Chapter 15
Data Collection and Management

15-1. General

a. Water resource studies tend to be data intensive. One reason is that the physical systems involved are often large and complex (e.g., watersheds, precipitation fields, river-reservoir systems, etc.), and substantial quantities of data are required for their representation. Another reason is that the investigations themselves are complex, with a variety of interdependent computational elements (e.g., precipitation-runoff simulation, and statistical, systems, and economic analyses, etc.). The transfer of data generated with one element to another is a significant requirement in such investigations.

b. The acquisition, processing, and management of data can require a substantial portion of the resources allotted for an investigation. Performance of these tasks in an efficient and reliable manner can be of considerable value. This chapter describes aspects of data management.

15-2. Data Management Concepts

a. Figure 15-1 illustrates components of a data management system for a water resource study. Elements of the system include a data loading module for entering data from various sources into the management system, "application" programs that read information from and write information to data storage, and utility programs that perform functions such as data editing, displaying data in graphical and tabular form, and mathematical transformations of data. With such a system, basic data can be loaded into storage, reviewed, and perhaps edited with utility programs. Interdependent application programs can be used to perform the analysis, using the data storage to pass information generated with one program to the next. Utility programs can be used to prepare summaries of results in various forms, including report-quality tables and graphs.

b. Common data types for flood-runoff analysis include individual element, time series, and paired function. Individual element data include items such as basin properties (e.g., drainage area, percent imperviousness, soil types), values for runoff parameters (e.g., unit hydrograph, kinematic wave parameters, baseflow, loss rate), structure dimensions, inventories of gauge types and locations, etc. Time series data consist of values of a variable for sequential points in time such as discharge and stage hydrographs, precipitation hyetographs, air temperature records, etc. Paired function data are sets of interrelated variables for which each value of one variable is paired with a value of another, such as discharge elevation, exceedance frequency-stage, reservoir storage-elevation, etc.

c. There are a number of commercial data bases that are well suited for the storage of individual element data. Such data bases are relational and permit queries such as "list all gauges within specified latitude and longitude bounds," or "list all subbasins for which the soils are in soil group A." Whether or not it is desirable to utilize a data base for individual element data depends on the data type and the frequency of use intended for the data.

d. Hydrologic studies generally make extensive use of time series data. Data bases that are designed specifically for this data type gain efficiency by treating such data in blocks (i.e., groups or sets) rather than as individual data items. Storage and retrieval is performed a block at a time. Block size might be, for example, one month for hourly data. A system designed for use with time series data is the data storage system (DSS) developed by the Hydrologic Engineering Center. It is configured as in Figure 15-1, with a number of water resource application programs having the capability to communicate with DSS files. A comprehensive set of data loading and utility programs supports the system.

e. Paired function data are also widely used in hydrologic studies. An advantage of central storage of discharge-elevation rating "curves," or any paired data, is that changes made to the data need to be made in one place only, and all application programs that use the data will have access to the revised data. The DSS system is designed to accommodate paired function data.

15-3. Geographic Information Systems

A powerful data management tool for spatial (i.e., geographically oriented) data is the geographic information system (GIS). Such systems enable the storage and retrieval of information that is associated with spatial elements such as rectangles, triangles, or irregularly shaped polygons. Variables such as slope, orientation, elevation, soil type, land use, average annual rainfall, etc. can be stored for each element. The data may then be retrieved and tabulated, displayed graphically, or used directly by application programs. Several commercial GIS's are available.

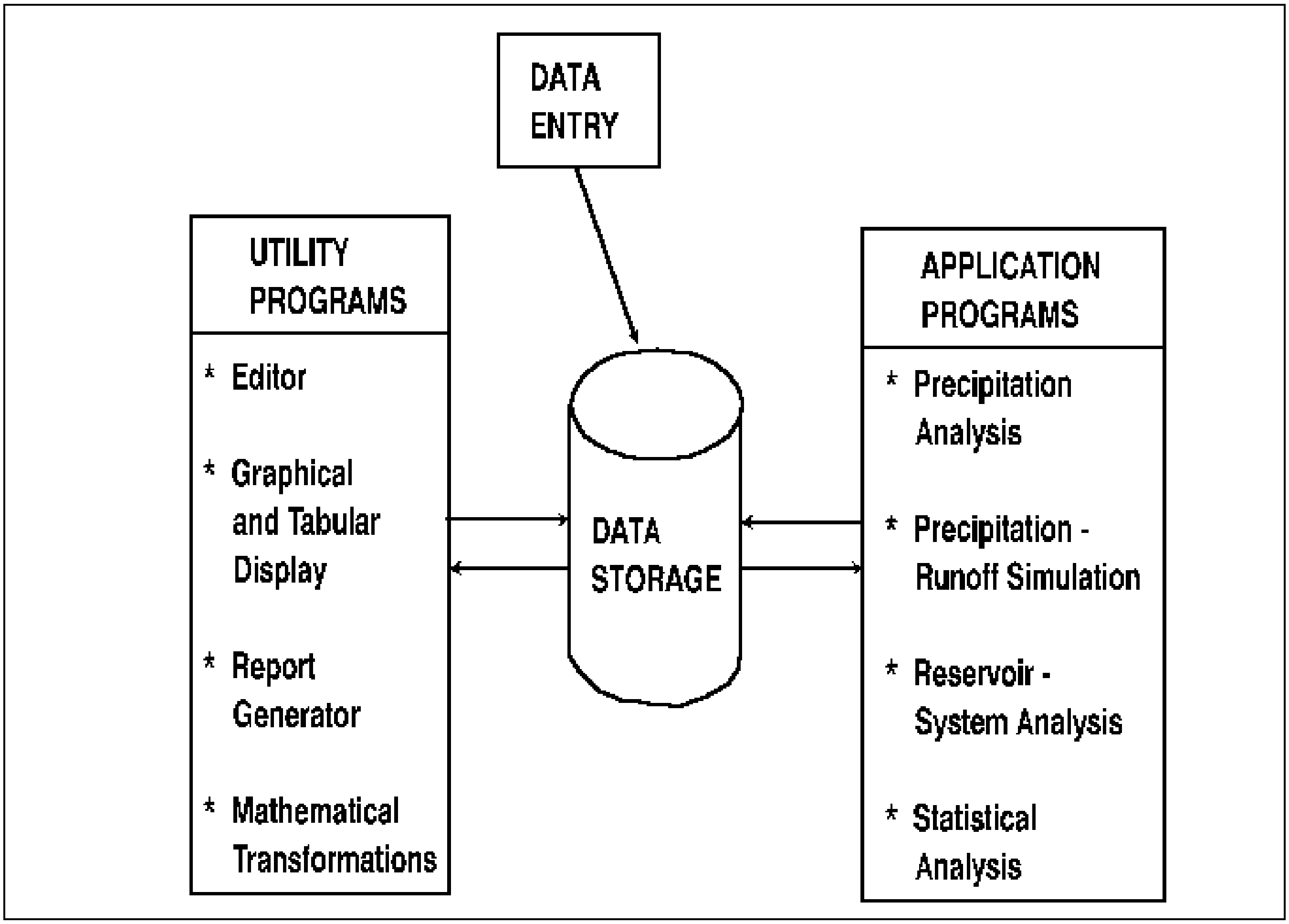

Figure 15-1. Data management

15-4. Data Acquisition and Use

a. The use of data typically involves the following: based on the purpose and scope of the study, determine the types of data that will be required; determine the sources and availability of the data; acquire and process the data; perform the analysis; and archive the data and study results. The first two steps are components of study formulation. As stated previously, the type, amount, and quality of data available for a study can have a significant impact on the choice of methodology and reliability of results.

b. Data for hydrologic investigations are generally obtained from several sources. For example, streamflow records are commonly available from the USGS, and daily and hourly rainfall values from the NWS. Also, commercial firms obtain such data from collection agencies and make them available in useful form (e.g., on compact disk). Various formats are used to encode the data, and these must be interpreted as part of the process of loading data into a data base. There are a number of data loading programs associated with DSS, including programs to read data formats used on commercially available compact disks, NWS data formats and U.S. Geological Survey WATSTORE formats. In addition, software is available to read the Standard Hydrometeorological Exchange Format (SHEF), which is accepted as a standard for data exchange by a number of agencies. The function of the loading programs is to read data in the appropriate format and enter that data into a DSS file. After the data has been loaded, utility programs can be used to graph or tabulate the data and perhaps edit or apply transforms to it (such as stage to discharge).

c. Application programs that have the capability to access data storage must be “told” what and how much data to retrieve. Such instructions are part of program input, as are instructions specifying the calculated information that is to be written to data storage. The capability to review application program results in tabular or graphical form with utility programs can be very powerful in facilitating the performance of a study. Final results can then be produced in report-quality form.

d. Upon completion of a study, data and study results should be prepared for long-term storage. Because formats used in specific data management systems may change over time, data should be stored in a system-independent format. For example, information from a DSS file can be transferred to a text (ASCII) file.