EPA-600/5-78-007
May 1978

Socioeconomic Environmental Studies Series

DATA BASE SYSTEM FOR STATE WATER QUALITY MANAGEMENT INFORMATION SYSTEM

Office of Air, Land, and Water Use
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the SOCIOECONOMIC ENVIRONMENTAL STUDIES series. This series includes research on environmental management, economic analysis, ecological impacts, comprehensive planning and forecasting, and analysis methodologies. Included are tools for determining varying impacts of alternative policies; analyses of environmental planning techniques at the regional, state, and local levels; and approaches to measuring environmental quality perceptions, as well as analysis of ecological and economic impacts of environmental protection measures. Such topics as urban form, industrial mix, growth policies, control, and organizational structure are discussed in terms of optimal environmental performance. These interdisciplinary studies and systems analyses are presented in forms varying from quantitative relational analyses to management and policy-oriented reports.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

EPA-600/5-78-007
May 1978

DATA BASE SYSTEM FOR

STATE WATER QUALITY MANAGEMENT INFORMATION SYSTEM

Grant No. S-801000

Project Officer

Harry C. Torno
Office of Air, Land and Water Use
Office of Research and Development
U.S. Environmental Protection Agency
Washington, D.C. 20460

Prepared for
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

DISCLAIMER

This report has been reviewed by the *Office of Research and Development*, U. S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U. S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendations for use.

ABSTRACT

The Pennsylvania State Water Quality Management Information System demonstration project (S-801000) is a jointly funded effort between the U. S. Environmental Protection Agency (EPA) and the Department of Environmental Resources, Pennsylvania Bureau of Water Quality Management (BWQM) The project was inaugurated to provide systems which would enhance the speed and precision with which decisions may be made in the Water Quality Management field and thereby increase the effectiveness of the program as a whole.

The objectives of the first grant period (starting January 1, 1969) were to enhance and demonstrate a State-wide Water Quality Management Information System which could be made available for use by Federal, other State, and inter-State water pollution control agencies, also, to provide a base for a Water Quality Management Data Systems techniques training program for Federal, State, and inter-State Water Quality Management personnel.

A portion of the proposed system was implemented utilizing standard keypunch card data entry for processing on a UNIVAC (RCA) SPECTRA 70/45 and provided predetermined periodic reporting capability. This portion of the project was documented in EPA Publication 600/5–74–022 entitled, Demonstration of a State Water Quality Management Information System. This portion of the system is referred to as WAMIS (Water Management Information System) Release I.

Systems concepts, systems design techniques, software, computer hardware and telecommunications capability have all experienced marked changes since the beginning of the project in 1969.

The Pennsylvania Bureau of Water Quality Management was encouraged to revamp its original systems design concepts based upon initial reports of the capabilities inherent in the EPA developed Data Base Management System (DBMS) known as the General Point Source File System (GPSFS). The modified portions of WAMIS Release I, together with the modules which were to be included in the system as part of the third year demonstration project, are collectively referred to as WAMIS Release II. This report addresses the current design concept, design technique, systems input/output capabilities, together with illustrative documentation of the Pennsylvania system.

TABLE OF CONTENTS

	PAGE
ABSTRACT	
INTRODUCTION	1
SECTION I	
RELEASE I (CURRENT) WAMIS	3
SECTION II	
WAMIS RELEASE II	
General	5
Systems Concepts	6
Data Entry	7
Information Retrieval	8
Data Management	9
Operations	10
Conversion	10
Resources	12
SECTION III	
SYSTEM DESCRIPTION	
Development	13
Analysis	13
Design	15
Implementation Plan	15
Implementation	16
System Specifications	16
SECTION IV	
THE EXTERNAL SPECIFICATIONS	
Description	18
Constructs	18
Relationships	19
Natural Hierarchy	20
Data Elements	20
Identities	21
Forms	21
Reports	22
Training and User Guides	23
General Documentation Standards	23
Development Concerns	24

BIBLIOGRAPHY 26

APPENDICES

A. External Specifications
B. Internal Specifications
C. Project Control Book
D. DMS 1100 Schema

Please Note: Copies of Appendices A through D are available on a loan basis from either the Department of Environmental Resources, Bureau of Water Quality Management, or the U. S. Environmental Protection Agency, Offices of Research and development

INTRODUCTION

In the early 1960's, it became apparent that the accelerating accumulation of information relative to water quality management required automated data management techniques. High speed electronic computers already had demonstrated the ability to handle masses of information for other purposes. Therefore, it appeared that water quality management data systems using computers would become essential in guiding many of the policy decisions of the coming decades.

The 1963 conference of Pennsylvania State Sanitary Engineers formed a Joint Committee on Water Quality Management Data. Organizations represented on the committee were the Army Corps of Engineers, Soil Conservation Service, State and Interstate Water Pollution Control Administrators, U. S. Geological Survey, U. S. Public Health Service, and later, the Federal Water Pollution Control Administration. In May, 1967, the Committee issued a report entitled Water Quality Management Data Systems Guide. The primary purpose of this manual was to provide a guide to the development of water quality management data systems for agencies with little or no experience in this field.

The Commonwealth of Pennsylvania has a massive investment in physical facilities to aid in the management of its water resources. In administering the wise use of this investment and in assuring compliance with federal and state statutes designed to protect the water resources, the Commonwealth Bureau of Water Quality Management is faced with managing a large amount of information.

Because of the magnitude of information to be collected, stored, retrieved and analyzed, the concept of a state-wide water quality management information systems (WAMIS) was developed. The WAMIS system reflects the objectives of the Joint Committee manual and additionally incorporates objectives related to the specific problem areas particular to the State of Pennsylvania.

Some of the areas of water quality management for which WAMIS provides information support are:

1. Facility inspections
2. Progress in terms of enforcement, construction and water quality upgrading
3. Permit processing activities
4. Identify problem areas and priorities
5. Determine specific treatment, research and budgetary needs

6. Treatment plant operation reports

7. Treatment plant operator certification

8. Word processing activities

9. Ground- and surface water quality

10. Planning

The subsequent sections of this report define the system objectives as perceived at the time of project initiation and describe how these objectives were modified through experience with the project during later years.

The Bureau of Water Quality Management's (BWQM) current version of WAMIS is described in Section 1 (Release I WAMIS). The system design objectives of the new version is described in Section 2 (Release II WAMIS). The development cycle, the external specifications, training, and user guides are introduced in Section 3. Section 3 also describes the general documentation standards applicable to all WAMIS Release II documents that were developed. The purpose of the standards is to insure that people working on WAMIS were able to develop the system and control changes. All communications regarding Release II WAMIS are published in formally identified documents prepared in accordance with this standard. These formal documents are presented in the appendices. The appendices are available from EPA or the Bureau of Water Quality of the Commonwealth of Pennsylvania.

SECTION I

Release I (Current) WAMIS

The Release I system as designed and implemented by Price Waterhouse in 1970–1971 established a workable water quality management information system. The systems were implemented and integrated into the BWQM on-going program efforts. As Bureau personnel gained familiarity with the capabilities and limitations of the system, it became apparent that increased capabilities were needed to meet the data processing requirements of the Bureau.

On the Release I system, the user is locked into utilizing keypunched cards as input with periodic and fixed reporting capabilities. The user has to wait an unacceptable length of time for software modifications for the production of new reports. Therefore, the Release I system cannot respond to changing information needs.

Shorter turnaround time between the input of information and output on a WAMIS report was required. The Bureau functions of grant processing, enforcement and permit processing require turnaround time as short as 24 hours.

An analysis of the current system showed that 29 official forms are used to record the values of approximately 390 data elements. The information is gathered principally by Bureau personnel at the seven regional offices throughout the state and recorded on card/column-oriented forms. Significant time is spent transferring data from working documents to the data processing forms. These EDP forms are mailed to Harrisburg for review and keypunching.

Thirteen reports are produced by the Release I system. These reports are distributed by the central WAMIS office to the regions at varying frequencies and are used as a source of information for decision-making responsibilities. Manual data files are maintained by each region and are used to obtain more current facility status and history than the Release I system can provide. There is also no provision for remote job entry to allow regions to specify types and contents of reports on a timely basis.

The existing WAMIS systems have the ability to produce reports containing information from very large files. It is very difficult to produce reports containing only desired data from selected records in the files. Therefore, the use of the existing printouts provides specific information about many data records in a very large quanity.

It is very difficult to relate information in one system to information in another system. This capability is needed by end users and management to support the decision-making processes required for responsible water quality management. The ability to relate data in different systems would also reduce data redundancy.

The activities and, therefore, the information requirements of the BWQM, are controlled by the Bureau's management environment. This environment is established by the state and federal laws, budget and public opinion. The dynamics of this environment require that the EDP system be flexible and can respond to changes in data management needs.

These observed requirements, fast turnaround, general reporting, related information, and flexibility were the basis for the development of the specifications for the new (Release II) WAMIS system.

SECTION II
WAMIS RELEASE II

General

The Commonwealth of Pennsylvania's water quality management system is described in detail in the EPA Publication No. 600/5–74–022 titled, *Demonstration of a State Water Quality Management Information System.* The objectives of the existing system were to demonstrate a state-wide water quality management information system, including case status reports, project status reports, water quality control and plant operation control systems.

Development and implementation of the system was performed in three phases. A comprehensive system design was developed and the facility status, water quality and contact modules (a name and address system) were made operational. The third phase anticipated that the plant operation control (POC), project status, grants and history modules would become operational. Time and funds did not permit the implementation of the third phase.

The system described and demonstrated in this report is the demonstration of the third phase. The report includes an overall coordinated systems design to interrelate the existing different modules, to demonstrate a plant operational control system (POC), and a PREP system which meets the needs of project status, grants and historical applications.

The third phase of this project includes an overall system design because experience with the existing water quality system has shown that data relationships among modules is very interdependent; therefore, the design integrates the various modules into a more cohesive system. This design recognizes the fact that data (i.e., data elements) have multiple users and should not be restricted by defining a computer file or records.

It was decided that the overall WAMIS system should be designed to operate on a generalized data base management system. The Bureau of Water Quality Management was provided with information by the U. S. Environmental Protection Agency (EPA) which indicated that the General Point Sources File (GPSF) developed by EPA and available as public domain software, would be able to handle the WAMIS design requirements. Support for GPSF was later discontinued by EPA and other data base systems were investigated. A Management Information System (MIS) is designed to retain information in a "data base." (A data base being defined as a central repository for interrelated network type data.) The input consists of transactions (add, delete, or modify), and the output consists of reports. The third phase demonstration Release II involves three interrelated information modules – Facilities, POC and PREP. These modules form the central data base for the WAMIS system.

The facility (FAC) module is included in the data base because all other modules are related to information defined by this module. The facility module defines data about any entity over which the Bureau of Water Quality Management has regulatory control. Examples are wastewater treatment facilities, water supplies, bathing places, industrial waste discharges, dams and encroachments. Data is maintained about the facility locations, constructions, inspections and populations served for each entity contained in the module.

The POC module was developed to support the EPA NPDES self-monitoring program. The objectives of POC are to monitor the reporting requirements for the plant operations for all wastewater dischargers. The POC module will compare daily plant discharge parameters and values with the permitted standards in the NPDES permit. It will then generate exception reports showing violations of the permitted standards.

The PREP module is designed to monitor and report on any generalized time-constrained activity. These activities include the processing of permit applications, processing of operator certification, and the monitoring of construction grants processing. The PREP module will also provide historical information about completed activities.

Since the scope of the total Release II is rather large, a smaller portion known as Release IIa will be implemented with batch input capability and on-line retrieval capability. The FAC, PREP and POC modules are now operational. The other modules that comprise the total Release II design, such as Water Quality, Contact, Planning, etc., will be implemented as resources allow. These modules will continue to be supported by Release I until they can be included in Release II. The information in Release IIa is detailed in Section III.

Release IIa will replace the facility status system of Release I. The data from Release I will be transferred to Release IIa as discussed in the section on conversion. The Release IIa system meets the requirements of the third phase of the WAMIS project.

System Concepts

The objectives of the Release II system are to provide more flexibility of data input and retrieval using current data base management system (DBMS) technology. The system will use a UNIVAC 1110 computer and the DMS 1100 DBMS as the basic data management system. Five major advantages that this technology offers over conventional processing techniques are:

Program Development – Data base management software will simplify application program development by performing automatically many routine functions under software control.

File Structures – Release II file structures offer powerful cross-referencing capabilities that eliminate redundant data and make possible fewer files.

System Design – Fewer 'overhead' programs for use with specific applications are needed. This cuts down on sorts, merges, file extracts, duplicating updating and scanning of files.

Operations – Fewer files means less set-up time and less external manual management of programs and data.

Flexibility – Changes required by users can be implemented with ease.

Data Entry

The input to a data management system is extremely flexible. The commonly used method of input, coding a form, keypunching and inputting data to the system in a batch mode is somewhat obsolete using data base management system software. (BWQM does intend to use batch input for large volumes of routine data to our system, particularly for input to the POC module.)

The most user-oriented approach to inputting data in a data management system is to have users directly key input to the data base by using a cathode ray tube terminal (CRT) in either an on-line (interactive) or demand (batch) mode of operation. The Release II system is applicable to terminal input and it is expected that as the Bureau of Water Quality fine tunes the Release II system, regional personnel will utilize the direct input mode for PREP and facility data. This will allow the user to immediately store his new information required in the data base without the time delay involved in keypunching and batch update.

Regional telecommunication capability will be utilized for data entry to the system and information retrieval from the system. To the maximum extent possible, Bureau working documents will be the source used to update the data base. Input errors will be displayed during the input process to enable immediate correction of errors by the terminal operator. (This will reduce time spent on encoding data onto keypunched forms.) Computer printed forms containing previously stored information will be used as turnaround documents. This will make separate reports unnecessary for activities that develop large amounts of update data for the data base. Examples of these activities are facility inspections and enforcement actions. We anticipate that the initial updates of the data base through a cathode ray terminal (CRT) will be done in a demand (batch) mode. The Bureau of Water Quality expects to utilize real-time capabilities (interactive) in the future. Real-time will be used where the user requires less than 24-hour turnaround between data input and output. Real-time capabilities with CRT input will be developed as implementation off the project progresses. Development of the real-time capability is expected in 1978.

In order for a user to develop a CRT input, a form will be projected on the CRT screen and the terminal operator will fill in the spaces and submit the data to the data base for storage. The design for this is included in appendix B. An editing program will inform the terminal operator when errors are found in the input. The editing validates codes, length of field and logic problems, where necessary.

Information Retrieval

Just as inputs in the data base environment are not constrained to keypunching, outputs from a data base environment are extremely flexible. The Bureau of Water Quality has designed standard reports to manage the routine large-volume type reporting requirements, particularly for the POC modules. These reports are found in Appendix A in the section titled Reports.

For much of the Release II reporting capability, the Bureau of Water Quality will be utilizing the DMS 1100 Report Writer. The Report Writer is a UNIVAC software product that can be used with the DMS 1100 data base system. It allows the user to select information from the data base and format the information into a useful report. Production of reports using the Report Writer is more expensive than producing a standard report to get the same information. The advantage that the Report Writer has over standard reports is that reports can quickly be developed to meet the users needs. This is necessary in an agency like the Bureau of Water Quality where information requirements are dynamic.

As the Bureau of Water Quality gains experience with reporting needs, a decision will be made as to which reports are to be standard reports and which reports will continue to be produced by the Report Writer. As some reports are shown to be produced on a routine basis, a program will be written to produce the reports.

Initially, the basic output from the data base is the retrieval of information resulting from the input of ad hoc queries at a telecommunication terminal. Using a few instructions in conjunction with selection criteria, the requested information is displayed on the CRT. For example, information on a project's priority, status and history will be retrieved using the ad hoc Report/Writer capability.

Requests for standard reports are also entered via terminals. If error-free, these result in searches of the data base for the requested information and formatting of this information is required. Sorting, totals, tallies and code conversions are performed and the report can be read from the CRT or printed at a high speed terminal.

On standard reports that will be produced by a program, the data that are coded for input to permit cost-effective storage will be displayed in expanded notation (e.g., codes will be translated into English before print) on the printouts. In this way users will be able to read the reports without referring to code lists. Reports produced by the Report Writer will not contain expanded codes because the Report Writer cannot translate codes before printing a report. The Report Writer reports are short, only one or two pages of printout; therefore, the users do not find coded output to be a problem on these reports. Standard reports would be much longer and, therefore, translation of codes on standard reports relieves the user of referring to a code list in order to understand the report.

Data Management

The data that is input to the Release IIa data base originates in the BWQM Regional Offices. The source documents are permits, enforcement reports, grant applications and other documents that are used for information and communication in the Bureau's ongoing program activities. Since the source documents are records of the Bureau's activities, the information on the source documents is considered accurate. Therefore, the Release II system only performs edits to check the data for acceptance to the data base.

The system performs edits on all coded input to check that the code is valid and of proper length. The identities of the records are checked to be sure that all appropriate records required to establish a proper hierarchy are in the data base. For example, a municipality record cannot be added to the data base unless the county record for that municipality is already in the system. Data fields are checked for length and type of characters allowed. Numerical fields are validated to make sure there are no letters included in the data.

Maintaining accurate and logical data is the responsibility of the regions that submit the data for input. The regions receive an error list for all data submitted. The listing describes each record submitted, its status (either accepted or rejected), and provides a message to tell the user why the system rejected any data. After an update is run, the regions receive a report containing the updated information which they use to assure that the data is correct and that it provides the information that was on the source document.

By using the same forms for both the source document and data input, coding errors are reduced. The data only has to be transferred one time, either during keypunching or during terminal input. This avoids coding errors that tend to occur when information on a source document is transferred to a special coding form.

The management of the data base operation including updates, deletes, report production, etc. is a responsibility of the EDP manager who is part of the WAMIS Section. His responsibilities include management of contracts, defining system work, and overseeing the data base operation. Data base administration (DBA) which is a team effort shared by the analysts, programmers, and the EDP manager is responsible for the data base activities. These activities include scheduling of updates, storage allocation, internal data structure, recovery, security, and any other activities required to maintain a cost-effective data base operation.

DBA activities also control the deletion of data from the data base. All requests to delete records from the data base must be approved and executed under DBA control. This is important because deletion of a record from the data base system also deletes any records that are related to the deleted records.

Operations

In the simplest case, "operations" are the following processes of Release II WAMIS:

Receipt of data from BWQM staff and rejection of unacceptable data.

Incorporation of received data into the data base.

Receipt of requests from BWQM staff or information (reports) and preparation of the same from the data base, using the Report Writer or a standard report.

For the non-systems staff member, these are all the operations there are, and the user manuals provide the information required to interface properly with the system. The Release II operations are diagrammed on chart 1.

Certain "system" staff are required to handle the other aspects of the operations:

Recovery in case of error

Monitoring of operating costs

Reorganization of data files.

As mentioned earlier, the process of using the system will make apparent a continual list of modification desires. These will be collected and, on a periodic basis, form the rationale for beginning a new development cycle which will consist of additions, modifications, or deletions to existing system components.

Conversion

"Conversion" is the process whereby BWQM people stop using Release I WAMIS and begin using Release II WAMIS. This involves distribution of new forms and procedures, training, incorporation of existing data into the new data base, and the destruction of old forms and procedures. Old programs (Release I) and data files will be removed to archives for reference.

This would clearly be a major activity even if it began with a completely operational version of Release II WAMIS capable of performing all of the processing required by the external specifications.

The users will input data into the Release IIa system as portions of the system become available to the user. The data will be input in small amounts at a time. This is being done as opposed to inputting or converting all of the Release I data at one time. Controlled input to the data base will give users time to become familiar with the data base environment. BWQM feels that users cannot deal with a data base as a black box, but must be familiar with the logical relationships between data and must understand the data base structure.

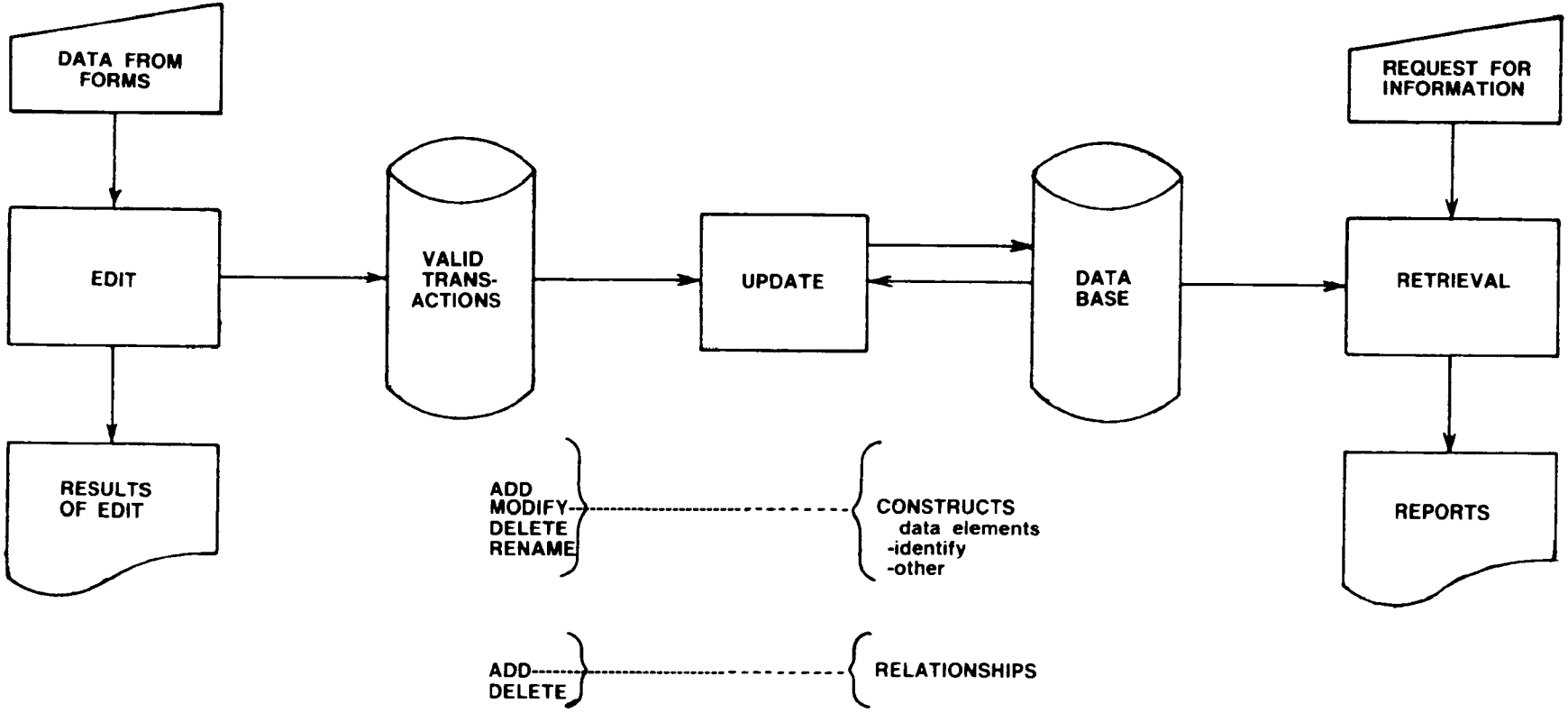

FIGURE 1 RELEASE II WAMIS OPERATION

Resources

The Release II system was developed in a time share environment using terminals in a demand mode for the writing and debugging of programs. The larger programs were coded and keypunched and input to the UNIVAC 1110 computer by remote card reader. A total of 210 programs were developed for the Release IIa subsystem. The total Release II system will require approximately 450 separate programs. Using the top-down modular techniques, only 10 of the total programs were large enough to require keypunching and batch input.

The Bureau of Water Quality has four programmers and one computer systems analyst working on the design and development of the Release II system. Combined with the other obligations, the Bureau has been able to commit approximately 2 1/2 systems persons to Release II work per year.

To develop a major system using a computer in a time share mode and with the limited number of staff available to the Bureau required that the Bureau's management accepted development times in excess of two years. In order to maintain direction and develop a system that would meet the needs of the Bureau over such a long period of time, a very rigid method of design and documentation was used. The design and documentation methodology is discussed in Section III and detailed in the appendices.

In order to maintain the Release II system, system staff will be required to operate, enhance and modify the system. Staff will also be allocated to provide data base administrative support. Data base administrative support is required to make decisions concerning update frequencies, operating cost analysis and scheduling. Systems staff will be required to perform the programming and other systems changes that the data base administrators identify. Total systems staff required to maintain Release IIa after implementation is completed will be no less than one analyst and two computer programmers. Contracted services may be used for additional systems support. Although contracted services will only be used for activities involving vendor supplied software, such as operating system and DMS1100 charges. BWQM will support all of the 210 programs written in-house.

SECTION III
SYSTEM DESCRIPTION

Development

The WAMIS Release II Development Cycle, diagrammed on Chart 2, consists of four basic activities:

Analysis	– The production of external specifications based on communication with BWQM staff.
Design	– The production of an Implementation Plan, given certain constraints such as the computers and languages to be used.
Implementation	– The production of the system (machine instructions, user manuals, etc.)
Operations	– The day-to-day operation of the system.

Although it is not readily apparent from the diagram, the development cycle is often continuous. Once operations begin, it is envisioned that user requirements will change. In this case, a new cycle begins at analysis and ends with upgraded operations, resulting in a modification to the Release II system.

Following is a discussion of the details of the development cycle.

Analysis

Analysis in the case of Release II WAMIS has been based on a very high confidence in the capabilities of the BWQM technical staff and in the belief that the system will be more useful if it is a step in an orderly evolutionary process of BWQM information requirements. As a result, a large portion of the analysis function has been delegated to BWQM technical staff by the formation of committees known as Task Forces. The Task Forces served to assure that BWQM staff had a major part in the Release II design.

Task forces were set up for each major information module. These Task Forces included technical and systems personnel from the central and regional offices. The major Task Forces are Facility, PREP, POC, Water Quality and FEUDS (planning and modelling). An Executive Task Force consisting of upper management and system staff coordinated and directed the activities of the other Task Forces. The major goal of the Task Forces was to define the external specifications of the system. This involved defining data requirements, reports, forms and aiding in fine tuning the data base. The definitions that were developed are listed in Appendix A.

The basic methodology of defining the external specifications can be summarized in a simple procedure:

Step 1.	BWQM needs as expressed by individuals and/or the Task Forces were defined in a rigorous form in the specifications.

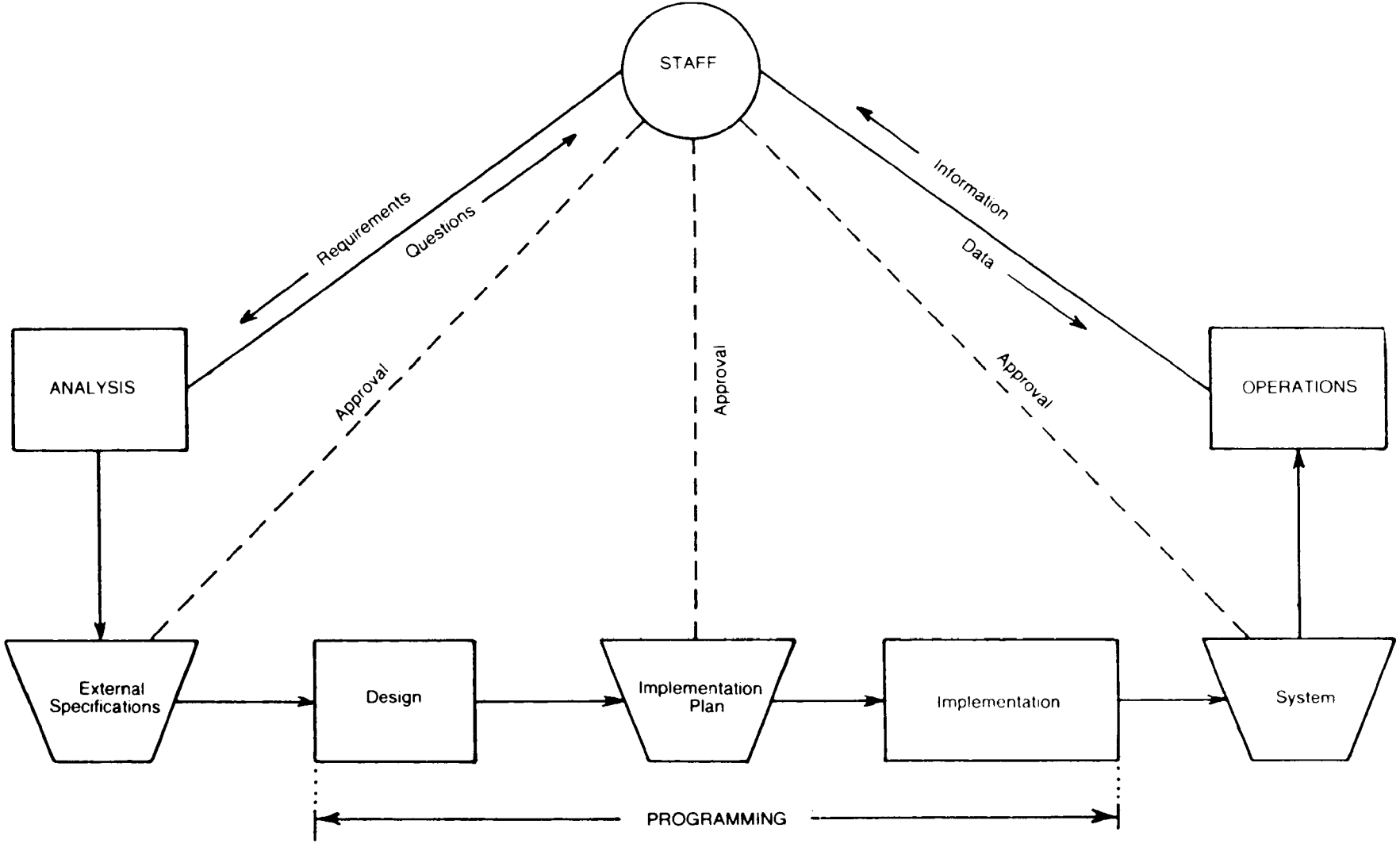

FIGURE 2 WAMIS RELEASE II DEVELOPMENT CYCLE

Step 2. The Specifications were published and the people involved determined whether their needs would be met by implementation of the specifications.

This procedure was repeated until each person was satisfied that the system as defined in the specifications would fill his particular needs. The final product is a specification of a system that will assist the Bureau in performing its functions but that will not have traumatic impact as it becomes operational because the user has been involved in the development of the specifications.

Design

The design of Release II WAMIS proceeded in a top-down manner (McGowan). This simply means that the design was produced in levels. The first level, level 1, being the overall systems design. Level 2 was a more detailed design of the overall system documented in level 1. Design was performed in levels until sufficient detail was provided for programming to be started. This process assures that the programs will work in the total system.

The top down process allowed continual monitoring by DER systems and Water Quality personnel to insure that Bureau's priorities and interests took precedence in all design decisions. The process consisted of:

Step 1. The designers were authorized to proceed with a given part of project. Initially, the project was the entire system design phase.

Step 2. The design was developed which resulted in the definition of sub-projects (levels).

Step 3. The design was evaluated and the design was approved or Steps 1 and 2 were redone.

Step 4. After approval, the new defined sub-projects were classified as follows:

Class 1. Projects for which design was authorized. These then could proceed, independent of each other, through Steps 1–4 for each project.

Class 2. Projects for which no further design was required. They were ready for implementation when authorized.

Implementation Plan

The implementation plan consists of two kinds of documents:

Internal Specifications (IS) See Appendix B.

Project Control Books (PCB) See Appendix C.

The internal specifications contain the description of how the system was designed. They were prepared in accordance with the WAMIS Doc. IS-STD Standard (Appendix B).

The control project books contain the instructions as to how the system was to be implemented. PERT diagrams and event definitions were the basic tools for directing the implementation effort. These documents were prepared in accordance with the standard: WAMIS Doc. PCB-STD (see Appendix C).

The implementation plan was divided so that the information required to maintain the system (IS) would not be confused with the information required to control the original development (PCB). After the system is operational, the internal specifications become the so-called "documentation" of the system – i.e. a guide to the people responsible for maintenance and operation of the system in the future.

Implementation

As all design tasks were finished and authorized to be implemented the implementation proceeded as follows:

Step 1. The project control books for all levels directly superior to the authorized project were obtained by the programmers and used to find the description of events, due dates, etc.

The project control books were used to identify the complete definition of the authorized project.

Step 2. References to the internal specification documents were also obtained. These contain definitions of routines (programs, macros, subroutines, etc.) and records (files, tables, work areas).

There are two kinds of implementation activity depending on the class of the project:

Class 1. Projects which have an approved design are implemented in strict accordance with the design, preferably by the designer or under his leadership. In the case of Release II, most of the programming and other implementation activities were done by the designer of the project.

Class 2. Projects requiring no design approval can be implemented in any manner consistent with good professional practices and general standards.

System Specifications

WAMIS Release II consists of the following components which are independent of the computer installation that will be used to operate the system:

1. A general introduction document (WAMIS Doc. Gen, Appendix A). This document is designed to be updated as required so that it always presents a current overview.

2. The External Specifications as described in Appendix A.

3. The Internal Specifications Appendix B.

4. The project control book Appendix C.

WAMIS Release II, implemented by BWQM is operating on the following configuration:

Univac DMS-1100 software

Univac 1110 hardware configuration.

WAMIS Release II consists of the following components which are developed to fit the configuration defined above:

Parameters, instructions (i.e. programs), data, etc. required to cause the configuration to perform in accordance with the external specification (in the form of cards, listings, etc.).

Internal specifications describing the above in a context suitable for maintenance.

Users Manuals. These, combined with (1) the external specifications and (2) the user manuals for the DMS-1100 and the UNIVAC 1110 configurations contain the information required by DER to effectively use the system.

SECTION IV
THE EXTERNAL SPECIFICATIONS

Description

The external specifications describe the system in terms of its inputs, outputs and storage content (data base). They do not contain any design decisions such as what computers (if any) or what kind of storage media (disk, tape) will be used.

The purpose in preparing a system definition which is independent of method of implementation is twofold:

1. Potential users of the system have a description of what the systems will do for them – which they can evaluate – as opposed to a description of how the system will operate – which they cannot evaluate without experience on the systems.

2. Designers can concentrate on a well-defined problem, producing the best solution within the budgeting and physical constraints imposed.

The specifications will be kept up-to-date after the system is operational. If events should force the redesign of the system for an alternate configuration, the external specifications insure compatibility. The external specifications can be used to develop a similar system for other water quality management organizations irregardless of the hardware configuration they may wish to use.

A detailed description of the WAMIS Release II External Specifications is provided in Appendix A. The external specifications of WAMIS Release II are contained in various formal documents beginning "WAMIS Doc ES-II" as identified in the following discussions. WAMIS Doc Gen. section, introduces and describes the external specifications; therefore, it should be considered a part of the external specifications and prerequisite reading for anyone using the external specifications. This section introduces the technical terminology required by the reader and introduces the individual parts of the external specifications.

Constructs

The data base contains information about places, people. projects, measurements, etc. as specified in the External Specifications (Appendix A). The totality of information retained about any single person, place, measurement etc. is hereafter called a "construct", consistent with the use of the term in physics and philosophy. A construct is the memory or conceptualization of a real or imagined entity (Infotech International Limited).

The term "construct" is preferred over more traditional terms such as record, repeating group, or entry in an array because the latter tend to describe physical storage areas. The purpose of the external specifications is to specify what data the system is to retain, not the detailed technical methodology to be employed in the retention process.

As shown in Chart 1, the input forms are edited and result in the DBMS performing the following transactions regarding constructs:

ADD A new construct is to be placed in the data base.

MODIFY An existing construct is to be modified.

DELETE An existing construct is to be removed from the data base.

RENAME: The identity of an existing construct is to be changed.

When it was decided that information about facility inspections was to be retained, a new construct type, FI, was defined. When the information regarding the occurrence of a given inspection is entered, a new construct (the given inspection) is thereby added.

Relationships

The basic information retained in the data base is:

1. There exists certain constructs (facilities, enforcement actions, municipalities, etc.).

These constructs have certain characteristics (data elements) (i.e., facilities have a location, a daily flow, etc.).

There are relationships between certain constructs.

Some examples of relationships are:

"A plan is associated with a demand center."

An establishment is responsible for a facility."

An operator works at a facility."

The general practice is to confuse relationships with data elements. Thus, one might define for facilities a data element called "responsible establishment". This works to a certain extent, but is basically wrong. It obscures the fact that a significant complexity of the system exists. The external specifications of WAMIS, therefore, contain definitions of relationships in lieu of defining data elements for one construct as being links or pointers to another. In this example, a relationship "establishment is responsible for a facility" was defined.

The relationships in WAMIS Release II are specified under both constructs entering into the relationship. A description of the characteristics of each relationship is found in Appendix A.

Natural Hierarchy

A special relationship exists between certain constructs; this is an identity subordination. For example, a sample (SM) is related to a facility (FC) by the fact that in order to identify a sample, the corresponding facility must be identified. When a construct must be logically related to another construct to be meaningful, it is described as being in Natural Hierarchy. If a construct is in Natural Hierarchy to another, the identities of both constructs must be provided to retrieve the subordinate construct. A relationship between the two constructs could have been defined. However, it is desirable to minimize the number of relationships in the system. Insofar as possible, constructs are placed in a natural hierarchy. In the case of a water quality sample, the hierarchy is:

FC – facility

SM – sample

CP – custodial period

SC – substrate component

VL – value

Therefore, to identify the pH of a sample at a facility, the identity of FC, SM and VL would have to be given.

The External Specifications show the natural hierarchy of all constructs.

In the case of the most simple report, a user selects constructs of a given type (facilities, for example), orders and displays the desired information. It is also possible to reference, in the same report, constructs of other types (samples, unit processes, etc.) provided they are in natural hierarchy.

Data Elements

For each construct type there is a list of its attributes called "data elements" contained in WAMIS Doc. ES-II-DED (Data Element Dictionary. Appendix A). Each data element is assigned an arbitrary identity number such as FC92477 (facility name) and is described according to type, length, etc.

The types of data elements are:

A/N – (alpha/numeric) – This means a sequence of characters as they were input without processing of any kind except to sort for reports. This is data such as names, addresses, descriptions, and other strings of letters.

NUM – (numeric) – This means a number. The value is assumed to have units and a decimal location of some kind.

CODE – This means the value must be one of a specific list. It is planned that for each code value there will be a corresponding expansion for reports defined in WAMIS DOC. ES-II-CODES (Appendix A).

DATE – Define with format DDMMYY, DDDYY, etc.

ID – Identity data element.

The data element identity number is used in all retrieval requests.

Identities

The construct VL (value) contains only three data elements called PARAMETER, FINDING and REMARK. At least one of these data elements must be an identity in order for a transaction (ADD, MODIFY, DELETE and RENAME) to take place. In the construct VL, the identity element is parameter. The non-identity data element (FINDING or REMARK) can be changed by the MODIFY transaction. If an identity element is to be changed, a special transaction type, RENAME, must be performed. Depending on the system methodology, a RENAME can be simple or complex. It could, for example, result in records being moved from one part of a file to another part of the file.

All identity elements have a description, for example, "(05VLSM)". This example says "this is an identity data element for the construct type VL. It is element 05 in a sequence of identities. Element 04 is an element of construct type "SM".

The External Specifications provides the complete list of construct types and identities.

Forms

Input to the system is on working forms whenever possible and is interpreted by the edit phase of the system. The basic content of all forms is be given in WAMIS Doc. ES-II-FORMS. The detailed layout of the form will not be shown because the system must be independent of layout up to final implementation. When input forms must be designed, they are designed to be the form that BWQM personnel use for information record keeping during their working routine (working forms). This reduces the activity of coding data from a source document or to a special input form.

Input for Release II will be collected on working forms and entered via terminals. The data is edited and errors (invalid fields, etc.) are communicated to the data collector for correction. These corrections can be mixed with ordinary subsequent input as desired. The Data Base Administrator decides when the edited data is to be entered in the data base.

All forms will be read and edited resulting in the basic update transactions: ADD, MODIFY, DELETE or RENAME a construct and ADD or DELETE a relationship.

Transactions will be rejected if they call for creation of a construct or relationship that already exists, the modification or deletion of one that does not exist, or have invalid identity elements. Individual data elements will be rejected if they do not conform to the standards for their type (for example VALUE=12A instead of 128 where VALUE must be a number.)

No attempt will be made to screen out illogical data, although data will be checked for length, type and valid codes. Instead, reports will be produced showing the information that was input. Errors will be corrected by updates. The reason for this is that the data entry clerk should not receive rejects which are not a result of typing errors. The latter will usually result in invalid data which unclear will be screened out. If it is found to be desirable to have logic checks, the external specifications will be modified to show, for each data element, the logical requirements it must meet. BWQM has found through past experience that logic checks are not very useful except in such simple cases as where a date field is checked to be numeric.

Reports

The principal outputs of Release II WAMIS are user-specified reports, most of which will be the result of ad hoc queries. A large number of people are to be trained in the method for obtaining information from the system, which, for the simplest case, is:

SELECTION – For a given construct type, name the criteria that each individual construct must meet to be selected.

Example: Select Facilities, Type = Sewage Treatment Plant.

ORDER – Name the data elements which control the order in which the selected constructs are to be listed.

Example: Design Average Flow – descending.

DISPLAY – Name the data elements to be shown on the report and describe report constants (headings, pagination) and the positioning of the output data.

It is possible to provide machine readable output for input to other systems (such as STORET).

There are systems outputs other than reports – these are edit messages, file maintenance reports, and lists regarding the user-supplied report specifications (errors, predicted results, etc.).

WAMIS Release IIa will produce the reports listed in WAMIS Doc. ES-II-REPORTS under the heading POC Module, PREP Module, and Enforcement Module. The constructs that are being implemented in support of this are:

```
CY---County
      MN---Municipality
            AF---Administrative File
ES---Establishment
      AP---Act 339 Payment
      NP---Need Project
FC---Facility and/or Resource Sampling Station
      FP---Facility Permitted Value
      OR---Operations Report
            OD---Operations Report Day
                  OV---Operations Report Value
            OS---Operations Summary
OM---Operations Report Parameter
PP---PREP Project
      GR---Grant
      PE---PREP Event/Task
            SE---Sub-event/Task
```

A loss of capability due to conversion of facilities from Release I to Release II will be avoided by including such reports as are provided at present if they are not made obsolete by Release II reporting capability.

Training and User Guides

User guides will be necessary to link the WAMIS information contained in the external specifications to the hardware and software. These will be provided as a set of documents identified as WAMIS Doc. USER. The exact list will be provided as part of this section when the design is complete.

The user guides are considered part of the system. Their preparation, their distribution and the initial training in their use is considered part of implementation. User guides are part of the design of the appropriate subordinate design module. See the edits and retrieval designs for examples. (Appendix B, WAMIS IS).

General Documentation Standards

In order for system documentation to have continuing utility, there must be a method for communicating changes. This section describes the method used for all WAMIS Release II documents. The documents are updated as they become obsolete so that at all times the formal documents contain the current information required to use, evaluate, or control the system.

Each page published will begin with a heading containing the following elements:

Document identity (WAMIS Doc. xxx)

Publication control number (level xxx)

Pagination (Page xxx)

Each document will have a distribution list and a document status page. The purpose of the latter is to show the publication history and the concurrence status.

The resulting library of documentation should contain all communication regarding Release II WAMIS. The complete detailed documentation is found in the appendices.

Development Concerns

The major problems encountered in the development of the Release II system involved the length of the project, the complexity of the data base, and the training of the system personnel in data base techniques.

The development time required from the finalization of the External Specifications to the implementation of the Release IIa sub-system was approximately two and one-half years. A project of this length is difficult to manage because of staff turnover, budget ramifications, and changes in priorities.

In order to reduce the effect of staff turnover on the project, extensive documentation and a modular programming approach was utilized for systems development. This facilitated the training of new staff to a productive level of systems work without incurring an extremely long training period (Maynard, J).

All of the critical systems development work was performed with the Bureau systems staff. Therefore, budget limitations had a limited effect on the project. If the development work on a project of this size is performed by contractors, changes in budget priorities and contract difficulties could adversely affect the development of the system.

The Release IIa data base, which is now in operation, contains data required by all Bureau programs. This avoids the situation of having a data base design for a special Bureau function, which may fluctuate in priority and, therefore, result in a system that is supporting a low priority activity.

The complexity of the Release IIa system was a problem because the DMS1100 data base system was released by UNIVAC during our development activities. Therefore, the DMS1100 system had not been implemented in many user sites, nor in support of a data base as complex as our design. This meant that there were no DMS1100 users that Bureau systems staff could contact for advice. Being a ground-breaking effort, Release II discovered numerous quirks and inadequate UNIVAC documentation for the DMS1100 system forcing us to become involved in trial and error activities.

A less significant problem, but still time consuming, was the retraining of our systems staff for data base maangement system development. Most of our personnel had experience on large batch systems oriented toward tape storage and sequential files. This meant that we lost significant time in retraining our people in the area of random storage data base management system concepts and the UNIVAC DMS1100 operation.

As stated before, most of the above problems arose because of the length of the project. If an agency with limited systems staff considers attempting a data base project of this magnitude, similar problems can be expected. The impact of these problems could be reduced by breaking the project into discrete steps. Each step would be implemented and placed into operation before development on the next step was started. For example, our Release IIa system, which contains three major modules, could be broken down into three steps. In this way, each development step would be less than a year's worth of effort which would avoid some of the problems with priority changes, budget, and personnel turnover.

An attempt to segment a large project into smaller steps would require good design and extensive documentation to assure that all the pieces of the system would fit together when all of the steps are completed.

BIBLIOGRAPHY

Data Base Systems Infotech International Limited, Nicholson House, Berkshire, U.K., 1975

Demonstration of a State Water Quality Management Inforamtion System. EPA Publication 600/5-74-022

Modular Programming, Maynard J., Petrocelle Books, New York, 1972

Top-Down Structured Programming Techniques, McGowan, Clement L. and Kelly, John R., Petrocelle/Charter, New York 1975

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/5-78-007		

4. TITLE AND SUBTITLE
Data Base System For State Water Quality Management Information System

5. REPORT DATE
11/77 Date of Preparation

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
John Kitch, Associated Staff

8. PERFORMING ORGANIZATION REPORT NO.

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Commonwealth of Pennsylvania
Department of Environmental Resources
Bureau of Water Quality Management
Harrisburg, PA. 17120

10. PROGRAM ELEMENT NO.

11. CONTRACT/GRANT NO.
S-801000

12. SPONSORING AGENCY NAME AND ADDRESS
Office of Air, Land and Water Use
Office of Research and Development
U.S. Environmental Protection Agency
Washington D.C. 20460

13. TYPE OF REPORT AND PERIOD COVERED
Final

14. SPONSORING AGENCY CODE
EPA/600/16

15. SUPPLEMENTARY NOTES
Appendixes will be updated preiodically.
Source Programs are available from performing organization.

16. ABSTRACT

This report describes the WAMIS Release II Date Base Management System as developed by the above performing organization. It includes System Design, Development procedures, Development Procedures, Overview of Data and a discussion of problems and recommendations. The appendixes, which are available from the performing organization, contain the System detail.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Data Base Management System Top Down Design Water Quality Information Management	Information Sciences Automatic Indexing Documentation	05/B Behavioral and social sciences

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release Unlimited	Unclassified	27
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

☆ U.S. GOVERNMENT PRINTING OFFICE: 1978—260-880/104