"a flexible data base structure becomes necessary to enhance accessibility."

Philippine Automated Water Information System: Concepts and Structure*

by

Alejandro N. Resurreccion, Jr.**

Abstract

The Philippine Automated Water Information System is a computer-based system for the storage, retrieval and analysis of water data. It is aimed to provide a systematic and more efficient means of data accumulation and dissemination. The system is divided into three components, namely: Data Management component which is responsible for the creation and maintenance of its requisite files, Data Retrieval component which is responsible in satisfying the user's request for data and the Data Analysis component which is responsible in providing the user with the results of processing and analyzing data.

Introduction

Major forces served as the impetus for developing a data system for water-related data in the Philippines. The first one was the recognition of the importance of planning. As a country in the threshold of development, the Philippines' political leadership has injected a sense of urgency in the preparation of plans. Numerous plans and project feasibility studies dealing with water had to be produced. The availability of comparable and retrievable data can significantly contribute to the shortening of the planning cycle. In its absence, aside from the difficulty of digging into historical data, planners simply gather data as demanded hence it is temporary and ad-hoc.

Water data per se are not lacking in the Philippines. In fact there are a lot of agencies, both in the private and public sectors, generating various types of water data and statistics. But the absence of a common integrating framework renders the data incomparable, confusing and oftentimes inconsistent with each other. The data are simply not organized in the form and aggregation satisfactory to diverse users because the generators are concerned mainly with their short-run needs. To overcome this difficulty and with the realization of the current growing activities in the field of water resources, the provision of a more standardized statistical information system which would eventually lead to an effective and efficient management of water resources data, was strongly felt. The management of a large

* Presented during the Symposium on Water Resources Research in the '80s, June 20-22, 1983, Quezon City. Reprinted with permission.

** Research Engineer, National Hydraulic Research Center.

amount of information and the ability of the data system to respond immediately to the needs of its users inevitably lead to a consideration of computerization.

The Origins of PAWIS

Guided by the foregoing influence factors, the National Hydraulic Research Center (NHRC), was commissioned by the National Water Resources Council (NWRC), as per a Memorandum of Agreement, to develop and execute a plan for the computerized storage and retrieval of water resources data in the Philippines. To this end, the NHRC developed the Philippine Automated Water Information System, better known as PAWIS.

The PAWIS system was formally introduced to its users in October 1980 and has been operating eversince. The operation, maintenance and continuing development of the system is an on-going project of the NHRC.

The design of PAWIS closely follows the proposal of the Preparatory Assistance Project (Leeds, Hill and Jewet, 1976) and models,in general, the National Water Data Storage and Retrieval System (WATSTORE U.S.G.S., 1975) which was established to store and retrieve water resources data in the United States.

Objectives

PAWIS has been established to accumulate all the possible water-related data in the Philippines.

More specifically, PAWIS aims to:

1. facilitate the gathering, processing, storage, retrieval and dissemination of relevant information on water;
2. organize and maintain a data processing system capable of manipulating data on hand;
3. standardize and effect quality control on acquired data prior to inclusion into the data bank;
4. keep an updated information of the various records stored in its files; and
5. define and initiate the information network by which PAWIS can serve its users most efficiently and grow steadily.

Structure

Figure 1 shows the system of information flow among the different operational units involved in the implementation of PAWIS and provides an insight on the range of services and tasks each one is to perform.

Collector

There are many agencies at all levels of government and in the private sector that are generating and producing information in one form or another. The aggregation of information from these diverse sources is the main task of NWRC so that from the viewpoint of the whole system, it acts as Data Collector. NWRC shall monitor all the data gathered from these sources and shall be responsible for assembling and documenting the data at its specified format for integration into PAWIS.

Data Coordination Center

The NHRC which developed the whole system acts as the Data Coordina-

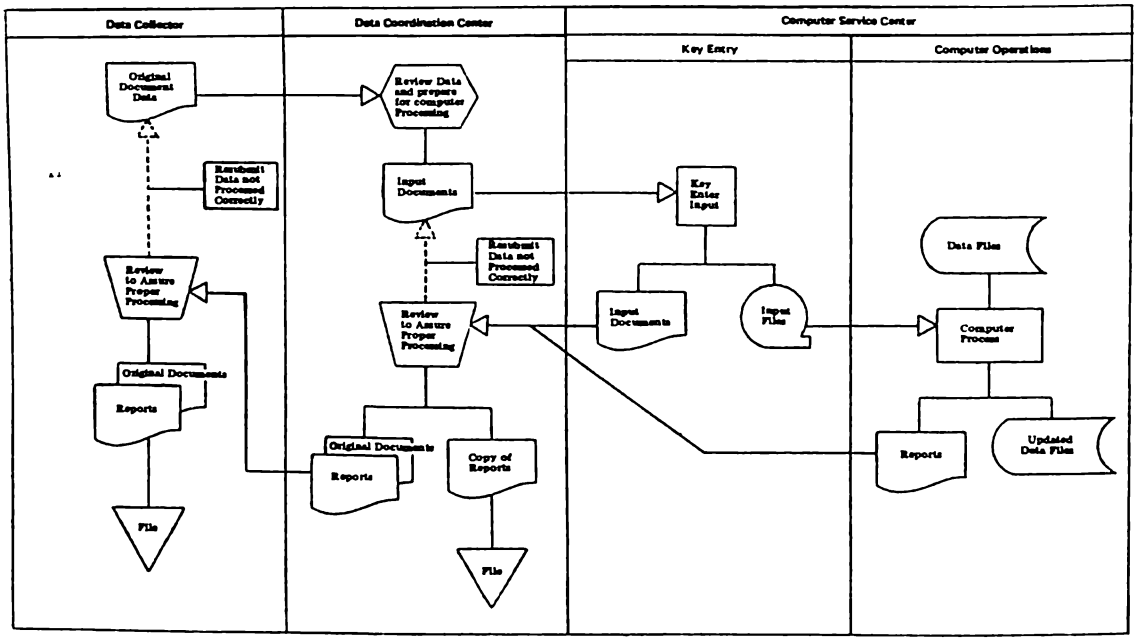

Figure 1. National Water Resources Council Data Bank
Data Bank Data Flow

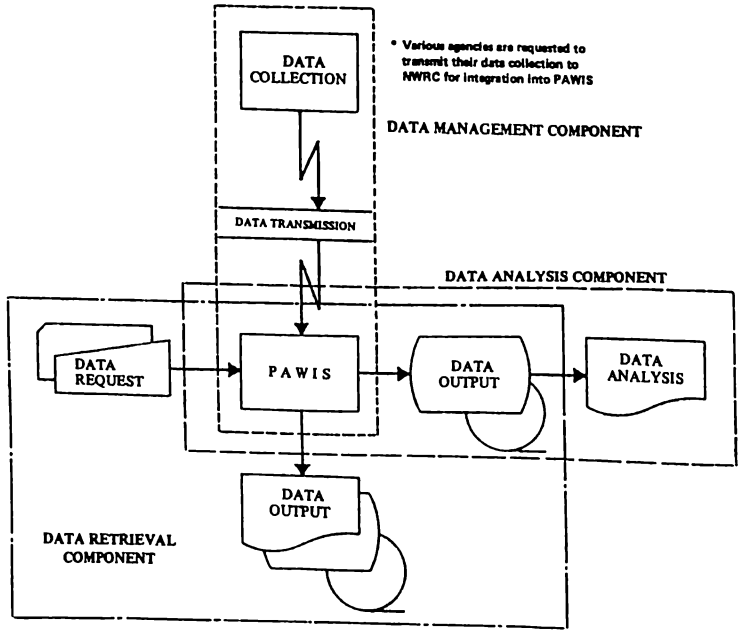

Figure 2

tion Center and is mainly responsible for monitoring both the performance of the system and the maintenance of its component programs which consist of job entry to and review of results from the data processing facilities and the correction and re-entry of erroneous jobs.

Computer Service Center

The Infrastructure Computer Center of the Ministry of Public Works and Highways has been designated to perform all the EDP operations of the system from encoding of input documents into machine readable form to processing reports in the form of computer printouts and machine readable output.

System Description

Under the program, the various agencies are requested to transmit their data collection to NWRC for integration into PAWIS. PAWIS will then provide specific data requirements to the users upon request.

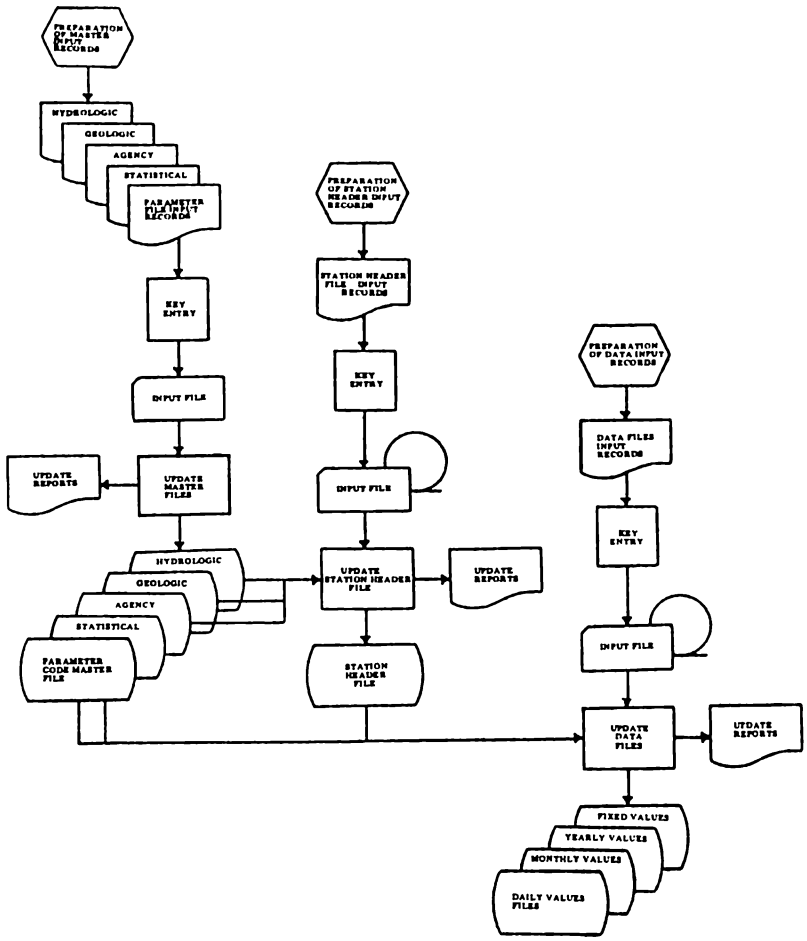

Figure 3. Data Management Component

By its very nature, PAWIS will have a huge amount of data and numerous users with varying needs. A flexible data base structure becomes necessary to enhance accessibility.

The system has three main components, namely: Data Management, Data Retrieval and Data Analysis (Figure 2).

Data Management – consists of the creation and maintenance of three files, namely: Master Files, Station Header File and Data Files. This involves adding, changing or deleting records in the said files, ensuring the validity of the data items entered, as well as reporting the contents of the existing files. Figure 3 is a systems flowchart of the Data Management component.

1. Master Files – validates and defines codes used for the various attributes of the data collection sites or of the data itself. Five files comprise the Master Files, namely:

   a. Agency Code Master File – contains the various agencies which are responsible for the operation of the different data collection sites and their corresponding agency code as assigned by NWRC.

   b. Hydrologic Code Master File – contains the hydrologic location names and their hydrologic unit codes as assigned by NWRC to the various regions, sub-regions, basins and rivers in the entire country.

   c. Geologic Code Master File – contains the various geologic unit names and their corresponding geologic codes as assigned by NWRC.

   d. Parameter Code Master File – contains the different parameter constituent as to type of water data and the corresponding parameter codes as assigned by NWRC.

   e. Statistical Code Master File – contains the different statistical natures and their corresponding statistical codes as assigned by NWRC.

2. Station Header File – is used to provide a single source of station identification information for all PAWIS data processing operations. The various information stored in the Station Header File are listed below:

   a. *Station Identification* – an identification number assigned to the data collection site which contains the region where the station is located, the principal type of data collected, the approximate latitude and longitude of the station and the agency operating the station.

   b. *Agency Code* – code of the agency responsible for the operation of the station.

   c. *Station Name* – name by which the station is known.

   d. Geologic Unit Code – code assigned to the area where groundwater station is located.

   e. Aquifer Type – code assigned to the acquifer type of the groundwater station.

   f. *Station Locator* – longitude and latitude of the data collection site.

   g. Rurban Code – code assigned to the political area where station is located.

   h. *Site Code* – code assigned to the principal type of data collected on the station.

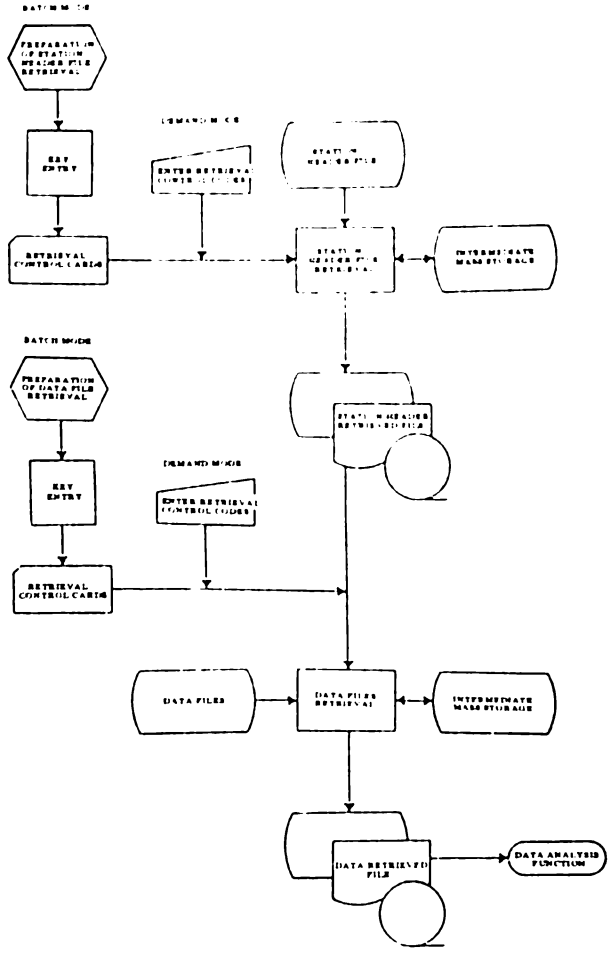

Figure 4. Data Retrieval Component

i. *Hydrologic Unit Code* – code assigned to the river where station is located.

j. Drainage Area – total drainage area of the station.

k. Contributing Drainage Area

l. Datum – gauge or land surface datum of the station.

m. Well Depth – greatest depth at which water can enter the wells of the station.

The six *italicized* items are mandatory entries to the Station Header File for each station. The combination of the six mandatory entries plus the seven optional entries makes up a Station Header record.

3. Data Files – these files will be the primary concern of any user of the PAWIS system. It contains information on water related data measurements gathered all over the country at different periods of time. The data are stored on the corresponding file depending on their frequency. No record will be stored in these files without a corresponding record in the Station Header File. The

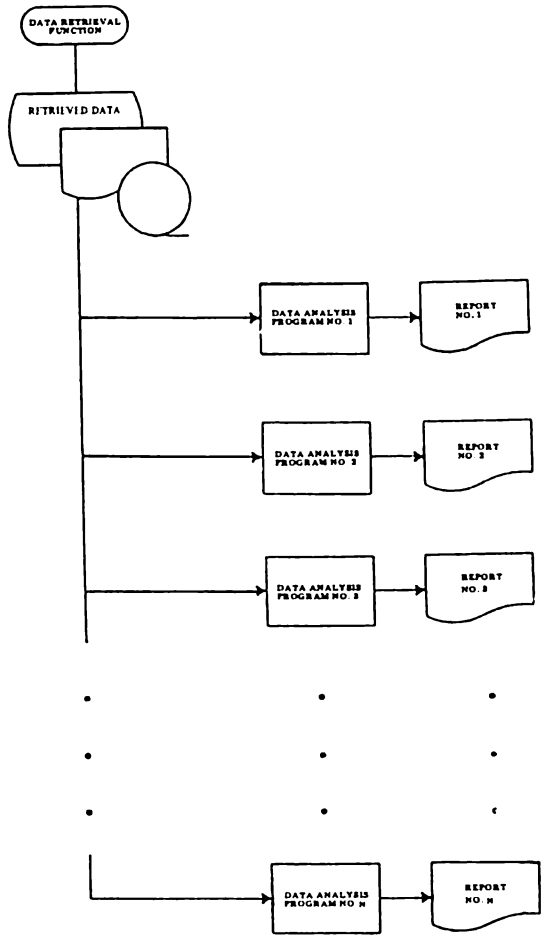

Figure 5. Data Analysis Component

different files and the various information stored on these files are the following:

a. Daily Values File

1) Agency Code

2) Station Identification

3) Cross-section Locator – provides the horizontal coordinate when multiple sampling points are at the same station.

4) Sampling Depth – provides the vertical coordinate when multiple sampling points are at the same station.

5) Parameter Code – code assigned to the type of data collected.

6) Statistical Code – code assigned to the statistical measure of the data collected.

7) Year of Data

8) Daily Values for a 12-month period

b. Monthly Values File
   1) Agency Code
   2) Station Identification
   3) Cross-section Locator
   4) Sampling Depth
   5) Parameter Code
   6) Statistical Code
   7) Year of Data
   8) Monthly Values for a one-year period

c. Yearly Values File
   1) Agency Code
   2) Station Identification
   3) Cross-section Locator
   4) Sampling Depth
   5) Parameter Code
   6) Statistical Code
   7) Beginning Year of Data
   8) Ending Year of Data
   9) Yearly Values for a ten-year period

d. Fixed Values File
   1) Agency Code
   2) Station Identification
   3) Cross-section Locator
   4) Sampling Depth
   5) Parameter Code
   6) Statistical Code
   7) Date data was taken
   8) Value

Data Retrieval – this handles the users request for data. Records stored in the Station Header File and the Data Files can be retrieved and used as data sets for the application programs. Figure 4 is a systems flowchart of the Data Retrieval component.

The retrieval of records can be done in the real-time or batch mode. It is capable of selective retrieval. Records are retrieved by supplying a list of values of a code or by specifying a range of values of a code. The retrieval procedures permit the listing of the retrieved records and/or storing these retrieved records on magnetic tape or magnetic disk which can be used as data sets for the application programs.

1. Retrieval from the Station Header File

   Selective retrieval of records from this file can be done based on the

following fields of a Station Header File record:

a. Agency Code – for a given agency code, all stations under the responsibility of the agency would be retrieved.

b. Station Identifier – for a given list or range of station identifiers, all data collection sites whose identifier is on the list or within the range would be retrieved.

c. Station Locator – for a given list of station locators, all data collection sites whose station locator is on the list would be retrieved.

d. Site Code – for a given list of site codes, all stations whose principal type of data gathered is on the list would be retrieved.

e. Station Name – data collection sites whose name is the same as the given station name is retrieved.

f. Geologic Unit Code – for a given list or range of geologic unit codes, all groundwater stations whose geologic code is on the list or within the range would be retrieved.

g. Aquifer Type – for a given list of aquifer types, all groundwater stations whose aquifer type is on the list would be retrieved.

h. Hydrologic Unit Code – for a given list or range of hydrologic unit codes, all data collection sites whose hydrologic code is on the list or within the range would be retrieved.

i. Drainage Area – for a given range of drainage area, all stations whose drainage area is within the range would be retrieved.

j. Datum – for a given range of datum, all data collection sites whose datum is within the range would be retrieved.

k. Well Depth – for a given range of well depth, all stations with a well depth which is within the range would be retrieved.

l. Polygon – for a given rectangular area, all data collection sites located within the specified area would be retrieved.

A combination of the different retrieval options of the Station Header File is possible. One can have an intersection or a union of the records retrieved using various fields for retrieval.

Records retrieved from the Station Header File can be sorted to produce an output based on any or a maximum combination of three of the following fields:

* Agency Code
* Hydrologic Code
* Station Locator
* Station Identifier
* Rurban Code

After records from the Station Header File have been retrieved, all or some of the corresponding Data Files records of the stations can be retrieved.

2. Retrieval from Data Files – selective retrieval of records from the Data Files can be done based on the following fields of a Data File record:

a. Agency Code – for a given agency code, all data file records collected by stations under the responsibility of the agency would be retrieved.

b. Station Identifier – for a given list or range of station identifiers, all data file records of stations whose station identifier is on the list or within the range would be retrieved.

c. Parameter Code – for a given list of parameter codes, all data file records whose type of data is on the list would be retrieved.

d. Statistical Code – for a given list of statistical codes, all data file records whose statistical nature of the data is on the list would be retrieved.

e. Year – for a given range of years, all data file records whose data are within the specified years would be retrieved.

A combination of the different retrieval options of the Data Files is possible. One can have an intersection or union of the records retrieved using various fields for retrieval.

Records retrieved from the Data Files can be sorted to produce an output based on any or a maximum combination of three of the following fields:

* Agency Code
* Station Identifier
* Parameter Code
* Statistical Code
* Year

Records of the Data Files can also be retrieved directly without going through the retrieval of records from the Station Header File.

Data Analysis – this will extend the services of the PAWIS system from not only providing the user with a copy of the raw data required for his need but also providing the results of processing and analyzing these data using known and generally accepted techniques developed within the sphere of water resources technology. Figure 5 is a system flowchart of the Data Analysis component.

The data analysis component involves a continuing program for the acquisition, adaptation and application of computer program packages related to water resources technology. A library of these programs is being maintained by the NHRC for possible integration into the system as data grow through the years.