THE INGEST, QUALITY CONTROL, AND PROCESSING OF HYDROMETEOROLOGICAL DATA AT NATIONAL WEATHER SERVICE FIELD OFFICES

Mark J. Glaudemans
AWIPS Software Group Leader
Hydrology Laboratory, Office of Hydrologic Development,
National Weather Service, National Oceanic and Atmospheric Administration
Silver Spring, MD

INTRODUCTION

The primary mission of the National Weather Service (NWS) is providing hydrologic forecasts and warnings for the protection of life and property. An additional mission component is to provide data for a national information database and infrastructure. Both mission components are completely dependent on timely and reliable hydrometeorological data, in a form usable by internal applications and external users.

As part of the Advanced Weather Interactive Processing System (AWIPS) operating at each of the NWS field offices, there exists a set of software components for handling hydrometeorological data. These systems ingest and decode the incoming data, perform quality control on the data, monitor the data for indications of potentially hazardous conditions, and post the data in a relational database from which applications can use the data, and distribute and disseminate the data to external customers. These data processing and storage systems provide the foundation for the Hydrology program at NWS Weather Forecast Offices (WFOs) and River Forecast Centers (RFCs). The robust, application-independent design of the database architecture and data processing infrastructure facilitates development of local applications.

This paper describes the flow of hydrometeorological data in the NWS field office from its entry point into the local AWIPS systems, to the processing and storage of the data in the respective database(s), and to its exit to external systems. The focus is on the data operations of WFOs, with some discussion given to additional data operations of the RFCs.

OVERVIEW

At its simplest level, the data flows as a constant stream of operational data decoded and stored in a database, residing alongside existing parametric and reference data. Applications use the data and eventually, the operational data becomes out-of-date and are purged from the database. To allow multiple applications to access the data and to provide an efficient, manageable method for storing and retrieving data, a relational database is used to store the data. The WFO Hydrologic Forecast System (WHFS) which serves the WFO Hydrology program, makes extensive use of this database, referred to as the Integrated Hydrologic Forecast System (IHFS) database (Glaudemans, 2002).

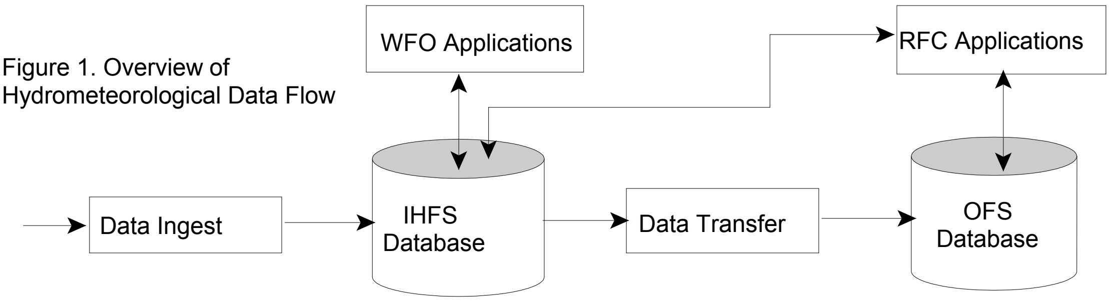

Figure 1. Overview of Hydrometeorological Data Flow

River Forecast Centers operate the NWS River Forecast System (NWSRFS), a large-scale river modeling system that was designed long before the advent of commercial relational database systems. Most of the data used by RFS applications are stored in random access, indexed file systems which comprise the Operational Forecast System (OFS) database. The data in the IHFS are transferred to the OFS database using a regularly-run data transfer function. RFC applications use the OFS database as their primary database, with some applications accessing the IHFS database. This high-level flow of data is depicted in Figure 1.

DATA INGEST

A large, steady stream of hydrometeorological data is ingested into the AWIPS IHFS database. These data are received via AWIPS communications mechanisms. The bulk of the data is sent via the Satellite Broadcast Network (SBN), which provides point-to-multi-point data communications between a central Network Control Facility (NCF) and field sites. The data sent via the high-bandwidth SBN include satellite data, model data, text data, graphic data, and point data. The Wide-Area-Network provides point-to-point and point-to-multi-point communications between all sites and has a much smaller bandwidth than the SBN.

While some radar data is distributed nationally via the SBN, field offices receive a large amount of their radar data via dedicated radar modem lines which connect the local AWIPS system to nearby radars. The radar data products are ingested by the WHFS and stored in the IHFS database by the DecodeDPA application. The primary radar product of interest is the Digital Precipitation Array (DPA) product, which provides hourly precipitation estimates updated every 5-6 minutes.

A significant amount of hydrometeorological data are received via the SBN from the Hydrometeorological Automated Data System (HADS), which is a real-time data acquisition and data distribution system that processes data from approximately 8500 data collection stations measuring environmental data at locations throughout the country. The HADS system receives the raw binary sensor data, decodes it, and transmits it to field offices via the SBN in the Standard Hydrometeorological Exchange Format (SHEF) (NWS, 1998). Another important source of data is the METAR-encoded reports from the over 900 Automated Surface Observation Systems (ASOS) data collection sites scattered around the county, primarily near airports. These METAR reports are converted into SHEF format by the WHFS Metar-To-SHEF application.

Each NWS office has local data collection systems that provide critical data not available on the national SBN. The Local Data Acquisition and Dissemination (LDAD) system is connected to these local data systems, which are usually operated by local governments. LDAD communicates with the systems using the appropriate protocols, often using two-way polling capabilities to request data. The collected data are decoded from their local format and encoded into SHEF. All SHEF data, regardless of its source, are decoded and posted into the IHFS database (Office of Hydrological Development, 2000b).

Products generated at NWS offices are made available to all other offices. Examples of this are river forecast products and flash flood guidance products generated by the RFC that are sent to the WFOs via the SBN. All AWIPS sites receive all products available on the SBN; an office-configurable product identifier filter is applied at the front end of this data stream to minimize the processing of unwanted products.

DATA STORAGE

The success of the NWS hydrology program depends in large part on the data processing and software applications provided to the NWS forecaster. These applications, in turn, depend completely on the availability and usability of real-time operational data and relevant reference data. The storage of the these data has evolved considerably over time and today includes the IHFS database and the OFS database (Roe, 1998). The IHFS and its associated features are discussed in the following subsections, followed by a brief discussion of the OFS database.

IHFS Database: The IHFS database is the centerpiece data repository for the WHFS applications which run at both the WFOs and RFCs, and for select RFC applications. It also serves as the front-end database feeding the data to the OFS database used at RFCs. The IHFS database provides an enterprise rather than an application-specific approach to data management, whereby data are stored in a logical structure dependent upon the data, not on any given application. The database design follows modern database design rules by using a “normalized” form for data structure and relationships, in which no data are duplicated and data dependancies are established as part of the data structure. In a few aspects, the IHFS database table design is not in the “normalized” form, because the normalized form resulted in unacceptable performance. In these cases, a de-normalized form is employed that results in greatly improved performance.

The physical structure of the IHFS database consists of tables and views, database constraints such as primary and foreign keys, table indexes, and database procedures and triggers. The database is documented in detail using Entity-Relationship Diagrams, Data Flow Diagrams, Data Dictionary, and Release Change Notes available at the following web address: http://hsp.nws.noaa.gov/oh/hrl/ihfs/database/html/databases.htm.

The database contains 168 tables, each of which have primary keys to prevent the table column(s) data from having duplicates. Foreign keys are also used to restrict the referencing values in specified columns to those referenced values in the same column of another (i.e.

foreign) table. There are indexes placed on selected columns in large tables for which fast retrievals are important. Because of the (mostly) normalized form of the database design, data are contained in one and only one table, even if the data relate to multiple entities. To allow for quick retrievals of data that are spread among multiple tables, there are 19 data "views" in the database, which act as virtual tables. There are 11 stored procedures which process and distribute data using an algorithm defined in a stored procedure language. Lastly, 6 triggers are defined, on 3 separate tables, to "trigger" some specific action upon an insert or update of data in the table. Typically, the action taken is to invoke one of the aforementioned stored procedures.

The IHFS database contains these major categories of data: operational observations and forecasts for stations; operational radar and precipitation analysis grids; reference data for stations, areas and polygons; and application and database control parameters. In the WHFS, the HydroBase application is the primary means by which reference and parameter data are managed, while the HydroView program provides the primary viewing and editing tool for the operational data. Each of the data categories are discussed in the following subsections.

Operational Data: The operational data stored in the IHFS consists of station data and gridded data. The operational data dominates the IHFS in terms of using the most space, as the reference data and control parameters occupy relatively little space. The efficient storage of the station observation and forecast data is critical for the proper execution of all applications. The storage of this data is based on the SHEF data model, where the database table columns correspond to the fields defined in the SHEF-encoded data.

With each station data value are stored the following SHEF attributes: the station identifier, the time of the observation/forecast, the physical element (e.g. temperature or river stage), the duration (usually instantaneous, but non-zero for elements such as precipitation), the type (e.g. observed or forecast), the source, the extremum (i.e. whether it's a minimum or maximum), an external qualifier code, and a revision indicator code (i.e. whether the value has been revised). For forecast data, additional attributes indicate the time of the forecast, and any associated probability for the forecast value. Non-SHEF attributes associated with each value include the quality code indicator, the time the value was posted, and the time and identifier of the data product that contained the value.

A summary of the station data stores is shown in Table 1. The physical storage of the SHEF data is based on the value's type code, which indicates whether the data value is either Observation, Forecast, Processed, or Contingency. Processed and Contingency data are each stored in a table dedicated to that SHEF data type. Because Processed data are generally always for observations, an option exists to store the data as Observation data. In practice, the Processed and Contingency data represents a very small fraction of the station data. Processed data is not used often, and Contingency data are used primarily to store Flash Flood Guidance (FFG) values.

Operational Station Data

Observed			Forecast	Processed	Contingency
Height/Stage	Moisture	Pressure	Height/Stage	--	--
Discharge	WaterQuality	Radiation	Discharge		
Precipitation	Lake	Agriculture	Precipitation		
Temperature	Ground	Weather	Temperature		
Snow	GateDam	Wind	Other		
Ice	Y Codes	Power			
Evaporation	FishCount				

Table 1.Operational Station Data Stores

The Observation and Forecast data are not stored in a single table for each type; rather, each type of data is distributed into separate tables based on the value's physical element code. Specifically, the Forecast data are split into the following tables: Forecast Height, Forecast Discharge, Forecast Temperature, and Forecast Precipitation. These four forecast data types represent the bulk of the forecast data processed. Any other forecast data are grouped together in a single table simply called Forecast Other. The observed data are stored in one of the following tables: Agricultural, Discharge, Evaporation, FishCount, GateDam, Ground, Height, Ice, Lake, Moisture, Power, Precipitation, Pressure, Radiation, Snow, Temperature, WaterQuality, Weather, Wind, and Y Codes. Distribution of the observed and forecast data into the appropriate tables is performed by stored procedures, allowing applications to easily distribute data by invoking the stored procedure.

The IHFS database stores all data that can be encoded into SHEF, and SHEF is capable of representing essentially all of the hydrometeorological point data. Besides these general methods for storing any point data, the IHFS database includes a collection of supporting data tables. There is a table for storing the raw text products which contain the encoded data, and a table for tracking the linkage between products and the station data within them that complements the product information stored with each value. The only text products ingested are those that pass through the product filter. Within a product, there are stations or specific station data that are of no interest to the local office. An additional filter is applied to filter out stations and or specific physical element-type source data from further processing. If desired, "unknown" station data can be stored in a separate table for later review. A Comment Value table exists for storing any text comments that SHEF allows to be associated with each value. As part of the WHFS quality control operations, data identified as invalid can be stored in the Rejected Data table. The Alert Alarm Value table stores any data that are identified as indicating hazardous conditions as part of the WHFS Alert/Alarm operations. Lastly, to allow quick retrieval of mission-critical data, there is a Latest Observed Value table for storing the latest data, and a River Status table for storing the latest observed and maximum forecast values for river stations.

Gridded data containing precipitation grids are stored in the database in a special manner. For performance and space reasons, the database tables only contain information referencing each grid; the actual gridded data are stored in binary files. A column in the database table provides the name of the file containing the data, although in the event of the normal case where there is no rain, then no file is generated since it would simply be a large grid of all-zero values.

Reference Data: The IHFS reference data are the data sets that define the stations, areas, and their static attributes. Stations contain a large volume of reference information, including detailed descriptions of the station location, data sensors, reporting characteristics, geophysical qualities, historical events, interested parties, cooperative observers, etc. For stations at river or reservoir locations, there are additional reference data sets. Much of this type of reference data is itemized in the NWS Form E-19.

Geographic reference data include coordinate and other information for areas such as hydrologic basins, counties, and NWS zones, and for vectors such as rivers, streams, highways, and roads, and for point entities such as cities, towns, and radar locations. A set of data tables also exist for describing the physical, hydrological, and dam-break scenario data for dams, in the event of a catastrophic failure of the dam.

Parametric Data: The IHFS application control parameters dictate how the operational data sets are processed. Some of these parameters are specific to individual stations, such as the thresholds used for the quality control operations and the alert/alarm monitoring operations. The Ingest Filter information is the primary method by which data for known locations is filtered by posting and processing applications. Other parameters control how various programs operate; the user can change these parameters to control program behavior, such as computational methods or colors for displays. For many fields in the database whose values are limited to a certain domain of possible values, lookup tables are used to contain the valid values.

File Storage: For WHFS data that does not fit into a relational model, a file-based storage method is used. The RiverPro product formatting application contains template phrase file and product content control files, and the Time Series application uses a configuration file that allows local control for pre-defining time series graphs to generate. As mentioned, the radar precipitation estimates are also stored in binary files. Lastly, essentially every program and script in the WHFS generates a log file, which are stored in file directories. A job that cleans out log files is typically executed four times a day to prevent the excessive buildup of these files.

OFS Database: The IHFS database contains observational data which is transferred to the OFS database by a transfer operation that executes regularly, typically on an hourly or sub-hourly basis. The transfer program writes the data into the Pre-Processor database, which is one of the four primary databases in the OFS. Using the operational data in the Pre-Processor database and the processing instructions in the Pre-Processor Parametric database, a set of programs generates time-series of data that are written to the Processed database. The Processed database is then used by the hydrologic and hydraulic river forecast models to generate the forecasts stored in the Forecast database. Other databases such as the Calibration database and the Extended Streamflow Prediction database are used for special purposes at the RFC.

DATA PROCESSING

The IHFS database has evolved over the years to be a rather comprehensive data storage entity capable of storing every type of hydrometeorological data set. It has grown more rapidly than the applications which use the data for performing the NWS mission. New and enhanced applications are being regularly added to the WHFS application suite to exploit the incredible volume of information embodied in the data sets. As new applications are developed, the relational database is well-suited to evolve with the needs of the applications. A few of the application features that relate to the database design and data flow are discussed in this section.

Program Controls: The numerous applications which use the IHFS data are designed to be flexible to the user needs, whether the applications be interactive or non-interactive, graphical or command-line driven. The control parameters in the database allow applications to be configured for different offices, for different users, for different scenarios, etc. Some classic example of parameters are display colors, or quality control and alert/alarm thresholds.

For those parameters which can not be readily modeled, the hydrometeorological applications use a “token” file concept. A token file is simply a list of tokens and their alphanumeric values; a single file contains tokens for all applications. An application that needs instructions or information on how to perform a function references this global token file and derives its value; if a value is not found for some reason, the application has an application default value it can use. The token file is actually a collection of three hierarchical files, where the application first checks an optional user token file, and if the token is not found, checks an office-configurable file. If the token value is still not found, then it checks the nationally-provided token file which is always available.

Quality Control: The IHFS database has many facets of its design for support of various Hydrology program requirements. These include management of station versus gridded data, management of reference and parameter data with logical relational links, support for alert/alarm monitoring. One particularly noteworthy area of its design is the quality control infrastructure for station data.

As part of the IHFS quality control model, every data value in the IHFS database has an associated quality control code. Applications test the quality of the data value and store the results in the bit-encoded quality code attribute. Initially, the value of the quality code attribute is based on the data qualifier code defined externally via the SHEF encoding of the value. Often, this data qualifier is not used, in which case the initial quality code attribute is set to “good”.

The model employs a three-tier strategy, in which the quality of a value is defined as either Good, Questionable, or Bad. Bad indicates that the data value is known with certainty to be invalid, and Questionable indicates that the value has failed at least one test that makes suspect the validity of the value, but is not known with certainty to be Bad. If a quality code value indicates Good, then the value is either truly “good”, or the value has not been identified as being Questionable or Bad by failing either an internal test or being identified by external sources as being “bad”. Incoming data are tested and if the data fails the quality test, the quality code is set accordingly. Bad data is either commingled with the regular data, or is directed to a table for storing rejected data (i.e. a "trash can"), because on the local preferences.

The bit-encoded nature of the attribute allows the results of up to 20 tests to be stored in a compact manner. As data are ingested, each value is checked against pre-defined thresholds to determine whether the value is within gross limits, or is within reasonable limits. If a value is outside the gross or reasonable range, the quality code is set to indicate that it failed the test and the value is considered Bad or Questionable, respectively. Regularly-scheduled rate-of-change checks are performed that compare the numerical difference between two instantaneous values, separated by a known time duration, against a pre-defined threshold. If the threshold is exceeded, then the quality code is set to indicate the value is Questionable. Other checks can be easily added, either as a local application developed using the published knowledge of the quality code infrastructure, or as a nationally provided application.

As applications access data, they are instructed to access (or ignore) the data based on the quality level within which the value lies or on the results of specific tests. The WHFS applications provide comprehensive tools for the user to review and redistribute the data based on its quality. These tools can move data between the "regular" data tables and the Rejected data table. There are user interfaces to review all data in detail; they allow the user to filter data so as to view only Questionable or Bad data, to view only Rejected Data, etc.. The data can be deleted or edited. Any modified data are marked as being manually edited, which by the rules of the WHFS quality control model, is assumed to indicate a Good value.

External Distribution: The hydrometeorological data stored within AWIPS probably represents the most complete repository of high quality operational data for the local NWS office area. While much of the data are received from national sources, a large amount of data is received locally from the LDAD system and from cooperative observers. All these data are quality controlled and can be provided to external users, including other government agencies, the private sector, the global community, and of course, the American public. The data values and their associated attributes can be encoded into SHEF or other formats by the WHFS or locally-developed applications. The data can be distributed via LDAD or via the Message Handling System (MHS) features of the AWIPS Communications Network (ACN). Once the data are available on the ACN, they can be provided to any interested party. The data are also available for presentation on office web pages and can be formatted into public products by WHFS applications. Other methods can also be employed to allow the external customers to benefit from these valuable data sets.

Local Applications: Nationally developed enhancements and additions are being made to the application suite which uses the IHFS. However, the needs of the NWS offices are often so unique that the national software cannot possibly meet all the local needs. The development of applications by local offices is encouraged and aided by the presence of the well-documented and designed database IHFS.

In addition to the detail design documents for the database itself, there are operation guides available for the SHEFdecode application which manages the complex data posting into the

IHFS, and for the Quality Code operations (Office of Hydrologic Development, 2000a; Office of Hydrologic Development, 2000b). Software libraries for handling the many general and specific needs of data handling are also available for distribution to local offices.

SUMMARY

The operational requirements for hydrometeorological data of the NWS Hydrology program have resulted in the evolution of a well-designed relational database that serves the WHFS applications used at WFO offices and provides data to the OFS applications at RFC offices. This formally documented database contains station and gridded operational data, and reference and parametric data. A high volume of data are ingested from a variety of sources and in a variety of formats, posted into the database, and made available to the scientific applications and distributed to the public and external customers. The database will continue to adapt to the changing needs of the hydrology program to meet the goals of the NWS mission.

REFERENCES

Glaudemans, M.J: R.A. Erb; E.B. Wells; J. Zimmerman; J.J. Hill-Maxwell; K.S. Mack, 2002: Overview and Status of the Hydrologic Forecast System in the National Weather Service Weather Forecast Offices. 82th American Meteorology Society, Annual Meeting, Interactive Symposium on the Advanced Weather Interactive Processing System, Orlando, Florida, January 13-17, 2002.

National Weather Service, 1998: Standard Hydrometeorological Exchange Format Version 1.3; National Weather Service, March, 1998.

Office of Hydrologic Development, 2000a: IHFS Quality Code Operations Guide, September, 2000.

Office of Hydrologic Development, 2000b: SHEF Decoder Operations Guide, September 2000.

Roe, J., G. Bonnin, M. Glaudemans, C. Gobs, and P. Tilles, 1998: Recent Database Developments at the National Weather Service Office of Hydrology. Special Symposium on Hydrology, Phoenix, AZ, Amer. Meteor. Soc., Jan 11-16, 1998.