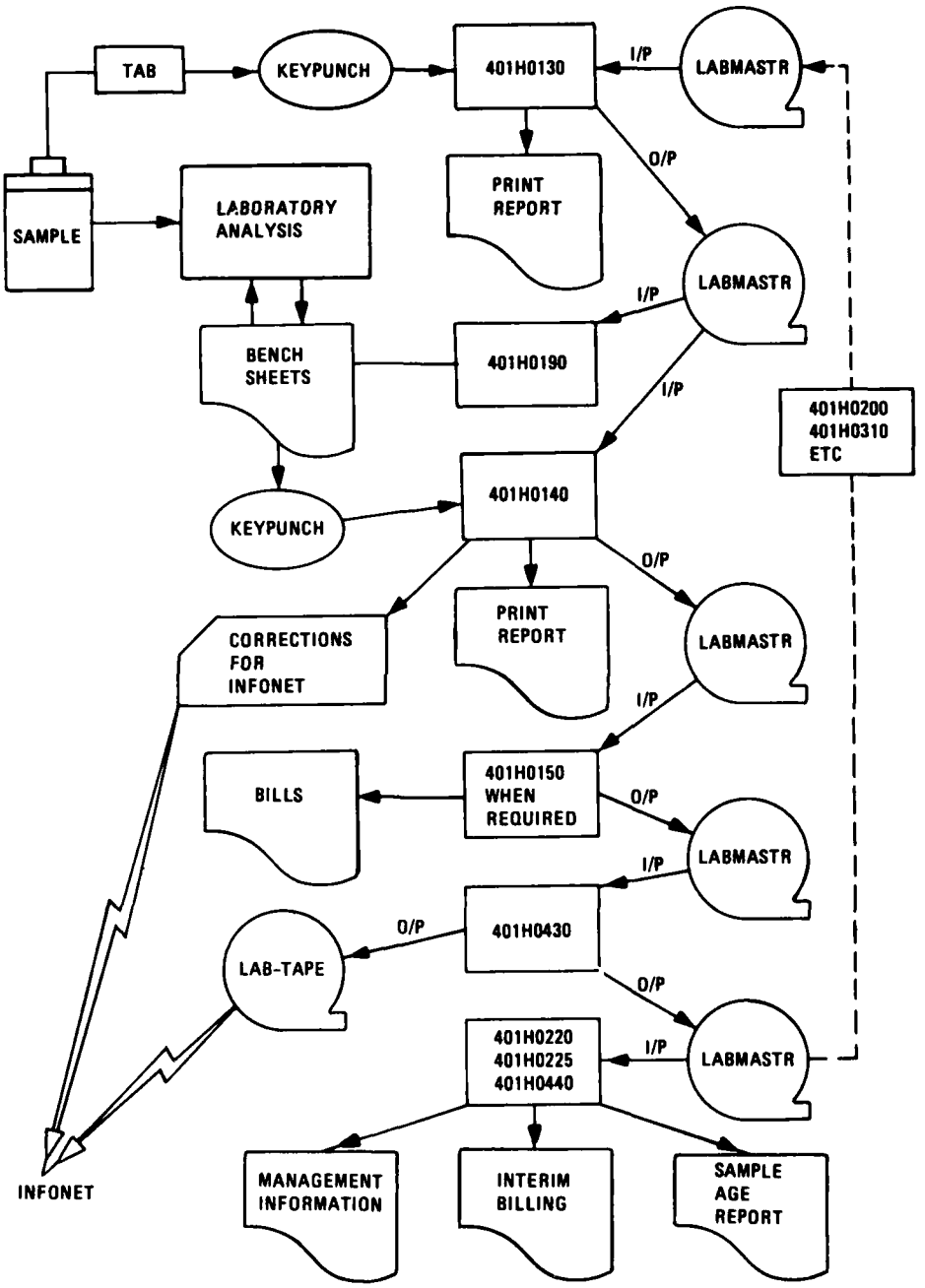


Figure D1. System flowchart of lab portion of AURAS

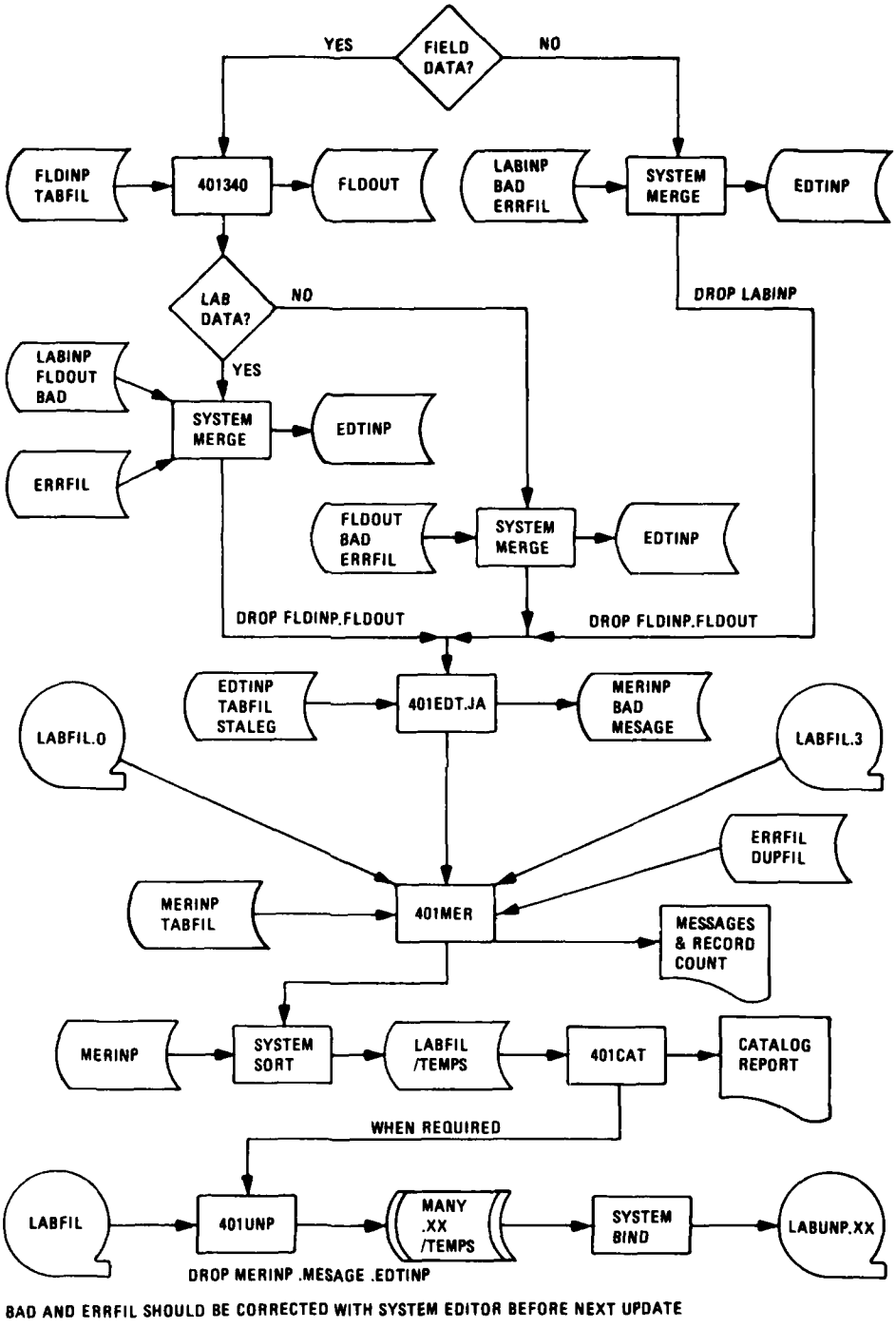


Figure D2. System flowchart for INFONET update

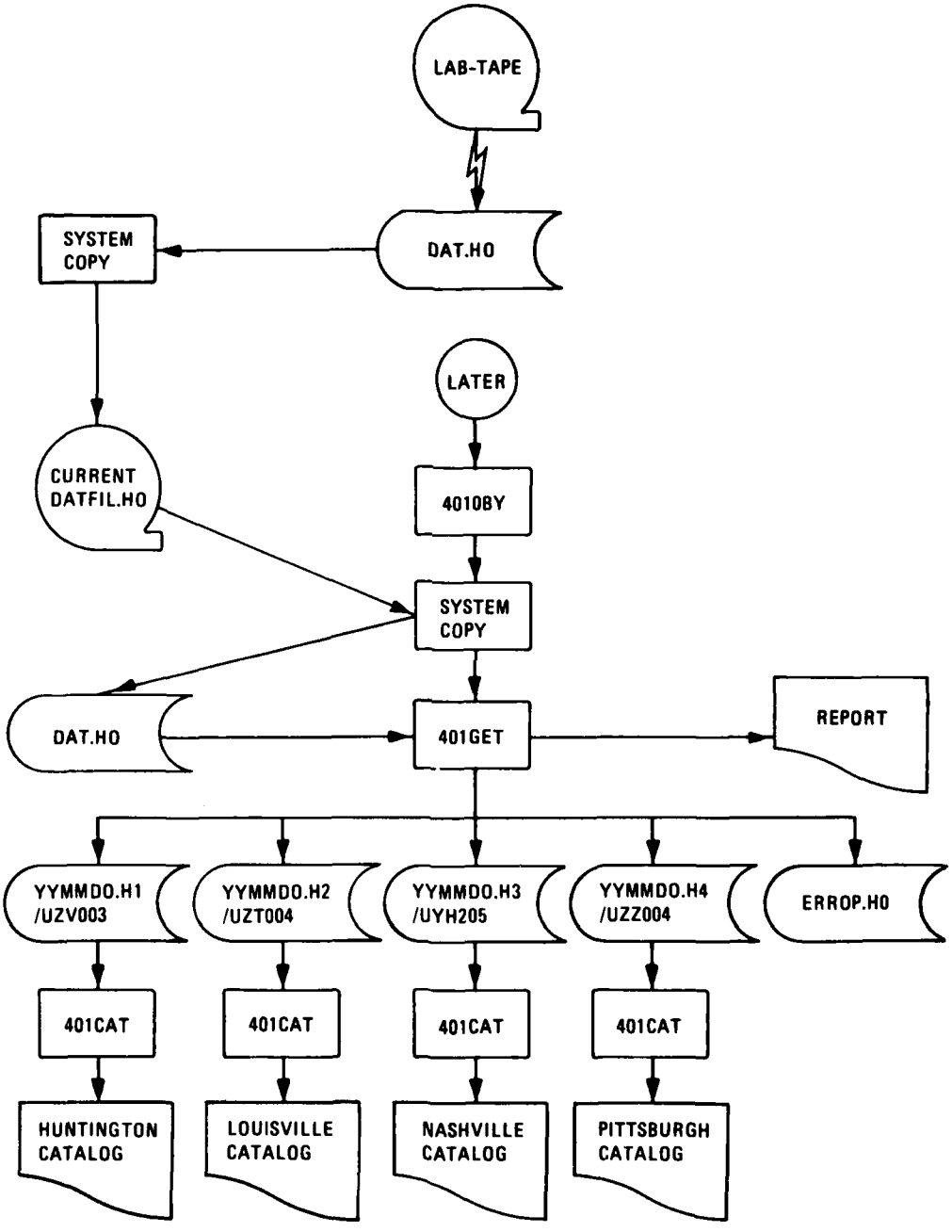


Figure D3. System flowchart for data transfer to INFONET

Table D1

Programs in LABMASTER

The LABMASTER system is currently supported by 12 programs on the G-437 for Division laboratory use and 14 programs on INFONET for storage and retrieval by Districts. Other programs are available, but not supported on the G-437. The programs are as follows:

G-437

401H0130	Laboratory Master File Addition
401H0140	Laboratory Master File Update
401H0150	Laboratory Master File Billing
401H0170	Laboratory Master File Station List
401H0190	Laboratory Master File Bench Sheet
401H0200	Laboratory Master File Station Correction
401H0210	Laboratory Master File Table
401H0220	Laboratory Master File Sample Age
401H0225	Laboratory Master File Management Information
401H0310	Laboratory Master File Station-Identification Correction
401H0430	Laboratory Master File INFONET Transfer
401H0440	Laboratory Master File Interim Bill

INFONET

401265	Laboratory Master File Sorted General Report
401340	Laboratory Master File Field Data Input
401CAT	Laboratory Master File Catalog

(Continued)

Table D1 (Concluded)

INFONET (Continued)

401EDT.JA	Laboratory Master File Pre-edit
401GET	Laboratory Master File G-437 to INFONET Transfer (Translate and Separate)
401GEX	Laboratory Master File G-437 to INFONET Correction Transfer
401LOD	Laboratory Master File System 2000 Data Base Load
401MER	Laboratory Master File Update and Merge
4010BY	Laboratory Master File Define and Equate Files
401SE1	Laboratory Master File Data Selection by Type
401SEL	Laboratory Master File Data Selection
401STD	Laboratory Master File Statistics
401SUB	Laboratory Master File Station-ID Selection
401UNP	Laboratory Master File Unpunctuated File Creation

APPENDIX E: SCIENTIFIC INFORMATION RETRIEVAL SYSTEM (SIR)*

1. The Scientific Information Retrieval (SIR) system was developed to fill the needs of scientific researchers.

System Identification

System Title: Scientific Information Retrieval

Code Name: SIR

Writer: Barry N. Robinson et al.

Organization: SIR, Inc.
P. O. Box 1404
Evanston, Indiana 60204

Availability: SIR is available through commercial data systems such as Boeing Computer Services, or can be leased for use on the user's own system.

Abstract

2. The rapid advancement of research technology has led to an increased gathering of data. This huge amount of information has, in many cases, a complex structure that has overwhelmed the researcher. In order to handle this amount of data, management systems have been developed. Until the development of SIR, most of the large integrated data base systems have been in the business area. SIR was developed for scientific researchers.

3. An evaluation of the system is as follows.

a. Advantages:

(1) SIR uses a language very similar to the language used by the Statistical Package for Social Sciences (SPSS); therefore, a person familiar with SPSS can use SIR without too much trouble.

---

* Robinson, B. N. et al. 1979. "Scientific Information Retrieval User's Manual," 2 ed., Printing and Duplicating Department of Northwestern University, Evanston, Ill.

(2) SIR interfaces with SPSS and the Biomedical Data Package (BMDP).

(3) SIR appears very flexible and powerful as a data management tool.

b. Disadvantages: No resident data base exists in the SIR system; therefore, one would have to be built by the user.

Technical Description

Major elements of the system

4. The major elements of the system are entering, manipulating, and retrieving scientific data. SIR is a hierarchical data management system. The user can establish hierarchical relationships between groups of data items, and SIR will use these relationships to store, manipulate, and retrieve data.

Procedure to store data

5. SIR provides the user with a variety of batch input commands. These commands enable the user to do the following:

a. Add new records to a data base.

b. Replace old records.

c. Update only certain variables within a record.

6. The SIR Interactive Editor enables the user to perform various SIR tasks interactively on a computer terminal. These actions are:

a. Input a set of SIR commands and correct and edit them.

b. Execute the commands.

c. Save the commands for later use.

d. Execute a previously defined set of commands.

e. Specify run-time parameters for a given run.

f. Retrieve data by use of the SIR prompts and acceptance of input during execution.

Output data procedures

7. The retrieval of data using SIR allows the user to extract data from one, some, or all the records belonging to each data analysis job or case. There are two types of retrieval available: "retrieval-by-case" and "retrieval-by-record."

8. Among the most important retrieval features are:

   a. An SPSS-like command syntax.

   b. A complete set of arithmetic and logical operations.

   c. Many built-in functions to search for data across a record or records.

   d. Built-in functions for performing mathematical and string manipulating operations.

9. With a retrieval task, the user can use the data in the summary records to:

   a. Perform simple statistical procedures.

   b. Create an SPSS or BMDP save file.

   c. Create a new SIR data base.

   d. Produce a complex hierarchical report.

Analysis available

10. SIR can perform simple mathematical calculations on data; however, the best analysis results are obtained by the creation of a data file that can interface with SPSS and BMDP for the actual analysis.

APPENDIX F: COUNCIL OF ENVIRONMENTAL QUALITY'S DATA BASE MANAGEMENT SYSTEM (UPGRADE)

1. The User-Prompted Graphic Data Evaluation (UPGRADE) system is a versatile system for analyzing computerized data on the environment, natural resources, public health, and related topics. It employs ordinary English language instructions, step-by-step analysis, and graphic display. UPGRADE is designed for efficient use by managers and scientists without computer training.

System Identification

System Title: User-Prompted Graphic Data Evaluation

Code Name: UPGRADE

Writer: President's Council on Environmental Quality (CEQ)

Organization: UPGRADE User Support
Council on Environmental Quality
Executive Office of the President
722 Jackson Place
Washington, D.C. 20006

Availability: UPGRADE is available for use by all authorized users. UPGRADE resides on a commercial computer system.

Abstract

2. CEQ is responsible for overviewing information dealing with the protection of the environment, safeguarding human health, managing natural resources, and working with other agencies to improve the quality, timeliness, and usefulness of environmental and technical data analyses. CEQ developed the UPGRADE system for several purposes: (a) to provide easier access to computerized environmental data; (b) to make these data available to a larger portion of the Nation's scientists and managers; (c) to facilitate more efficient and convenient environmental assessments; (d) to increase uses for available environmental data; (e) to provide better capabilities for identifying correlations between

factors represented in different computerized data banks; and (f) to improve environmental research and data collection programs through the insights and feedback provided by users of the UPGRADE system.

3. An evaluation of the system is as follows.

a. Advanatges:

(1) Anyone who understands the data and principles of analysis can use UPGRADE. No computer training is needed.

(2) The system is very flexible in the type and usage of data that require graphic display and/or statistical analysis.

b. Disadvantages: The only type of terminal that can access the system is a Tektronix 4014 CRT or equivalent terminal with a Tektronix 4631 hard-copy unit.

Technical Description

Major elements of the system

4. The UPGRADE system is divided into three major sections: Operational Mode Selection, Data Selection, and Graphic and Statistical Analysis specification. The system is described by flowchart as shown in Figure F1.

5. The Operational Mode Selection consists of three modes available to the user:

a. Verbose (01), which supplies the user with detailed explanations of prompts and response options throughout the session.

b. Terse (02), which greatly cuts down on the amount of explanatory material supplied with each prompt.

c. Superterse (03), which allows the user to profile a terminal session by specifying a set of responses to prompts he wishes to maintain throughout the session.

6. In the Data Selection portion, the user will specify the exact portion of the data upon which he wishes to perform analysis. The end product of data selection will consist of variables to be entered into the Graphic and Statistical Analysis section.

7. In the Graphic and Statistical Analysis section, UPGRADE will prompt the user for the type of analysis to be performed for each data

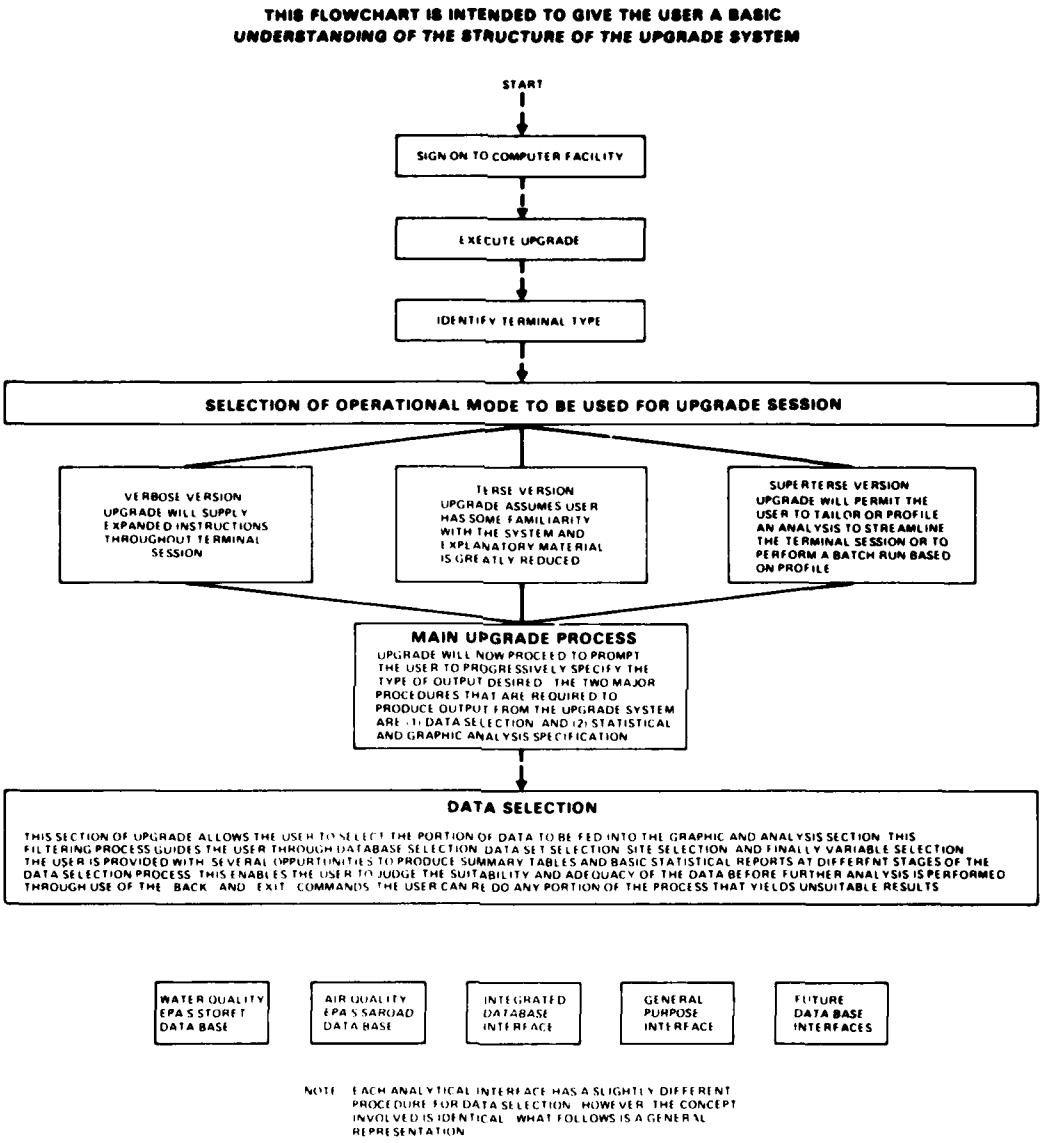


Figure F1. Flowchart of UPGRADE (Continued)

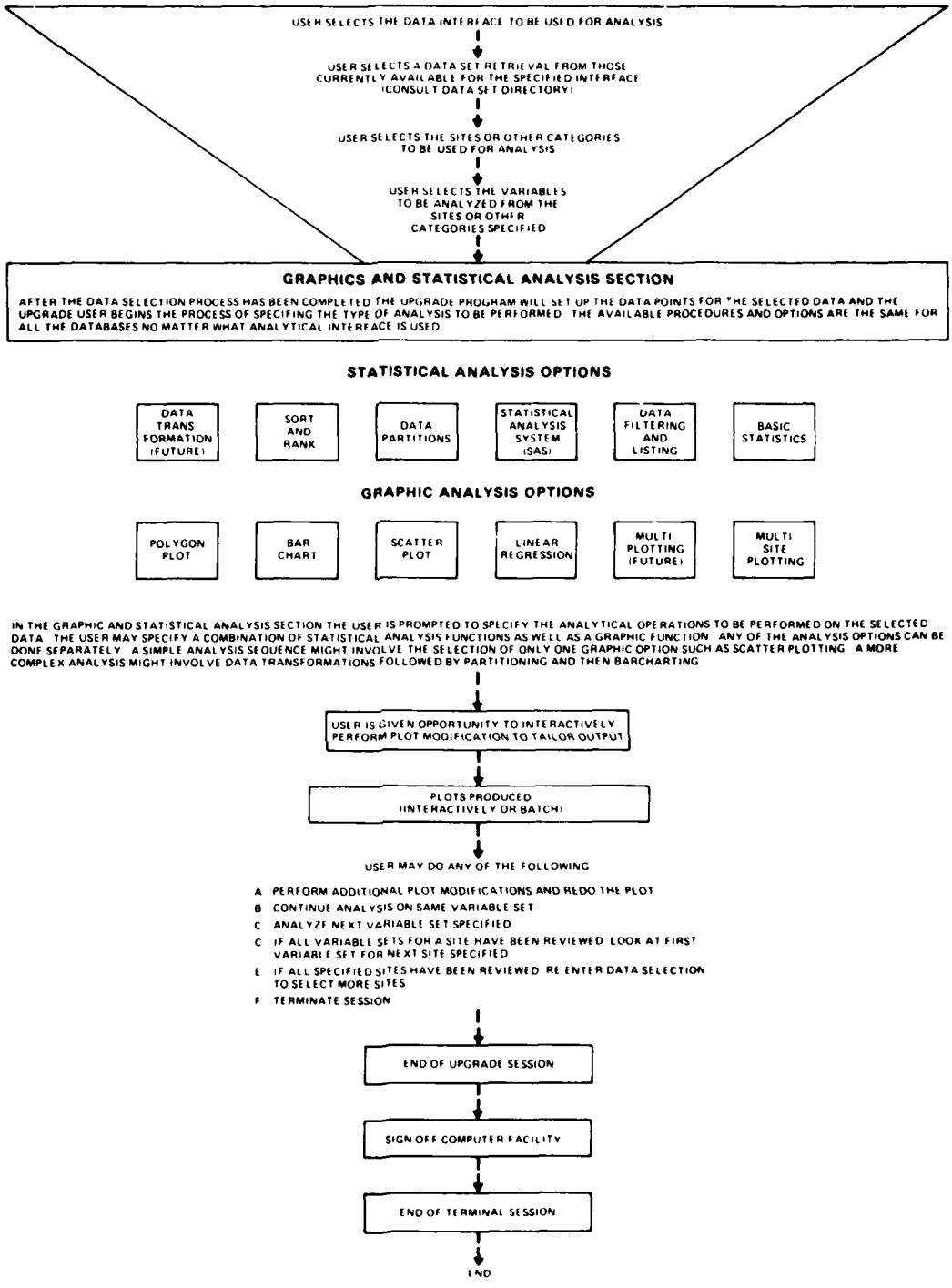


Figure F1 (Concluded)

set. There are two main categories of analysis that can be performed by UPGRADE, data manipulation and statistical analysis, and graphic analysis.

Procedure to store data

8. The general purpose interface of UPGRADE allows small-scale input of any type of data into the UPGRADE system. This can be accomplished either by entering data values interactively during the terminal session or by creating and storing a data set on a computer file to be called by the UPGRADE system. UPGRADE can also interface with the following data bases.

a. STORET. EPA's* water quality data base consisting of data from over 200,000 monitoring stations maintained by local, State, and Federal agencies in the contiguous United States.

b. SAROAD. EPA's air quality data base consisting of data from monitoring stations maintained by local, State, and Federal agencies in the United States.

c. NCHS. Sixty-nine cause census mortality data tabulated by the National Center for Health Statistics (NCHS).

d. NCHS-STORET.** Allows comparison of 53 cause NCHS mortality data with STORET water quality data, demographic data, U. S. Geological Survey and state groundwater data, etc.

Output data procedures

9. Data selection in the UPGRADE system is limited to the data bases and data sets that have been interfaced to the UPGRADE program. However, since UPGRADE was designed as a flexible analytical tool for all types of data analysis, other data bases and data sets can be adapted to the system with varying degrees of difficulty depending upon the structure and idiosyncracies of the data. The retrieval of data is done interactively during a terminal session.

Analysis available

10. The different types of UPGRADE analysis options are shown in

---

* EPA = U. S. Environmental Protection Agency.

** NCHS-STORET is an integrated data base interface.

Table F1. The information was taken from the manual, "The UPGRADE System - User's Overview."*

* President's Council on Environmental Quality. 1977. "The UPGRADE System - User's Overview," Washington, D. C.

Table F1

UPGRADE Analysis Options

UPGRADE Graphics Options

a. Scatter plotting--a plot of data points and their distribution.

b. Polygon (point-to-point) plot--data point plot with a straight line connecting successive points.

c. Bar chart--each data value is represented by a bar (numerous shading and density options are available).

d. Polynomial fit--scatter plot with up to sixth-order least-squares fitted line and table of m, r, t, and f values.

e. Multiplotting (FY 1978)--to allow up to five y-axis variables on same graph.

f. Multiple-site plotting (FY 1978)--to allow the plotting of sites (instead of a variable) on the x-axis.

Plot Modifications

A large variety of analytical and cosmetic options are available, allowing the user to tailor graphic output from the UPGRADE system. Future development will include the addition of even more plot-mod options.

a. Interchange axes--x becomes y and vice versa.

b. Reverse axis scaling--scale from maximum to minimum instead of vice versa.

c. Change scale factors--allows user to specify scale ranges. This mod can be used for "windowing" a plot so only data points within a specified range are plotted.

d. Change number of axes annotation--to modify precison of scale divisions.

e. Change number of axes tick marks--to modify precision of scale divisions.

f. Add or delete grid lines--to divide a plot into quadrants.

g. Change letter size of axes annotation--to modify legability of scale annotations.

(Continued)

Table F1 (Continued)

Plot Modifications (Continued)

h. Change symbol and symbol size of data points--will also be used to differentiate between different variables when multiplotting becomes available.

i. Change graph title.

j. Add second line to graph title.

k. Eliminate or restore plotted data points--if a regression line without the plotted points is desired, for instance.

l. Change line type--will also be used to distinguish between lines for different variables when multiplotting becomes available.

m. Change axes length--to modify dimensions of entire graph.

n. Change axes scale type--to allow use of log and probability scales.

o. Change degree of polynomial fit--to use up to a sixth-order fit for regression analysis.

p. Eliminate outlying data points from a fitted plot--to "window" a regression line to a selected range of data values.

q. Eliminate or restore current date printout that appears on every plot.

r. Change number of decimal places used for axes annotation--to modify precision of annotations.

s. Change number of decimal places used for bar annotation--to modify precision of bar chart annotations.

t. Suppress statistics printout on fitted plot--if m, r, f, t, and ndp values are not needed.

u. Change bar density and shading specifications for a bar chart.

Statistical Options

a. Sort and rank--to obtain a table of median, quartiles, tertiles, and 15th and 85th percentiles for any one variable, or complete, sort, and/or rank for each data point.

b. Data filtering and listing--to eliminate outliers, select a range of data values, or obtain a listing of data point values.

(Continued)

Table F1 (Continued)

Statistical Options (Continued)

c. Linear regression--to produce a table of coefficients of variance for regression analysis.

d. Data partitioning--to group data values for the x-axis into class intervals and plot against a partitioned y-axis variable. Statistics for class intervals and partitions can be produced.

e. Basic statistical summaries of selected data, including minimum, maximum, mean, standard deviation, number of data points, and historical period of record.

f. Data transformation--to allow user to perform arithmetic operations on variables to obtain ratios, etc. User will also have capability of using a variable for exclusion purposes to obtain a selected range of data values for that variable.

g. SAS--an integrated system for data management and statistical analysis.

Highlighting SAS's statistical capabilities are its versatile least-squares procedures, which produce a wide variety of linear and nonlinear regression analyses, analyses of variance and covariance, and multivariate analyses of variance. One can produce highly specialized analyses with comprehensive matrix manipulation procedures.

SAS can also produce multiple and partial correlation coefficients, Spearman's and Kendall's correlation coefficients, and contingency table chi squares.

It has several procedures for analyzing time-series data. One can calculate summary statistics and print them or use them directly for further analysis. One can obtain frequency and cross-tabulation tables and analyze them as well as perform discriminant analyses, factor analyses, and cluster analyses. One can construct and evaluate Guttman scales and can perform t-tests or tests of goodness-of-fit or probit analyses.

Procedural Options

a. Automatic sequencing--to perform same graphic and statistical

(Continued)

Table F1 (Concluded)

Procedural Options (Continued)

operations on a range of sites and have graphics produced automatically.

b. Station combination--to combine sites or a range of sites to produce composite analysis.

c. Batch processing--to specify a large UPGRADE run and have output produced on line-printer, CALCOMP plotter, or microfiche.

d. Mapping interface for off-line production of shaded maps.

APPENDIX G: NATIONAL WATER DATA INFORMATION EXCHANGE SYSTEM (NAWDEX)*

1. NAWDEX was established as a national confederation of organizations active in the field of water resources; its primary objective is to make information about the availability of water data available to those who need it.

System Identification

System Title: National Water Data Exchange

Code Name: NAWDEX

Writer: U. S. Geological Survey

Organization: Chief Hydrologist
U. S. Geological Survey, WRD
420 National Center
Reston, Virginia 22092

Availability: NAWDEX is available for use by Federal and non-Federal organizations active in the field of water resources. It is not for sale or lease.

Abstract

2. NAWDEX provides the linkage among those who acquire and those who use water data. NAWDEX is neither a water data bank nor a coordinator of water data collection activities, but rather a resource for the location of water data. To the maximum extent possible, all water data in the NAWDEX system are to be retained by the collector organizations.

3. An evaluation of the system is as follows.

a. Advantages: NAWDEX is a central depository for information

---

* Edwards, M. D. 1978. "Status of the National Water Data Exchange (NAWDEX) - September 1977," U. S. Geological Survey, Reston, Va.
U. S. Geological Survey. 1976. "First Interim Report on Digital Geographic Data Handling Activities in the U. S. Geological Survey," Reston, Va.

about the type of available water quality data and where the data can be found.

b. Disadvantages:

(1) NAWDEX by itself is not a water quality data management system. It will still be necessary to have access to the various data bases in order to obtain water quality data.

(2) NAWDEX does not contain any analysis procedures.

Technical Description

Major elements of the system

4. The major elements of the system are entering, storing, and retrieving information indicating where water data reside. Two major computerized water data information bases have been developed and implemented on NAWDEX. The two data bases were implemented on the U. S. Geological Survey's IBM 370/155 computer at its National Center in Reston, Va. These two data bases are The Water Data Sources Directory and The Master Water Data Index.

5. The Water Data Sources Directory identifies organizations that collect water data and locations within these organizations from which data may be obtained. The Master Water Data Index identifies specific sites for which water data are available, the types of data available at each location, the period of record of the data, the major parameters for each type, the frequency of measurement of each parameter, and the collector organization.

Procedures to store data

6. Access to NAWDEX is provided to all participating agencies through computer terminals compatible with the IBM 370/155. Procedures, instructions, and forms for the manual encoding and keypunching of data for entry into NAWDEX are available for all NAWDEX users.

Output data procedures

7. A retrieval system for NAWDEX is available for use in batch mode. This system provides for (a) selected data to be retrieved from The Master Water Data Index (MWDI) in a machine-readable form which can

be used as input to application programs; (b) a report listing procedure that allows retrievable data to be printed in columnar tables; and (c) a plotting procedure that can be used to produce site-location plots to be used as overlays for maps at a variety of scales.

APPENDIX H: STATISTICAL ANALYSIS SYSTEM (SAS)

1. The Statistical Analysis System (SAS) is the result of a need by personnel in the Statistical Department, University of North Carolina, for a very flexible and powerful statistical analysis system on the computer. As the system grew, more modules were added to aid in the manipulation of data for the statistical routines. Now a variety of data management systems can be put together by selecting and combining the desired modules available in SAS.

System Identification

System Title: Statistical Analysis System

Code Name: SAS

Writer: Anthony Barr, James Goodnight

Organization: SAS Institute Inc.
P. O. Box 10066
Raleigh, N.C. 27605

Availability: SAS can be leased to operate on an IBM 360, IBM 370, IBM 4331, or an IBM-compatible machine such as Amdahl, ITel, CDC Omega, and Riad, with the standard IBM operating systems, OS/MFT, MJT, JSI, VS2, SVS, and MVS (not CMS).

Abstract

2. SAS is an integrated system for data management and statistical analysis. By combining statistical versatility with extensive capabilities for data manipulation and report writing, SAS gives a total system to help solve problems. SAS is designed to be a computer tool for the research worker--not only analyzing the data but also managing it. SAS is capable of performing the following tasks:

a. Manipulate and transform data.

b. Store data in a user-produced data base.

c. Retrieve data from a user-produced data base.

d. Statistically analyze data.

e. Produce reports.

3. An evaluation of the system is as follows.

a. Advantages:

(1) SAS is a very complete data management system and is easy to use.

(2) SAS is very powerful and versatile because it takes advantage of the features of the operating system for storage allocation, access methods, linking, etc.

(3) Data may be stored in any form and retrieved in any sequence the user desires.

(4) Data entered into SAS are available immediately for use.

b. Disadvantages:

(1) SAS is not portable. It is limited to the use of an IBM or IBM-compatible machine.

(2) SAS does not maintain a data base as an integral part of the system.

Technical Description

Major elements of the system

4. The major elements of the system are entering, manipulating, analyzing, storing, and retrieving data. SAS is a system with half of the coding in Assembler language, nearly half in PL/1, and the remainder in FORTRAN. The SAS library has 150+ modules which are designed to help the user in analyzing and manipulating data. New modules are added to the system as they are developed and tested.

Procedure to store data

5. SAS accepts data in virtually any form from any input device. In a single SAS job, a user can enter new data into the data base; edit and transform the data; analyze the data statistically; further manipulate the data; perform more analysis; store the data permanently on disk; and so forth--without special Job Control Language or formatted parameter cards.

6. The steps listed below illustrate how new data can be entered into an SAS data set.

a. The data statement is used to begin the formation of the SAS data set.

b. Use an INPUT, SET, MERGE, or UPDATE statement to tell SAS where to find the information for the data set. The data entering the new data set can also be generated by SAS statements.

c. A user can optionally modify the data with program statements before putting it into a data set. Program statements can change data values, make new observations and variables or delete old ones, produce reports, and perform many other operations.

Output data procedures

7. The SAS user has a simple language communication with the system. With it, he can transform or delete variables and observations, form new variables, and form new data sets from existing ones. Users can merge data sets, concatenate them, update them, and form subsets of them. Users can print what they want of a data set in their own specialized format. Users can attach explanatory labels to variables. SAS also keeps track of how a user has built a data set and makes that history available to him. Adding these features to SAS's procedures for printing, sorting, ranking, and plotting data gives the user a very wide choice of reports.

Analysis available

8. The analysis of data in the SAS system is handled through the use of the procedures shown in Table H1 (Reference b). It must be emphasized that SAS is not a data base but rather a very powerful set of data management and analysis tools that can be used to build a data base to serve the needs of users.

References

9. References used in this appendix include:

a. Barr, A. J. et al. "A Users Guide to SAS 76," SAS Institute, Inc., Raleigh, N. C.

b. _____. 1979. "SAS Users Guide, 1979 Edition," SAS Institute, Inc., Raleigh, N. C.

c. Helwig, J. T. 1977. SAS Supplemental Library Users Guide," SAS Institute, Inc., Raleigh, N. C.

d. Helwig, J. T. 1978. "SAS Introductory Guide," SAS Institute, Inc., Raleigh, N. C.

e. Reinhardt, P. 1979. "The SAS Supplemental Library User's Guide, 1979 Edition," SAS Institute, Inc., Raleigh, N. C.

f. Sail, J. 1978. "SAS Views," 2 ed., SAS Institute, Inc., Raleigh, N. C.

Table H1

Analysis Procedures Available in SAS

The ANOVA Procedure
(analysis of variance for balanced data)

The AUTOREG Procedure
(autoregression)

The BMDP Procedure
(BMDP interface)

The CANCORR Procedure
(canonical correlation)

The CHART Procedure
(bar charts (histograms), pie charts, star charts)

The CLUSTER Procedure
(cluster analysis)

The CONTENTS Procedure
(contents and history of SAS data set)

The CONVERT Procedure
(BMDP, Data-Text, OSIRIS, SAS72, SPSS file conversion)

The COPY Procedure
(copying SAS data sets)

The CORR Procedure
(correlation analysis)

The DATASETS Procedure
(listing, deleting, and renaming SAS data sets)

The DELETE Procedure
(deleting SAS data sets)

The DISCRIM Procedure
(discriminant analysis)

The DUNCAN Procedure
(Duncan's multiple range test)

The EDITOR Procedure
(interactive and batch editing of SAS data sets)

(Continued)

The FACTOR Procedure
(factor analysis)

The FORMAT Procedure
(defining formats and printing value labels)

The FREQ Procedure
(frequency and cross-tabulation tables)

The FUNCAT Procedure
(categorical analysis)

The GLM/Introduction Procedure
(introduction to the General Linear Models procedure)

The GLM/Reference Procedure
(simple regression, multiple regression, analysis of variance for unbalanced data, analysis of covariance, response surface models, weighted regression, polynomial regression, partial correlations, multivariate analysis of variance)

The GUTTMAN Procedure
(Guttman scaling)

The MATRIX Procedure
(*matrix language*)

The MEANS Procedure
(means and other descriptive statistics)

The NEIGHBOR Procedure
(nearest neighbor discriminant analysis)

The NESTED Procedure
(nested analysis of variance)

The NLIN Procedure
(nonlinear regression)

The NPAR1WAY Procedure
(one-way analysis of variance on ranks)

The PDS (partitioned data set) Procedure
(listing, deleting, and renaming PDS members)

(Continued)

Table H1 (Continued)

The PDSCOPY Procedure
(copying load modules and libraries)

The PLAN Procedure
(randomized plans for experiments)

The PLOT Procedure
(printer plotting)

The PRINT Procedure
(print SAS data sets)

The PRINTTO Procedure
(routing procedure output to disk or tape)

The PROBIT Procedure
(probit analysis)

The RANK Procedure
(ranking)

The RELEASE Procedure
(releasing unused space at the end of a data set)

The RSQUARE Procedure
(all possible regressions)

The SCORE Procedure
(linear combinations of coefficients and data values)

The SORT Procedure
(sorting SAS data sets)

The SOURCE Procedure
(printing source library contents)

The SPECTRA Procedure
(spectral analysis)

The STANDARD Procedure
(standardization)

The STEPWISE Procedure
(stepwise regression)

(Continued)

Table H1 (Concluded)

The SUMMARY Procedure
(summary statistics)

The SYSREG Procedure
(ordinary least squares, two-stage least squares, limited information maximum likelihood, three-stage least squares, seemingly unrelated regressions)

The TAPECOPY Procedure
(copying tape volumes)

The TAPELABEL Procedure
(printing contents of tape volumes)

The TTEST Procedure
(t-tests)

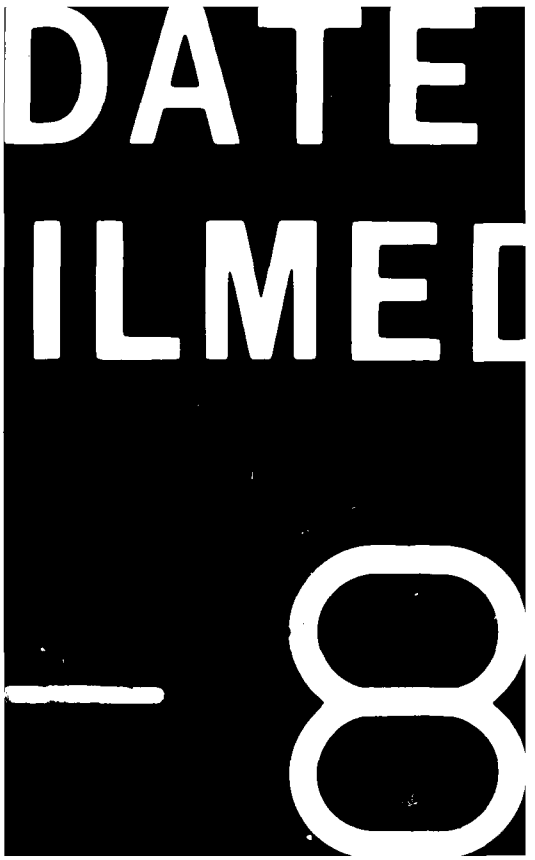
The UNIVARIATE Procedure
(univariate descriptive statistics, including percentiles)

The VARCOMP Procedure
(variance component estimation)

In accordance with letter from DAEN-RDC, DAEN-ASI dated 22 July 1977, Subject: Facsimile Catalog Cards for Laboratory Technical Publications, a facsimile catalog card in Library of Congress MARC format is reproduced below.

Smith, Margaret H
Management of water quality data within the Corps of Engineers, 1979 / by Margaret H. Smith, Phillip L. Doiron. Vicksburg, Miss. : U. S. Waterways Experiment Station ; Springfield, Va. : available from National Technical Information Service, 1980.
16, [65] p. : ill. ; 27 cm. (Miscellaneous paper - U. S. Army Engineer Waterways Experiment Station ; E-80-2)
Prepared for Office, Chief of Engineers, U. S. Army, Washington, D. C. under Work Unit No. 31608 (IVB) (EWQOS Task IVB.1)

1. Data collection systems. 2. Water quality control. 3. Water quality management. I. Doiron, Phillip L., joint author. II. United States. Army. Corps of Engineers. III. Series: United States. Waterways Experiment Station, Vicksburg, Miss. Miscellaneous paper ; E-80-2.
TA7.W34m no.E-80-2


DATE
ILMED
8