CHAPTER 10

DATA STORAGE, ACCESS AND DISSEMINATION [HOMS G06]

10.1 INTRODUCTION

The availability of sufficient good quality data underpins all aspects of hydrology, from research, to water resources assessment through to a wide range of operational applications. The definitions of "sufficient" will vary from one application to another, and are discussed elsewhere. Discussed in this chapter are the issues of "good quality" and access to data, as well as availability of data to a wide range of users.

10.1.1 The importance of data

The culmination of data collection in hydrology, from precipitation measurements, water-level recordings, discharge gaugings, to groundwater monitoring and water quality sampling, provides a data set that can be used for decision-making. Decisions may be made directly from raw data measurements, derived statistics or the results of many stages of modelling beyond the raw data stage, but it is the collected data that form the basis for these decisions.

Raw data – whether field forms, charts or reports – must be available after processing. Some errors in reporting and processing may not come to light until scrutinized by users. It may also be necessary to check transcriptions from the original or to re-assess the collector's interpretation of a doubtful trace. Records from a particular site may need to be verified by re-sampling in response to development activities, or changes in technology may result in an upgrading of standards. In either event, the data may require reprocessing. Thus, the original data must be securely archived. The storage should be kept away from the electronic database and should be physically secure.

A data set is clearly of great value as it is inevitably collected through a huge commitment of time and money. The management of these data is therefore important work in itself and this work must be performed effectively in order to maximize the results of this investment. A well-established and well-managed hydrological archive should consolidate the work put into data collection to provide a source of high-quality and reliable data for tens or hundreds of years into the future. A poor-quality archive, due to lack of forethought in its foundation or poor management, can lead to years of excess data collection or modelling work, and subsequent poor decision-making. The archive could become redundant within a very short time. Moreover, poor-quality data and databases will result in suboptimal planning decisions and poorly designed engineering structures.

Data integrity is also a major issue. Often the key to understanding the success or limitations of work in all fields of hydrology is an understanding of the quality of the data on which this work is based.

Of course the scale of data management depends on the scale of the operation: a large-scale and detailed hydrology project will require more complex management techniques and computer storage than a smaller scale project measuring a limited set of variables over a short period of time. Other factors affect the scale on which data are managed; as well as volumes of data there are often budgetary constraints, where limited staff time can be spared for archiving and funds are unavailable for large data management systems. The capacity of staff to manage data can be a constraint and the level at which data management is performed is dependent on the experience or skills of the personnel.

Despite the potential variety due to scale, however, there are a number of essential aspects that are common to all hydrological data management systems. This chapter as a whole describes each of these aspects of data management in detail, with a focus on the general approach, and occasionally highlights the reality at each of these extremes of scale.

10.1.2 Data management processes

There is a definite flow path that hydrological data must follow from the point of collection, as input into the system, through validation to dissemination and use in decision-making processes. This path is essentially the same regardless of the scale of operation and the level of technology used for data management, and is demonstrated by the schematic diagram in Figure I.10.1. The list in Table I.10.1 summarizes some of the data sets existing at various stages of the data management process. This is an overview of the entire data

management process. Some of the aspects of this process are discussed in detail in Chapter 9. In the present chapter, the focus is on data storage, access and dissemination and demonstration of where they fit in the data management process.

A full description of the recommended procedures for storage and cataloguing of climatological data is given in the *Guide to Climatological Practices* (WMO-No. 100). While hydrological data require a somewhat different treatment for storage efficiency, many of the same considerations apply.

The vast quantities of climatological and hydrological data being gathered by many countries may preclude storage of all original data. However, copies can be made in media (for instance, electronic scan) that require a small fraction of the space required for the original documents; the original materials may then be discarded. Storage conditions for any of the media on which data are stored should minimize deterioration of stored records by excessive heat, temperature fluctuations, high humidity, dust, insects or other pests, radiation and fire.

Where possible, duplicate sets of records should be kept, one in the main collection centre and the other at a regional centre or at the observer's office.

The various types of input data and the data-processing and quality-control process applied to them are described in Chapter 9. Input data can originate from manuscript observer records, chart recorders, automatic data loggers and manuscript sheets or digital files storing instantaneous discharge measurement (gauging) information, including river section, depth and velocity profiles, often with associated descriptive text information.

Table I.10.1 summarizes the processes involved in data management from inputting raw data measurements through to disseminating processed data, as well as the data involved in these processes.

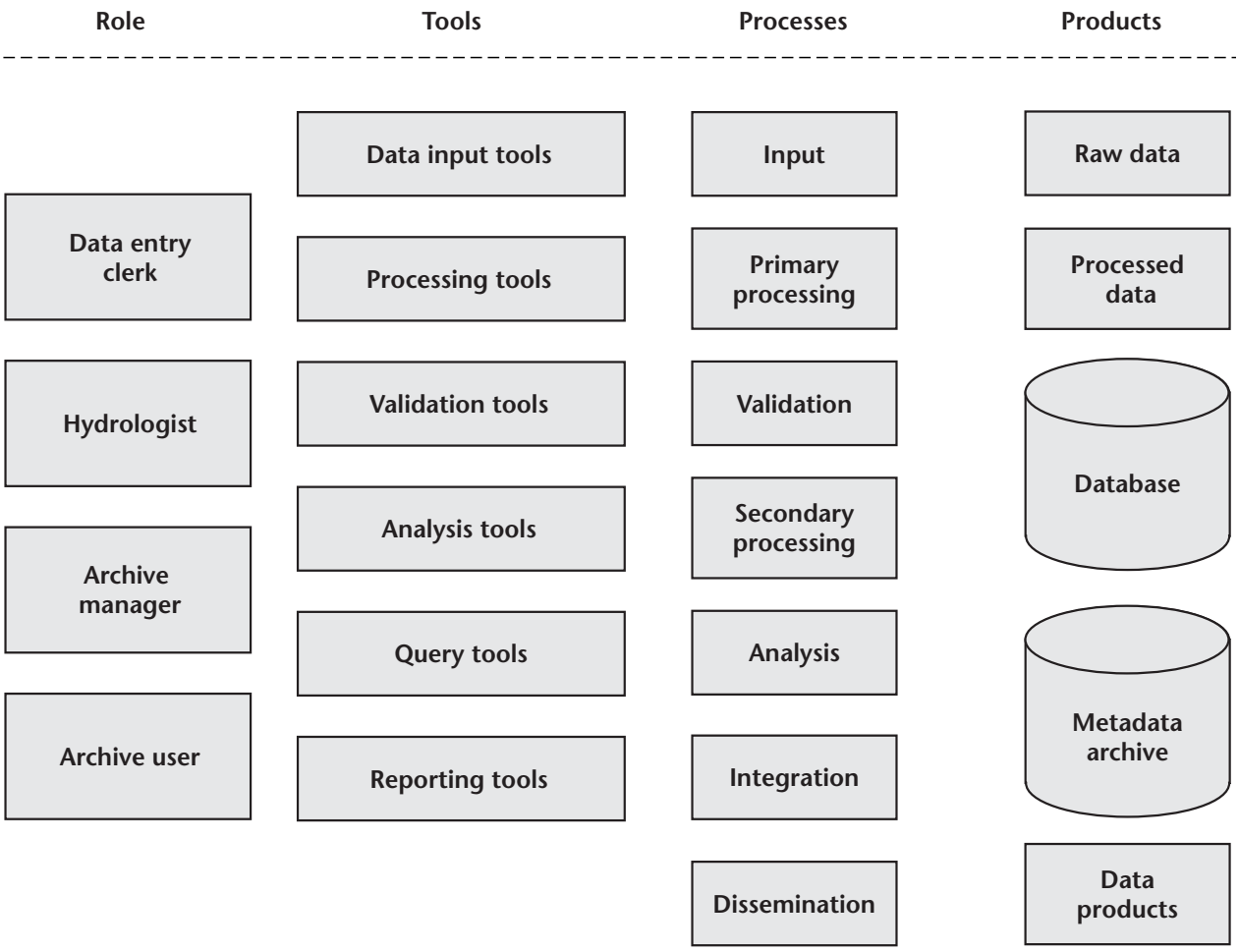


Figure I.10.1. Data management scheme

Table I.10.1. Processes involved in data management

Process	*Description*	*Examples of data type involved in the process*
Input	Data are obtained from the source or instrument, via manual recording, data logging, digitization or other method, and transformed to the appropriate format for storage.	Gauge readers' book/sheet (daily or sub-daily stage read manually from a staff gauge, with additional notes to describe any necessary context)
		Raw data file from automated data logger linked to measurement probe – often in binary/custom format
		Charts from water-level recorder
		Sheets/digital files storing instantaneous discharge measurement (gauging) information, including river section, depth and velocity profiles and descriptive information
Processing of raw data	Data stored in manual or digital forms, often both	Digital files of data digitized from charts
		Data series within databases or digital files of manually entered gauge records and gauging information
		Digital files of data series converted from custom logger files, usually text files
Validation	Raw data are checked for erroneous values and edited to produce a corrected raw data series.	Data series within database at each stage of editing/validation
		Text describing changes including methods used and reasons for editing
Secondary processing	This stage includes infilling of missing data where appropriate, conversion to secondary intervals (for example, calculations of mean and total series), creating rating curves from new discharge measurements and conversion of raw stage data to flows or reservoir storage.	Data set for each new series created in conversion process
		Rating equations created from discharge measurements, and associated text describing decision-making process in creating rating equation. Rating equations will generally change over time, and so a good rating history must be maintained.
Security and archiving	Data must be archived in such a way as to be accessible but secure, and well documented and indexed.	Metadata allowing quick and easy access to data sets, as well as a full index of the information available
Integration with other data	Allowing display of data with other sources of data such as GIS data sets	New datasets such as spatial rainfall coverage derived from point raingauge data or groundwater surface maps from borehole data
Data dissemination	Data are distributed as required and in appropriate forms to modellers, decision makers, public bodies, etc.	Summaries of data, yearbooks, etc.
		Increasingly, data dissemination is via websites.

It is clear that there is potentially a large number of data sets at each of the stages of the data management process, and decisions must be made concerning which data should be stored, and how to do so in an effective hydrological archive. A description is provided below of the storage of data, analysis of these data and the production of information, access to the information and the dissemination of the information to the variety of users.

Having completed the data-processing and quality-control phases described in Chapter 9, the data will be archived in a form that will have associated with it a range of analysis and product derivation tools, as well as access and dissemination tools.

10.2 DATA STORAGE AND RETRIEVAL

10.2.1 Storage of data

One of the most important considerations within the field of hydrological data management is which of the numerous data sets produced should be stored. There are many stages in the process of data management from recording to dissemination and each of these stages can represent one or more distinct data sets. If one were to store every possible permutation of data in this process, the result would be a confusing and unwieldy archive. At the other extreme, if a hydrological archive existed only as a static data set of processed and validated data, there would be no means of understanding how the data had been derived or measured, or the potential limitations of the final data set. For example, a processed flow data set provides no information about the means of measurement, the process of deriving flow data from water-level data, or whether the data had been edited and how this editing was performed. It is therefore necessary to decide on a feasible data storage mechanism that falls somewhere between these two extremes.

The basic consideration for the level of detail of data storage is that of reproducibility. In any hydrological project, however large, it is necessary that the steps from raw data to final processed data set be understood, and reproduced if required. Users of the processed data should be able to quickly and easily see the process through which data have passed, and understand the potential limitations. This does not imply that every change to the data set must be retained for posterity, rather that raw data sets should be retained, and that changes and assumptions made during validation and processing should be documented and stored. It is also important that users of the data be able to differentiate between original data and data that have been added to fill gaps, or edited data.

Once again, the level at which this data storage is performed will be determined by a number of factors, such as available storage space, availability of funds for storage and documentation, and the availability of staff. There will inevitably be a trade-off between completeness of archive and resources spent. At the most complex end of the spectrum, a large and complex hydrology project may use a data storage system that allows fully automated auditing of all changes to data sets in the system, storage of dates and times when changes are made as well as identification of the user making them, and allows edits to be sequentially rolled back to recreate any version of the data set that existed. A simpler system may contain merely the raw data set and the final data set with a file of notes documenting the decisions and edits made. However, in both cases the process is essentially the same:

(a) Raw data files must be kept, whether these are in a hard format (gauge reader books, recorder charts) or digital format (raw data logger files or telemetered data);
(b) All processed data sets should be associated with descriptive metadata records detailing the origins of the individual data set and linking each to the data set from which it is derived;
(c) Important stages of data processing should be stored, even if the processed data are only an interim stage between raw data and disseminated data. The decision about importance will be determined by the scale of the data management system. For example, if a raw water-level data series is to be converted to a series of monthly mean flows only, it would be sensible at least to store the validated water-level data set, and the derived daily flow data, in addition to the raw data and final monthly mean flow;
(d) Wholesale changes to parts of each series should be documented against the data set, for example, noting the application of a datum to a period of a stage record, or conversion of a period of a stage record to flow with a rating curve, which will itself exist as a data set;
(e) Changes made to individual data values, for instance interpolating missing data values or editing values separately, should be documented against each data value changed, with notes against the record as a whole to indicate to the data user that the series has been altered;
(f) The resulting data set will then have a comprehensive catalogue of what has been edited and why, allowing any data user to be able to both understand the reasons for and methods of changing raw data values, and to reproduce the data set from the raw data.

10.2.2 Storage methods

One of the most important considerations when archiving data digitally is the database to be used. The term database is often used misleadingly, and in hydrological circles and elsewhere is often used in reference to both the database system itself and the software for querying the database and displaying and analysing the data.

Both of these are important aspects of any archive individually and will be dealt with separately in this chapter.

A database can be described simply as a filing system for electronic data. Any organized assembly of digital data is, in effect, a database. Several important aspects of these assemblies define which database is most appropriate in a particular case, relating directly to the principal concerns of data managers set out in 10.2.1.

10.2.2.1 Important criteria for data storage systems

When developing data storage systems, a number of important criteria must be considered, including:

(a) Security – this includes management of access and administrative rights for the various users;
(b) Ease of maintenance;
(c) Costs, including initial outlay and recurrent costs including any software licences required, maintenance and storage;
(d) Ease of query;
(e) Power of existing data query tools;
(f) Ease of development of additional query tools;
(g) Ability to include/link to other data sources or data display software, such as GIS;
(h) Suitability alongside existing information technology (IT) infrastructure/requirements and staff capabilities;
(i) A metadata system that provides adequate information on the data in the database;
(j) Ability to allow networked/remote access – linking to network and web servers.

Of course each instance of a hydrometric archive will have different levels of importance for each of these aspects and again the extremes are depicted. The advanced requirements of a large national network, such as automated, real-time data loading, links to sophisticated analysis tools and multi-user access from numerous distributed organizations in turn require substantial expensive technical support, training of users and often bespoke development of tools. The database must be run securely on a high specification machine and be able to be backed up automatically to tapes in a fireproof safe. A small project database may need to be operated by a single hydrologist. In this case loading, editing and analysis of the data must be simple operations that can be quickly learned. The resulting database may need to be small enough so that it can be sent by e-mail to other users. A small nation's hydrometric archive may be data essential to the country's social, environmental and financial future, but may, of necessity, have to be run on a very limited budget. Collecting data is expensive and money spent on overly specified computing systems might detract from the purpose of the archive: that of measuring and publishing good quality hydrometric data. However, a database must be sustainable: secure, simple to manage on the available infrastructure, whilst providing the necessary tools.

The types of database (here, electronic data management systems) can be divided into the following categories.

10.2.2.2 Simple ASCII files

The simplest type of database could be a set of ASCII files containing data, indexed on a PC or network drive. A separate file could be used for storing data for a particular time series, perhaps with a separate directory for storing the data for each station. Advantages of such a system are that it costs no more than the computer on which data are stored, is very simple to set up, with little or no knowledge of computers, and that files can easily be found, with the text format allowing any user to read data immediately, and to store any sort of data that can be subsequently read. Disadvantages include the obvious insecurity of the system, the limitation of a single-user storage system, the lack of existing analysis and graphing tools for data, and the difficulty to develop tools to work with the data. However, many organizations still maintain such a system, which could be considered appropriate for a small company storing copies of data that are archived elsewhere when no analysis is required, security is not an issue but low maintenance is of paramount importance.

10.2.2.3 Bespoke database formats

Many data storage systems, particularly those developed before the surge in computer technology in the late 1990s, use their own format for storing data. These are often highly compressed formats allowing large volumes of data to be stored on the small disk spaces that were available then. In addition, writing customized methods for accessing data from specific formats can allow very fast retrieval and saving of data, as all of the overheads associated with allowing generic data access are avoided. Besides these advantages of bespoke systems, an organization that has compiled its own database has also developed a considerable body of knowledge and would be able to cater efficiently to its needs in data storage and viewing, as well as analysis tools. Disadvantages include the inability to interface with or incorporate other available technologies (a feature of more generic systems) and the cost of maintaining the developed tools as the platforms and operating systems on which they

run evolve. In addition, there is the risk of the organization relying on detailed in-house knowledge, which may lead to difficulties if there is no system of knowledge transfer, or this knowledge somehow becomes lost.

10.2.2.4 Relational Database Management Systems

Relational Database Management Systems (RDBMS) are, as their name suggests, more than just databases. They are generally a specific file format for data storage (the database itself) together with management protocols and software access tools. The most complex of these can feature integral query, reporting, graphing and publishing tools. Several well-known RDBMSs are available on the market. They are well used around the world and are therefore well tested and supported, both by the vendors and the users. Skills for development of additional tools are readily available. The security, level of support, availability of query tools, price and the like vary between systems.

10.2.2.5 Specialized hydrometric database systems

The database systems described above are just that: generic tools for storing data. They must be adapted by the user to meet his or her specific needs. It is most likely that the needs of almost all hydrologists can be fulfilled by a specialized hydrometric database system. These are essentially off-the-shelf software products (although in some cases there may still be considerable work involved in installing the software), which can be purchased or otherwise acquired. They mostly comprise a database system, of one of the types described above, which has been adapted to cope specifically with common types of hydrological data. For instance, database tables and access routines that specifically handle hydrological data types and store appropriate descriptive information and metadata may have been created. Commonly the system is provided with software to allow the data to be managed, edited and graphed, and this software is much easier to use than the database itself. In addition many of the data management tools are extended to include analysis tools, for instance, tools to produce flow duration curves from flow data, and statistical tools for fitting distributions to flood peaks. These database software systems include HYDATA, HYMOS, TIDEDA, HYDSYS and WISKI.

The pros and cons of these systems fall into the categories mentioned in 10.2.1. The commonly available systems generally vary in scale. The smaller systems can be easier to install and run and are less expensive to buy and maintain. Larger scale systems tend to be more expensive, but have more advanced functionality, and are often built around a larger scale database with increased security, although this can often have associated licence costs for the database software. The choice of system therefore depends on user needs, and ability to purchase and maintain a database system.

10.2.2.6 Database management skills

Databases can be managed by a single person or by teams of many people, but the processes performed generally require particular skills, which determine the role of the person involved in the process. Some of these skills are listed in Table I.10.2.

10.2.2.7 Summary

To summarize, there are numerous types of digital data storage systems. While most hydrologists' needs would be filled by one of the specialist hydrometric database systems available, some advanced users may have additional requirements that would be better met by a customized database system. Most importantly, a database system is purely a means of storing digital data. Good archive management can only be achieved by good management of data.

Table I.10.2. Database skill requirements

Role	*Description*
Data entry clerk	Little understanding of hydrology or IT required, though data may often have to be downloaded from data loggers or extracted from different file formats
Hydrologist	Validation work requires expert knowledge of hydrology and the local hydrological regimes. Analysis work requires expert hydrological knowledge.
Archive manager	General archive management and dissemination requires hydrological knowledge. Integration of archive data with other processes requires both hydrological and IT skills.

10.2.3 Types of data and information to be stored

This section describes in more detail the particular information that should be stored in a hydrological archive. Perhaps the best way to consider how an archive should be arranged is to imagine approaching an archive with no prior knowledge of the meteorological conditions, sizes of rivers, catchment characteristics, gauging station network, water use within catchments or volumes of data. It should be possible to quickly gain an understanding of the entire contents of the archive, then to easily retrieve exactly the data required. Archive users should be able to fully appreciate all of the changes that have been made to data and should be able to retrieve information on data availability, summary statistics and full data sets quickly. This allows users to start work on a data set at any point in its management process with minimal effort. In addition, the archive system should make the documentation of work a simple and efficient process. The production of further data sets, for inputs into models, into further data processes, for distribution to separate data users, or for the creation of publications such as yearbooks, should also be a simple and quick process for the archive manager.

10.2.3.1 Archive metadata

On viewing a hydrological archive, the first level of data seen by a user should be data describing the archive itself. These are in fact metadata – information about the archive itself that should actually be published by the data manager as a means of disseminating information about the archive. These data could take the form described in Table I.10.3.

These archived metadata could be provided by a complex computerized system, perhaps with a GIS interface to allow access to the data and automatically updated summaries of data availability through which the user could browse, or this could be as simple as a folder of papers, which is the responsibility of the archive manager. In the case of the latter, the folder should be regularly updated as new stations or new data are added.

10.2.3.2 Station metadata

When users of the archive are familiar with the data holdings they will require further information. Station description data are important to provide the context within which the station operates. These data can also provide a shared resource for data users to understand, for instance, the implications for the data of measurement devices used, or the morphological setting of the station, and for the station management staff to store information concerning location of station, access information, datum and addresses of local operating staff. Most of the data in Table I.10.4 can relate generally to meteorological stations, gauging stations or other measurement sites, though some fields are specific to hydrological river gauging stations.

Information on the development of the WMO core metadata standard can be accessed at: http://www.wmo.int/web/www/WDM/Metadata/documents.html.

A summary of the status of hydrological metadata systems can be found in Global Runoff Data Centre, Report 31 (Maurer, 2004).

Table I.10.3. Descriptions of data held in archive

Data type	*Description*	*Examples*
Archive description	Brief text describing the background and aims of the data monitoring project	Name and description of the project, start date of the project/archive, aims of the project, summary of dissemination routes
Geographical maps	Maps providing the physical context for the archive's data	Catchment boundaries, gauging station/meteorological station location and other data measurement locations, river network, lakes and other features of importance
Data summaries	List of data sets and availability	Summary, by data type, of data held in the database, referencing locations of measurements, plus additional data held, for example, derived spatial data and GIS data from other sources, plus a summary, for each data set, of data availability over time at an appropriate scale

Additional information on metadata database standards is available from the following Internet sites:

– USGS – Federal Geographic Data Committee's "Content Standard for Digital Geospatial Metadata: http://www.fgdc.gov/metadata/metadata.html
Dublin Core Metadata Element Set, Version 1.1 http://dublincore.org/documents/dces/
– ISO 8459-5 Information and documentation – Bibliographic data element directory – Part 5: Data elements for the exchange of cataloguing and metadata:
http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=27176&ICS1=35&ICS2=240&ICS3=30

A number of examples of metadata for hydrological systems are available and can be accessed at:

– Global level: http://www.watsys.unh.edu/metadata/

Table I.10.4. Examples of station metadata

Metadatum	*Description*	*Examples*
Identification	Current identification information for the station, and summary of the purposes for which the station is used	Station name(s), station number(s), catchment name, waterbody name, hydrometric area name, elevation, catchment area, primary purpose, secondary purpose, primary measurement method (for example, weir type), high flow measurement method, general description of station
Location	Information about the geographical position of the station	Latitude/longitude (or position in local coordinate system), nearest town/landmark, benchmark location and height, information about landowner, routes, accessibility, appropriate access time, information concerning access in flood periods, etc.
Operator	Information about the organization operating the station, if operated by another, for example, regional organization	Operator name, contact details, responsibilities, etc.
Observer	Information about the staff taking measurements at the station	Observer name, contact details, responsibilities, starting date, frequency of visit, reporting method and interval
Station history	Description of the history of the station, showing any changes that may affect data measured	Date opened, date closed (for closed stations), location history, operator history, equipment history, datum history
Equipment/telemetry	Information describing any data loggers or automated telemetry systems used at the station	System names, manufacturer, purpose, reference for associated literature, date installed, antenna heights, etc., reporting interval and frequency, parameters reported, additional descriptive information
Statistics	Summary statistics of data at the station	Statistics, values, period to which statistics apply, date calculated, etc.
Graphics	Pictures of station and surrounding area	Picture, description, date, references to digital pictures files, etc.
Data set history	Information describing the data sets produced for the station	Parameters measured, derived series produced, flow path for data measured at the station, summary of data availability
Gaugings and ratings history	Descriptive information concerning instantaneous discharge measurements and rating equation development – actual gauging data should be stored in database	Description of river section(s) used for gaugings, history of changes to section through movement of section/erosion, etc., drawings of section, description of issues

– National level: http://www.epa.gov/Region8/gis/data/r8_hyl.html;
– State level: http://www.isgs.uiuc.edu/nsdihome/webdocs/st-hydro.html; http://www.wy.blmgov/gis/hydrologygis-meta.html

10.2.3.3 Time-series data

The majority of data used in hydrology is time-series data, the measurement of a variable at a fixed place over time, including rainfall, streamflow, water level, reservoir storage, borehole water level, soil moisture and pH. At a single station (or geographical location) there are often multiple time series of data measured and each of these may have different characteristics. Each station should store a summary of the time series measured at the site (data set history in Table I.10.4) and the attributes of time series should be noted against each of these (Table I.10.5).

10.2.3.4 Real-time data

Data collected by some form of telemetry and required for real-time operational use, for example, flood forecasting, reservoir operation or monitoring low flows for ecological purposes, may have to be archived and accessed in a different system than that collected for regular monitoring or long-term assessment of water resources. Such telemetered data generally must go through some fairly simple data-validation process before being archived for input to real-time models. Such validation may be as simple as checking that each incoming data value is within pre-set limits for the station, and that the change from preceding values is not too great. Thus, 15-minute rainfall data must always be a positive number, but less than the highest ever recorded 15-minute rainfall for the region of interest plus perhaps 10 per cent. River-level data must also be greater than bed level or crest level of the gauge weir, and a suitable maximum value can generally be set. In addition, from analysis of previous major flood events, an appropriate maximum rate of rise for any 15-minute period can be established. Where data fall outside of these limits, they should generally still be stored in the raw data file, but flagged as suspect, and a warning message displayed to the model operators.

Where suspect data have been identified, a number of options are available to any real-time forecasting or decision support model being run:

(a) The suspect data could be accepted and the model run as normal, although this is rarely a reasonable option;

Table I.0.5. Time-series data characteristics

Data field	*Description*
Name	Ideally, a time series should be appropriately named allowing instant recognition of what is stored, for example, daily mean flow or monthly total flow as opposed to fow series 1 or flow series 2
Time-series type	Data being measured, for instance rainfall, flow or water level
Measurement statistic	Indicates the derivation of the data, or the statistic being stored: mean, instantaneous, total, maximum, etc.
Unit	Indicates the unit in which the data are being stored
Interval	Frequency at which measurement is made, or period over which statistic is calculated, for example daily, monthly, every 15 minutes. Irregularly recorded data are also considered and often denoted as an instantaneous time series.
Period of record	Start (and sometimes end) date of the data series
Limiting statistics	It is often recommended that an initial estimate of the maximum and minimum data values for a series be set before measurement commences as a means to validate the data. This is particularly useful if automated validation methods can pick out values outside the recommended range. Following extreme measurements, these limiting statistics can be reset more accurately. A level for maximum rise and fall within the series can also be useful if the methods used for data validation can usefully use these statistics, as can setting limits to more complex derived statistics. These statistics should be for guidance only, as data outside these limits can be valid and should not be excluded.
Further time-series level data	Other information that can apply to a time series as a whole can also be stored at this level, such as the water day of measurements (indicating which period derived average and other values should be calculated) and datum if this is appropriate to the entire period of data.

(b) The model can be run treating the suspect data as missing, that is, assuming that there has been no further rainfall during the period in question, or having no observed river level and flow data against which to test a forecast flow;
(c) The missing data could be substituted with some form of backup data. Thus missing river levels could be extrapolated from previous values, and missing rainfall data could be infilled by reference to other operating gauges, or mean seasonal values assumed.

How missing data are dealt with will vary from situation to situation and will depend upon the modelling requirements. The topic of modelling is dealt with in Volume II, Chapter 6.

10.2.3.5 Spatial data

Almost all of the data discussed above are either descriptive metadata or time-series data of measured attributes. A further type of data is discussed here. Spatial data are data that have a substantial geographical component. Examples include maps of gauging station sites, digital elevation models and isohyets of rainfall. Spatial data can be displayed in GIS and these are often used to integrate hydrological and spatial data sets.

Geographical features are represented in GIS coverage in various forms (Figure I.10.2):
(a) Polygon – data exist as shapes of areas, such as countries or basins;
(b) Line – data appear as lines with associated attributes, for example, rivers;
(c) Point – data exist as individual points, for example, river-gauging stations and raingauges;
(d) Grid – region is divided into grid squares and the attribute (for example, rainfall) over this square is stored together with other attributes.

The characteristics of these geographical features are called attributes, for example, each polygon of a geological coverage may contain attributes such as lithology or aquifer type.

For the purpose of this discussion spatial data for hydrology can be divided into two simple categories:

Physical maps

Physical maps are an invaluable resource in hydrological studies and still constitute the principal source of spatial data in many countries. They can include specialist maps, such as those showing soil coverage, geology or rainfall, or they can be national maps showing multiple feature types such as towns, roads, contours and rivers. Physical maps should be considered a central part of a hydrological archive, and are a useful first point of reference providing valuable contextual to actual gauging station and meteorological data. They should be stored accordingly, ideally in appropriate map chests, racks or mountings. The map archive should be well documented, including:
(a) Map reference numbers and originator/source;
(b) Map title and description;
(c) Scale;
(d) Projection;
(e) Number and name in map series;
(f) References.

Dissemination of physical map data can be difficult, as there are often restrictions due to copyright. However if photocopies of maps, or portions thereof, are to be disseminated, the above information should be included to allow maps and legends to be traced and interpreted.

If physical maps are created in an archive, for instance runoff maps from gauging station flow data, the existence of these maps, and the details, should be published in the appropriate place. This could be in hydrological data books or through national or regional mapping agencies.

Digital data

Over the past 10 years or so there has been a widespread move from physical maps to digital maps. Developments in technology have allowed maps to be digitized and used in GIS. Within these systems it is far easier to manipulate and integrate maps. It is also easier to extract information from them and disseminate changes.

Many digital maps may be simply digitized versions of physical maps. For instance, the contours on a normal general usage map could be digitized to a line coverage, or a map of soil types could be digitized to a polygon coverage. As with any data management process, the origins of data, as well as any edits made, should be carefully documented so that users of the resulting data can be aware of its origins.

In addition, digital maps can be created. For instance, a gridded coverage of rainfall data can be created from point source raingauge data by the use of various processes. Contour lines, when accurately digitized, can be extrapolated to create a

digital elevation model grid of topographic heights. Using contours or a digital elevation model, gauging station catchment boundaries can be manually added to a new line coverage. If such derived maps are stored as part of a hydrological archive, the same precautions regarding reproducibility mentioned in 10.2.1, should be followed. Each derived map should have associated archived metadata describing the process used to create it. Any significant and useful intermediate data sets created should be archived as appropriate.

10.2.3.6 Management considerations

When managing hydrological data and information it is important to include the following:

(a) Validation or quality-control flags (9.8 and 9.9);
(b) Text comments from users/data processors (9.7 and 9.8);
(c) Audit trail – information on the introduction of data to the database and any following changes or adjustments (Chapter 9).

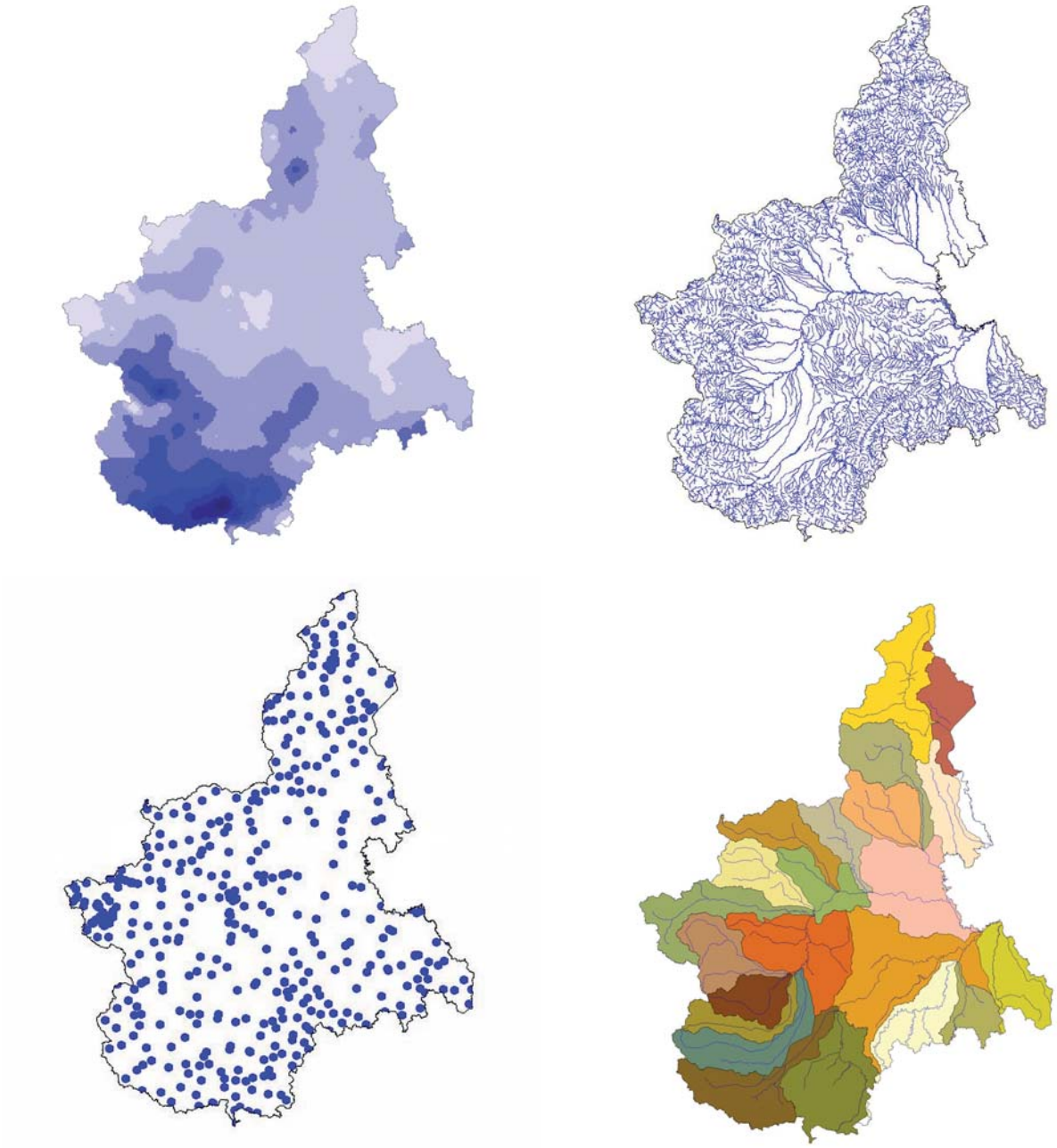


Figure I.10.2. Examples, clockwise from top left, of gridded (rainfall), line (rivers), point (gauging stations) and polygon (catchment) data

These types of data and information should also be stored and made readily accessible.

10.2.3.7 Controlling data flow

The importance of proper control of incoming data sets has already been mentioned with regard to data-entry operations. The need to be aware of the status of all data sets in the various stages of validation and updating is equally vital. This is particularly true when suspect data have been queried and some response is awaited from hydrological staff in charge of quality control.

Initially, the entire monitoring process may be manual, but ultimately some functions may be automated as part of the general computerized data-handling activities.

Automation allows routine monitoring of data-batch status, validation summaries and the physical disposition of data on the system, for example, the tape or disk volume numbers and the data set names. Such control is essential where large quantities of data are handled.

Data-control personnel should be appointed with the following responsibilities:

(a) Logging of incoming data batches and the routing of these batches to the appropriate data-entry system;
(b) Monitoring and logging data-entry status and the subsequent submission of data for initial validation and processing;
(c) Routing validation reports to appropriate hydrological personnel and receiving edited data;
(d) Repeating steps (a) to (c) until all data batches have been accepted for updating purposes;
(e) Forwarding monthly and annual summary statistics to appropriate agencies and personnel.

The exact nature of the tasks depends upon the extent to which individual users have access to data for editing purposes. In online systems where users are responsible for their own quality control, central responsibilities are reduced. However, such users must have some means to indicate that quality control has been completed and that data sets are ready for further processing.

10.2.3.8 Updating procedures

Most archival databases in hydrology are updated in at least two stages. These stages are shown in Figure I.9.2. The first stage is the cycle of monthly updates corresponding to a standard reporting period. The extent to which the four first-stage activities are split into separate computer runs is dependent upon the user and the physical resources of the system. If most files are archived on tape, it probably would be impossible to perform the complete set of monthly processing with one program because too many tape drives would be required. There may also be a policy not to compute derived values, for example, flows or potential evapotranspiration, until all the basic data have been checked manually.

For the end-user, the main outputs from this first updating phase are the monthly summary reports. For database management purposes, the most important results are the updated annual workfiles. If this first phase system handles data only in monthly blocks, it may be necessary to maintain incomplete data files. This need arises from the use of computer-compatible recorders, where the recording medium is normally changed at irregular intervals. Thus, when processing month 1, there may be several days of month 2 on the recording medium. In this case, the month 2 data are saved on a temporary file until the complementary data are available during month 3. The cycle is repeated, and a complete month 2 file and a new incomplete month 3 file are generated. This problem is rarely encountered with manual reporting or telemetry stations. If the computer-compatible medium requires pre-processing, there exists an option to perform the splitting and subsequent aggregation of months on the pre-processing (micro) computer before data are submitted to the main processing machine.

After passing the validation checks (and being subjected to any necessary primary processing) (Chapter 9), the monthly data batches are added to the current annual data file. Data not passing validation checks must be manually scrutinized, and, where errors are verified, relevant actions must be taken as indicated in Figure I.9.2.

In order to provide an adequate turnaround of data, it is generally necessary to start processing each monthly data batch from the tenth to the fifteenth day of the following month. If processing is not started by this time, there is a danger that the total data handling, entry and processing for the annual file updates may become backlogged.

The purpose of the annual updating cycle is to add the annual workfile to the historic database. This transfer carries with it a change in status of the data from a working data set to a quality controlled hydrological reference set. Thus, it must be ensured

that as many data queries as possible are resolved before the annual updating takes place. Output from the annual processing stage may be utilized for hydrological yearbooks.

10.2.3.9 Compression and accuracy

A significant operation in all database updating is the compression of data to make optimal use of storage space. The technique of packing is described in the *Guidelines for Computerized Data Processing in Operational Hydrology and Land and Water Management* (WMO-No. 634). However, packing techniques tend to be machine specific, and several other data-compression techniques are used in various hydrological database systems. These are:

(a) Integer numbers are used in storage, which are suitably scaled for output purposes. For example, daily rainfalls, measured to a precision of 0.1 mm, could be stored in tenths of a millimetre (an integer) and subsequently divided by 10 for output. The storage requirement is halved. A normal integer uses two bytes of storage compared with the four bytes required to store a real (decimal) number;
(b) The use of unformatted (binary) data files instead of normal ASCII files. In addition to requiring less space, binary data are more rapidly stored and retrieved;
(c) The use of a counter for repeated constant values. Thus, a period of 10 days without rainfall need not be stored as a set of 10 zeros, but as a repeat factor of 10 followed by the zero value;
(d) A more sophisticated version of the above method is to completely remove redundant data. Redundant data are derived from the over-recording of hydrological phenomena by some types of field instruments, in particular, by fixed interval recorders. For example, in the sequence 40, 50, 60, it is apparent that the central value can be derived by interpolation from the adjacent values. Thus, software can be developed to scan data, eliminating all those values that may be linearly interpolated within a defined tolerance range. This technique greatly reduces the storage requirements but leads to no significant loss in the information content of the data. In New Zealand, the use of the TIDEDA system (HOMS component G06.2.01) has resulted in two- to twelvefold savings in storage space;
(e) The use of relative rather than absolute data values. For example, water level in a borehole may be quoted in absolute elevation terms or, more economically, in relation to some local datum or average water level. Only the difference from the previous data value need be stored. These various forms produce smaller numbers that may be stored in correspondingly smaller storage locations. Some balance must be made in the levels of data compression employed. Increasing efficiency in the use of storage is gained at the expense of executing compression and expansion routines each time the data are stored or retrieved. The correct level of data compression should reflect the relative limitations of storage space and computation capabilities, and software development skills, at each installation. With regard to the accuracy of the data stored, it is exceptional for any hydrological data to be observed to an accuracy of greater than one part in 1 000. For this reason, many hydrological databases store data only to an accuracy of three or four significant figures. Thus, a flow computed as 234.56 $m^3 s^{-1}$ may be stored as 235. Such a practice is also used to save data storage space.

10.2.3.10 Physical-file organization

Sequential file organization is simple, may be used on all forms of storage medium and is suited to time-series data that are input and most frequently accessed in a sequential manner. Indexed sequential files are very attractive for the storage of most hydrological data as the inherent sequential nature of the data is preserved on the storage medium, but the ability exists to access directly individual, or groups of, records.

Random-access, like indexed-sequential organization, is only relevant to disk or diskette files, but requires higher system overheads in terms of storage volumes. Individual records may be accessed directly and more quickly if they are accessed in a random manner. By the use of cross-references (pointers), data in random-access files may be related in complex and effective ways.

If a hydrological database is being developed to support online (interactive) data manipulation, files must be available on disk, and the use of indexed-sequential or random-access files should be feasible. Indeed, their use is probably essential to obtain acceptable response times when handling large amounts of data.

Where online data access is not a priority, it may be worthwhile to keep single variable time-series data, such as water levels or rainfall, on sequential files because they are usually searched to abstract a time sequence of data. For multivariate time-series files, there are some advantages in indexed-sequential or random-access organization.

If a certain variable is measured at a few stations only, then all stations will need to be searched to locate the values in a sequential file. In some types of random-access file, it is possible to store a pointer with each variable value, and the pointer indicates the location of the next station record that contained a value for the same variable. This location could then be accessed directly. Such a technique is advantageous for water quality data where the variables observed vary widely both between stations and for the same station at different times.

Data stored on magnetic tape, the most common format for large database archives, must be held in a sequential manner. However, when files are transferred from tape to disk, any of the range of access methods described above may be used. Whichever access method is used, it is recommended that all large database files be unformatted (binary).

Some database systems utilize a mixture of techniques to maximize storage and retrieval efficiency. This is done by storing large groups of sequential data in single records of random-access or indexed-sequential files. By using this method, each daily or even hourly station-year data may be stored as one physical record in a random access, or indexed sequential, file. To retrieve the data for a given month, the relevant station-year record may be accessed directly on the disk. This record is then transferred to an in-memory buffer from which the data for the correct month may be rapidly read. Some mention should be made of the use of database management systems (DBMS). These systems invariably rely on the use of random-access files. Some caution is recommended in their use unless exact data input and retrieval formats are known (and relatively fixed), and there exists sufficient software support. An evolutionary approach to DBMS use is recommended.

Many agencies are now evaluating the use of RDBMS for the joint storage of data and other information. Advances in this field should be closely monitored.

10.2.3.11 Logical file organization

There are two aspects of the logical organization of data – the major groupings, which determine the number of files, and the sets of variable values that are included in the records of each file.

A comprehensive hydrological database will contain the following groups of files:

(a) System reference files that include the code lists (dictionary file) used to check data input, encode data for storage and decode data for output. If some form of spatial data coding is used, then hydrological and/or geographical referencing files will also be needed;
(b) Station description files ranging from simple files relating station numbers to station name, type, location and instrumentation, through to detailed files, such as the complete data for well or borehole logs;
(c) Calibration files containing the detailed background information necessary to compute derived variables, normally on a station-by-station basis. Examples include rating curves for river-flow stations and calibration coefficients for climatological and water quality sensors. Some data are independent of stations, for example, current-meter calibration coefficients and reference tables for theoretical incoming radiation and sunlight hours;
(d) Time-series files containing the series of observations made at hydrological stations. They may be single- or multiple-variable series and may be observed at regular or irregular intervals of time.

The relationship of these various groups of files is shown in Figure I.10.3.

From an organizational point of view, it is possible to combine all information of types (b) and (c) into common files or to split each type into current and historic files. This has the advantage of enabling a standard format and size to be used for the current files. The decision is largely governed by the amount of descriptive data to be held in the computer files compared to that held in manual files.

It is useful to consider the various alternatives available for storing different types of time-series data in the same physical file.

At the simplest level, all stations are allocated their own files with data ordered sequentially in time. This technique is suitable for small data sets or for keeping archived data on tape. However, because hydrological networks may contain several thousand stations of various kinds, this simple system becomes extremely difficult to manage and support with large numbers of files.

At a higher level, that used for most hydrological database systems is the use of files containing many stations, where each file contains data of a different type. This may be hydrological, for example, daily discharge values or mixed time-series, for example, several variables at fixed intervals. In the first case, a daily discharge file, for example, would contain

all daily discharge data for the entire hydrological network. The file, if sequentially organized, would be ordered by station and, within each station, by time. In the second case, all daily data would be included, regardless of the hydrological type, and the file would be ordered by both station type and station number. Both these cases are encountered in the Water Data Storage and Retrieval (WATSTORE) system (Kilpatrick, 1981), which comprises five large files. One file contains the station header (description) data. Of the remaining four, three are grouped by hydrological type (water quality, peak flow and groundwater-site inventory) and the fourth, grouped as time series, is the daily values file. This latter file contains data observed on either a daily or continuous basis and is numerically reduced to daily values. Instantaneous measurements at fixed time intervals, daily mean values and statistics, such as daily maximum and minimum values, may also be stored. In 1981, this file contained 190 million daily values, including data for streamflows, stages, reservoir contents, water temperatures, specific conductances, sediment concentrations, sediment discharges and groundwater levels.

At the highest level of integration (other than the utilization of DBMS) are systems that handle all types of time-series data in one common storage format and that store all time-series data in one physical file. Such an approach, used in the New Zealand TIDEDA system, greatly simplifies software development for data management and retrieval tasks because the storage format is standard. Similar data-processing and storage systems, both also HOMS components, are the United Kingdom HYDATA and the Australian HYDSYS systems. Details on how data are manipulated by these data-processing and storage systems can be found in the *Guidelines for Computerized Data Processing in Operational Hydrology and Land and Water Management* (WMO-No. 634).

10.3 DATA RETRIEVAL

10.3.1 Data analysis tools

Data analysis tools can be an integral system working from the same database, or separate manual and computerized tools for performing tasks required to create an archive (see Table I.10.6).

In the development of data extraction tools, it will be necessary to identify the needs/requirements of users and ensure that the tools developed meet these requirements. This will need to take into account data requirements for:

(a) A single series – for example, daily or monthly flow data for a defined period;
(b) A multiple series – for example, flow data from a group of stations or coincident rainfall and streamflow data;
(c) For a single value across a series (such as for modelling or GIS display) – for example, the annual peak discharge for a site or the average annual rainfalls for a number of sites.

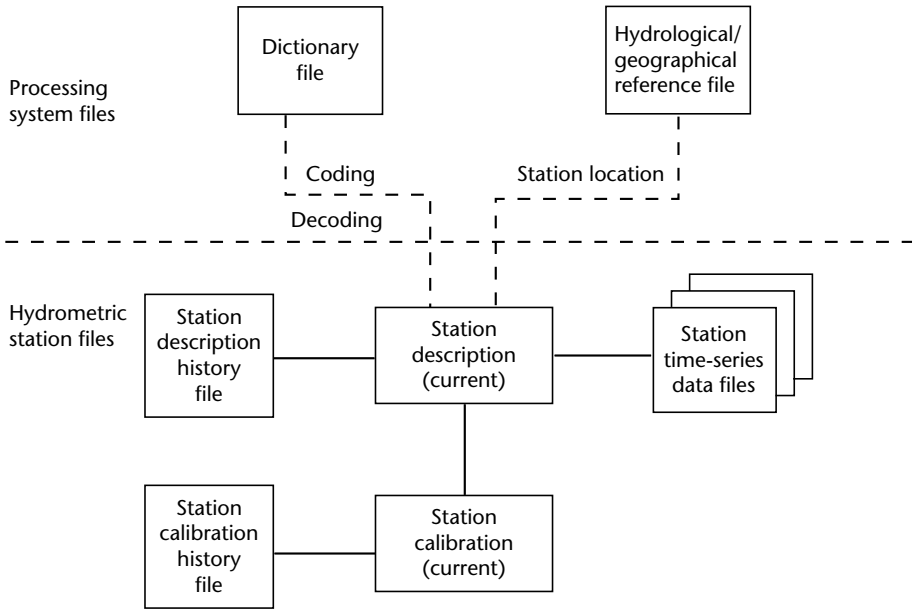


Figure I.10.3. Relationship of hydrometric station data files

Table I.10.6. Data analysis tools

Tool	*Description*
Data input tools	Manual typing, software and hardware for downloading data from loggers, reformatting software, standard spreadsheet tools for formatting and storing data, automated real-time data management tools
Processing tools	Primary processing: hardware and software for digitizing chart records, for example; secondary processing: software tools for conversion of water levels to flows, for example
Validation tools	Software for viewing graphs and editing data, software for producing double-mass curves/maximum-minimum hydrographs, etc.
Analysis tools	Software (including spreadsheets) for producing statistics such as flow duration curves
Query tools	Software tools for retrieving specific data values or statistics from archived data
Reporting tools	Software tools for producing reports/data for dissemination from archived data

Data and information should be able to be retrieved from the database in a range of formats, again targeted to meet the needs of the users, and can includ the following:

(a) Descriptive file – a range of information from different sources describing the data that are available and their characteristics;
(b) ASCII files (10.2.2.2);
(c) Comma-separated value (CSV) files – This is a delimited data format that has fields or columns separated by the comma character and records or rows separated by newline. Fields that contain a special character (comma, newline or double quote) must be enclosed in double quotes. However, if a line contains a single entry which is the empty string, it may be enclosed in double quotes;
(d) Other formats as defined by the user.

10.3.2 Extraction of single variable data

There are occasionally inefficiencies in storing data as multiple time series. These inefficiencies relate to the wide range of variables that may be observed at each site and the way in which data may be retrieved.

Consider climatological data that, after its initial use in the computation of potential evapotranspiration, may be accessed only to retrieve individual variables. Such retrievals are commonly required for the spatial interpolation and/or mapping of data, for example, temperature data for snow melt computations or radiation data for assessing crop-production potential. The retrieval process would be inefficient because all stations must be searched even though the variable was observed at only a limited number of them.

It has been seen (10.2.3.10) that such problems can be overcome by using data pointers stored with each value, which give the location of the record containing the next value of that variable. However, if this technique is used for many variables, the overhead for pointer storage becomes very high. A solution to this problem is to remove important variables – those frequently accessed individually – and to store them as single variable time series. This practice is standard with rainfall data observed at climatological stations. This extraction of important variables is best performed during the annual update when validated data are transferred to the historic archive.

It should be stressed that the decision to perform single variable extraction is dependent upon the anticipated way that data will be retrieved. The frequent retrieval of values for a specific variable suggests the extraction of that variable from the multiple variable set. The fewer the stations at which such a variable is observed, the more inefficient is the multiple variable search, and the stronger the case for a single variable format.

If, as is usually the case with water-quality data, retrievals are made for several variables relating to the same observation time, then the original multiple variable format probably remains the most convenient.

10.3.3 Data retrieval system

Data retrieval is discussed in detail in the *Guidelines to Computerized Data Processing in Operational Hydrology and Land and Water Management* WMO-No. 634). The ability to rapidly retrieve selected data sets is one of the fundamental advantages of electronic hydrological data processing. Efficient retrieval systems allow the hydrologist or water resources planner to concentrate on data analysis by minimizing the previously time-consuming tasks of locating, collating and

manually processing data. A comprehensive retrieval system should contain the following features:

(a) A wide range of data-selection criteria – Typically these should be by variable, basin, station, time period and variable value (or range). In particular, it should be possible to select data on the basis of any combination of these criteria;
(b) Data interpolation/aggregation in time and space – Perhaps the most important of these options are the interpolation of irregular into regular time series and the aggregation of short time-interval series into totals or averages of a longer time base, that is, conversion of hourly into daily values or daily into 10-day values.
If some form of geographical/hydrological referencing system is used, spatial data adjustments may also be made;
(c) Computation of simple statistics – Some facility should exist to enable the computation of simple statistics for the period(s) of record selected. This would include totals (if relevant), means, standard deviations and ranges. More comprehensive statistics – cross-correlations, multiple regressions, probability analyses, etc. – may be offered as part of the standard retrieval system, or the selected data may be passed to a statistical package (or user program) as described below;
(d) Selection of output format – This feature should allow for the direct output of data in (specified) tabular or plotted format and for the creation of data files in formats suitable for further processing. In this latter case, the retrieved data set may be stored for input to statistical packages or user-specific application programs.
A particular output format may be suitable for the interchange of hydrological data on a national or international basis;
(e) Selection of output device – There should be broad flexibility in the choice of output device. At a minimum, this should include a line printer, VDU and disk file. If available, a plotter should be selectable. Data to be transferred to tape or floppy disk is normally first stored on hard disk and transferred with a separate utility requiring several user-specified variables.

It is important that retrieved data, particularly that intended for printed tabular output, retain their codes and flags relating to status and reliability (9.3).

Background information relating to the general reliability of data and/or unreliability during specific periods should be available to the user through the station description file (2.5.2) or the data catalogues.

Data retrievals may be generated in three ways:

(a) Routine data retrievals – These are station data summaries and statistics produced on a monthly and annual basis;
(b) User-specified retrievals – After consulting hydrological yearbooks or data catalogues, users may request data retrieval by using an appropriate form, and the retrieval is submitted as a normal batch job. This relies on computer operators or other technicians to input the retrieval request using the data retrieval software.
The retrieval request form should allow for a wide selection of output media;
(c) Online (interactive) retrieval of data – There are several modes of online specification of data retrievals which, because of their potentially wide use, are discussed below.

As discussed earlier in this chapter and as shown in Figure I.10.4, the existence of an online master database allows the interactive retrieval of data. However, except for systems with small amounts of data or very large disk storage capacities, the major part of the database must be stored offline. Thus, the direct interactive mode is usually suitable only for retrieving limited quantities of most recent data. In some systems, remote users can send messages to the computer operators to request the mounting of a particular offline database volume. However, such requests are rarely satisfied immediately, and this can become very inefficient in terms of terminal usage and communication costs.

Probably the most efficient means of online specification of retrievals is the two-stage process. In the first stage, an interactive program allows the user to specify retrieval requirements, and in the second stage this request is automatically submitted as a batch job and the output is obtained later. The format of an interactive machine/user interface is called a menu system. Executing large data retrievals in batch mode is much more efficient in terms of the computer's ability to allocate its resources, particularly for the extraction of data from offline volumes.

The above discussion relates primarily to online retrievals of data from hydrological-inventory systems. However, the ability to review data being collected and stored for real-time systems is perhaps a more fundamental requirement. Retrieval options range from telemetry interrogation of individual or groups of field stations to the plotting and display

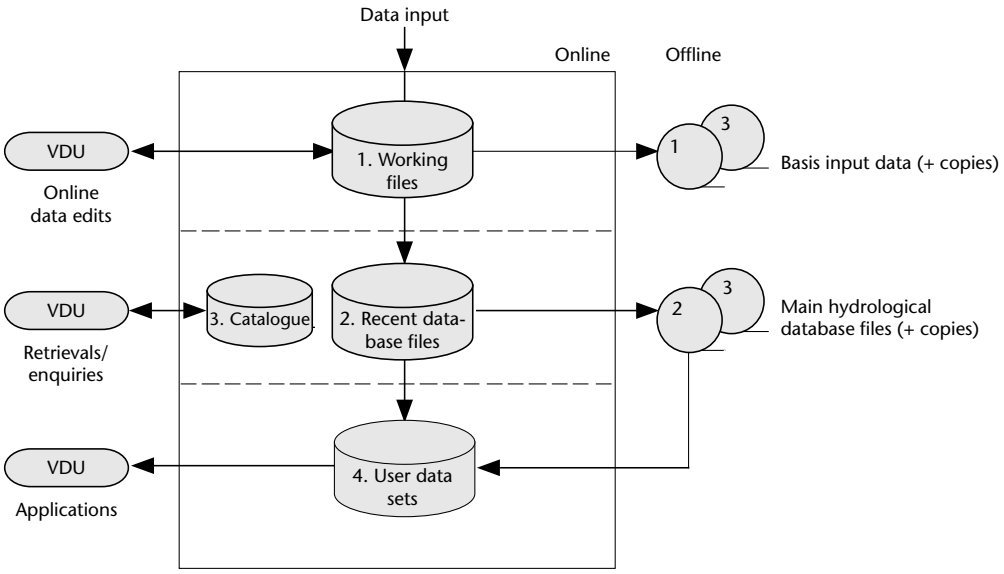


Figure I.10.4. Disposition of online and offline data sets

of recently collected data, and recently made forecasts, at the processing centre.

10.4 DATA DISSEMINATION

10.4.1 General

Data are of no value until they are used; only when hydrological data are analysed and used as part of the water management planning and decision-making process do they become really valuable. Good quality long-term records are needed in order to quantify the mean and variability (both seasonal and inter-annual) of any hydrological variable. Thus, the standard period of data required for 'reliable' estimation of mean annual rainfall is 30 years, and yet given very long observed periodicities in rainfall in some parts of the world, perhaps even this is not really sufficient. In addition, given clear evidence of global warming and associated climate change, scientists and engineers require long records to be able to detect and monitor trends in rainfall, river flows and groundwater recharge to enable the preparation of contingency plans for coping with changing water resources.

To be useful, good quality data must also be readily available to a range of users. Data are often collected by agencies that are themselves the primary users of the data, such that the data collection underpins the work of the agency, for example, public water supply, operation of irrigation schemes or operation of hydropower facilities. Such agencies are often, but not always, governmental bodies. In such cases, internal dissemination of data in the agency is a matter that will not be dealt with here. This Guide considers how potential users of the data who are not part of the collecting agency can access hydrometeorological data, assuming that the data have been entered into an appropriate database system as described earlier.

Potential external users of hydrological data might include staff of other government departments, private sector water supply or hydropower companies, engineering or environmental consultants, academics and researchers. A wide range of potential users exist, whose requirements for data are quite variable, with some needing solely data from a single point on a single river, and others requiring data over a region, whole country or even groups of countries in the case of transboundary rivers.

Access to data

International access to data, both meteorological and hydrological, is a subject that engaged the attention of the World Meteorological Organization and its Member countries over many years. This resulted in the adoption by the Twelfth WMO Congress in 1995 of Resolution 40 (Cg-XII) that sets out WMO policy and practice for the exchange of meteorological and related data and products, including guidelines on relationships in commercial meteorological activities. At its subsequent session in 1999, the Thirteenth WMO Congress adopted Resolution 25 (Cg-XIII) – Exchange of

hydrological data and products – thereby establishing the policy and practice for the international exchange of hydrological data and products. This Resolution provides the framework to facilitate international access to hydrological data and products (see WMO-No. 925). The full text of Resolution 25 (Cg–XIII) is also available at: http://www.wmo.int/pages/prog/hwrp/documents/Resolution_25.pdf.

10.4.2 Catalogues of data availability

The first requirement of any user of the data will generally be maps showing the location of all stations of various types, combined with tables showing the data held at each site, and the period of record covered by each. This type of information forms the metadatabase of the data set, and may be a separate set of data tables within a computerized database (as described earlier), or may take the form of hard-copy printed material. Thus users may either access this sort of information electronically via an Internet portal, or may have to search for the material using printed yearbooks.

The traditional means of providing such information to users was through yearbooks, often printed annually, although sometimes summary catalogues of data holdings might only be produced every three to five years, as data networks are generally fairly stable over time. This approach is simple, but can be expensive in terms of printing costs, and the printed catalogues may not be easily accessible to all users. However, for many years, printed catalogues have often been the most effective means of disseminating information on data availability and will continue to be effective in countries where Internet access is not commonly, or reliably, available.

However, in the future the most common means of making such catalogue data available to users is increasingly likely to be through a web-based browser linked directly to the metadata. This has the advantage of being available to all users with Internet access, and there is no requirement to have yearbooks. It is also potentially easier to maintain the system and to keep it up to date.

For each gauged catchment, the information provided should include:

(a) Details of the catchment, for example, its size, geomorphology, landforms, vegetation and land use;
(b) Climate zone and average annual rainfall and evaporation for the catchment;
(c) Location, type and quality of the gauging station;
(d) Details of any upstream regulation or factors that may complicate the use of the records;
(e) Period, completeness and quality of the streamflow and water quality (including sediment transport) records;
(f) Locations of meteorological stations in or near the catchment and their periods of record.

This information is grouped and discussed under three headings, namely descriptive information, catchment map and data availability.

In order to assist the users in identifying the gauged catchments that are appropriate to their purposes, a description of the characteristics of each gauged catchment and the principal features of the gauging facilities, and an indication of the quality and reliability of the flow record, should be provided.

Suggested headings and pertinent information are illustrated in Table I.10.7. In practice, all details may not be available or appropriate under each heading for each gauged catchment, but it is suggested that the same format be retained throughout.

An example which complements Table I.10.7 is provided in Figure I.10.5. A map for each catchment or group of catchments has proven to be valuable. The map should be produced at a scale that is convenient for displaying the information. Catchments of different scales may warrant maps of different scales. In the near future any information for the production of catchment maps will be retained within computer-based GIS for ease of presentation at a variety of scales. The information to be included on the map is described in Table I.10.8 and a basic example is provided in Figure I.10.6.

The data-availability page should present a relatively concise and easily updated summary of streamflow, precipitation and water quality data. It should be based on monthly data for flow and precipitation and on annual water quality data. For catchments with many precipitation stations, it is impractical to include a summary for each station. All stations and their period of record are shown on the map described in the previous section, so it would be sufficient to restrict the data availability to pluviographs and a selected set of key daily precipitation stations. Stations with long periods could require several pages to ensure adequate scales for legibility.

Table I.10.7. Outline of data–catalogue format

Identification	*Description*
Name	River name, station name and station number
River basin	Basin name and number
Location	Gauging station location in latitude and longitude and local grid coordinates
Catchment details	
Catchment area	The catchment area expressed in square kilometres
Climate zones	The climate over the catchment expressed in bioclimatic zones that reflect the amount and occurrence of precipitation
Average rainfall	An assessment of the mean annual rainfall at the centroid of catchment and, for large catchments, the range of mean annual rainfall across the catchment. The sources of the figures should be quoted
Pan evaporation	An assessment of the mean annual pan evaporation at the centroid of catchment. The source of the figures should be quoted
Geomorphology	Descriptive comments on the relief, landscape and underlying geology of the gauged catchment
Landforms	Quantitative estimate of proportions of major landforms within the catchment
Natural vegetation	Descriptions of the natural vegetation derived from vegetation surveys
Clearing	Proportion of natural vegetation cleared, or substantially altered by intrusive human activity. Source and date of clearing estimates should be included.
Present vegetation	Descriptions of the present vegetation cover across the catchment with a reference to the source
Land use	Comments on land use. Source of information should be quoted, be it field observation, map of rural land use, or more detailed evaluation
Regulation	Comments on upstream developments that could modify the runoff regime. Possible sources of detailed information should be listed.
General comment	Where the station does not measure total catchment runoff or the record cannot be corrected for upstream regulation, the catchment characteristics are omitted in favour of comment on the station's particular special purposes or functions.
Gauging station details	
Period of record	Month and year of opening and closing of the gauging station. When more than one station has operated near the same river a suitable reference is included.
Classification	The gauging station's current classification within the hydrological network (for example, project station or basic-network station)
Gauging installation	Description of stage-recording instruments and the features controlling the river stage at the gauging station. Changes in either of these facilities during the period of operation should be noted.
Stage record	Annual average percentage of data recorded and percentage of these data that require interpretation in processing (faulty record)
Rating curve	Brief comments on the method and quality of the stage-discharge relationship, together with maximum measured discharge. Where possible, the proportion of measured flow that the maximum measured discharge represents should be known.
Sensitivity measure	Some measure of the rating-curve sensitivity should be provided. The preferred method to indicate sensitivity is the percentage of flow volume that could be measured to within 1, 2 or 5 per cent with a 1-mm error in the stage record. Note that this measure is based on the slope of the rating curve and the cumulative flow duration curve. Alternatively, it may be defined for a 10-mm or 100-mm error in stage.

607003 Warren River	Wheatley farm		
River basin	Warren river		
Location	Latitude S 34˚22′ 14″	AMG. Grid	N 6196500
	Longitude E 116˚16′ 34″		E50 433450
Catchment characteristics			
Catchment area	2 910 km^2		
Climate zone	Mediterranean climate; intermediate to low rainfall.		
Average rainfall	735mm/annum (Range 950–550).		
Pan evaporation	1 275 mm/annum (Range 1 250–1 400		
Geomorphology	Low to moderate relief; undulating plateau with incised mainstream valley, bauxitic laterite soils over Archean grantic and metamorphic rocks.		
Landforms	Map units; Atlas of Australian Soils (Ref. 8) 16% – Ub90 dissected laterites; rolling country with yellow mottled soils and gravelly ridges 14% – Cb43, Tf6 swampy flats; shallow drainage lines with leached sands and podzolic soils 57% – Cd22, Tc6 laterite plateau; uplands with sands and ironstone gravels over mottled clays 13% – Tf6, Ta9 incised valleys; moderate slopes. mainly yellow podzolic soils		
Natural vegetation	Map units; vegetation survey of WA (Ref. 1) 20% – eMi woodland; marri-wandoo woodlands on dissected laterites 70% – eMc forest; jarrah-marri forest on laterite plateau 10% – mLi low woodland; paperbark woods on swampy flats		
Clearing	About 40% area cleared (only 27% cleared in 1965)		
Land use	About half catchment in state forest, cleared areas used for sheep and cereal production in upper catchment and beef production in lower reaches		
Regulation	Small farm dams on minor water courses		
Gauging station details			
Period of record	May 1970 to date		
Classification	Hydrological network – primary mainstream catchment		
Gauging installation	L&S servo manometer and continuous graphical recorder to date. Rock bar control for low and medium flows; channel control for high flows		
Stage record	96.5% recorded, 7.6% faulty		
Rating curve	Low to medium rating fair due to nature of control, medium to high flow rating good, but theoretical beyond measured range. Numerous discharge measurements to 97.04m^3 s^{-1}, which represents 99% of total recorder flow yield		
Sensitivity measure	99% of flow < 1;100% of flow < 2		

Figure I.10.5. Sample data-catalogue page

Table I.10.8. Outline of map details

Identification	*Description*
Catchment boundary	Scale and source of map from which the catchment boundary was defined
Streamlines	The number of streamlines to be included should be a function of catchment area. Source of streamline data
Catchment scale	Variable – a function of catchment size
Rainfall stations	Location and station number, period of operation and type of raingauge, for instance, pluvio, daily read or storage
Rainfall isohyets (optional)	Average annual rainfall isohyets for the catchment with the reference
Land use (optional)	Where applicable the boundaries of the main land uses should be known. Forest, agricultural and urban boundaries would be one example.

It is suggested that the information in Table I.10.9 be included on the data-availability page.

10.4.3 Summary reports

Many organizations publish summaries of data. Some examples include climate averages, rainfall statistics, streamflow statistics/records and water quality records or surveys.

Typically, such publications consist of station information, including station number, latitude and longitude, type of data collected, other site specifications (name, river name, grid reference, catchment area, etc.), period of operation, period of data processed, and instantaneous, daily, monthly and annual data summaries (including minimum, maximum and mean values). Data may be presented as part of the text or attached as microfiche or provided on a computer-compatible form, such as a disk or CD-ROM.

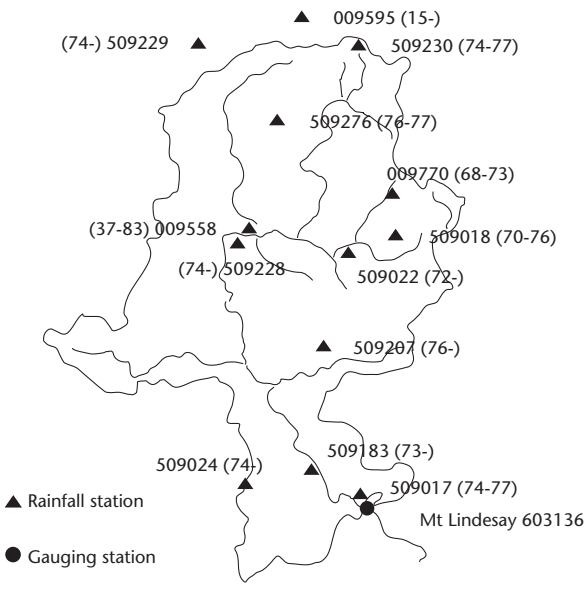


Figure I.10.6. Denmark river catchment

10.4.4 Yearbooks

A yearbook is a very effective means of disseminating hydrological data, although only certain types of data can be published. Thus, using modern data loggers or telemetry, rainfall and river levels (and hence flows) are increasingly observed every 15, 30 or 60 minutes, leading to between 8 760 and 35 040 values per year. It is neither practicable, nor generally necessary, to publish this fine time resolution data, and yearbooks generally only publish daily or even monthly rainfall totals and mean daily flows.

Groundwater data vary only slowly with time and are also sometimes monitored only intermittently, perhaps weekly or monthly. Such data sets may therefore be published in their entirety. Other climate variables, such as temperature, wind speed and radiation, are often published only as monthly means.

Examples of typical yearbook output taken from the United Kingdom National River Flow Archive (NRFA) are shown in Figures I.10.7 to I.10.13.

10.4.5 Data export on request

National hydrological data sets are increasingly being disseminated to users via the Internet, where

Table I.10.9. Outline of data-availability page(s)

Identification	*Description*
Flow data	Available record and record quality clearly presented in a month-by-month form
Rainfall data	Available record and record quality clearly presented on a month-by-month basis for the key pluviograph and manually read rainfall gauges. The period of record covered may be restricted to the period covered by the stream-gauging station for practical reasons.
Water quality	Number of samples analysed each year within a meaningful set of analyses groupings. The groupings suggested are as follows: (a) Samples with basic analysis only (any or all of conductivity, pH, river temperature, colour, or turbidity parameters); (b) Samples analysed for major icons; (c) Samples analysed for nutrients; (d) Samples analysed for heavy metals or other trace constituents.

users can use a map and tabular dialogue box interface to select stations of interest and types of data they wish to download. Internet access enables users to browse the data set and determine what types of data they require from a selected station or set of stations.

Some systems may then allow users to download the selected data directly to their own PCs, or alternatively, users may only be permitted to place an electronic request for the data onto the website. One good reason for not allowing users to download whatever data they request is that data volumes can be large, and data transfer may be unacceptably slow through some Internet service providers, particularly where slow modem links are used. For the same reason, providing the data to users in the form of attached e-mail files can be problematic given file size limitations on some e-mail portals.

Figure I.10.7. Map showing United Kingdom gauging stations held in NRFA

In many cases, the preferred option is for users requiring data to post a request on the website, with data being provided by CD, or possibly by being placed onto a File Transfer Protocol (FTP) site. The user would then be able to download the data from this site.

Data may be freely available from the website, particularly where the user is able to download data directly. However, in some cases, users may have to pay a handling charge for the data to cover the staff costs associated with preparing the CD. Although some users may object to paying for data, charging is often justified as the data-providing agency often has to justify its continued existence to its funders and managers. The fact that users are paying for data can provide at least part of an agency's funding needs, but perhaps more importantly, it demonstrates that its work is valued by external users or customers.

A good example of a web-based data retrieval system is the United Kingdom National Water Archive: http://www.nwl.ac.uk/ih/nwa/index.htm, or the website for hydrological data of USGS: http://water-data.usgs.gov/nwis/, or as an example of data from WHYCOS projects: http://medhycos.mpl.ird.fr/ and http://aochycos.ird.ne/HTMLF/ETUDES/HYDRO/INDEX.HTM.

039002 1996 Thames at Days Weir

Measuring authority: EA | Grid reference: 41 (SU) 568 935 | Catchment area (sq km): 3 444.7
First year: 1938 | Level stn. (m OD): 45.80 | Max. alt (m OD): 330

Daily mean gauged discharges (cubic metres per second)

Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1	56.600	25.800	42.600	24.800	20.400	11.000	5.380	3.970	3.730	4.120	3.520	10.000
2	57.500	24.200	40.400	21.800	23.200	8.850	5.490	3.710	3.050	3.960	4.020	9.970
3	60.600	24.200	39.100	21.100	19.300	9.250	5.320	3.640	2.770	3.550	4.060	11.100
4	56.300	22.600	38.600	20.800	13.100	9.160	5.140	3.630	2.700	2.860	7.280	12.600
5	54.700	22.400	37.700	21.100	13.900	9.130	5.170	3.430	3.120	3.090	6.880	15.400
6	54.900	24.000	33.800	20.800	15.500	8.860	6.050	3.200	3.100	3.060	8.280	13.500
7	63.400	24.000	33.700	20.500	15.500	10.100	5.640	3.060	3.060	3.040	5.620	11.200
8	77.900	23.600	33.500	20.100	16.400	14.600	5.040	2.750	2.980	3.300	5.350	9.310
9	101.000	30.700	35.000	20.100	15.500	11.800	5.270	2.760	2.970	4.110	6.450	9.320
10	109.000	53.600	35.900	21.500	14.200	11.200	5.410	3.720	2.880	3.250	5.810	8.870
11	97.600	57.200	33.300	22.500	14.100	9.190	3.870	4.850	2.990	2.980	3.210	8.810
12	84.100	81.600	34.600	22.500	14.200	8.880	5.500	4.320	3.220	3.520	4.470	8.840
13	77.800	99.800	36.000	48.400	14.100	8.410	4.730	4.220	3.250	3.430	5.480	8.340
14	69.400	90.200	35.200	40.600	13.800	7.440	3.870	3.790	3.020	3.130	4.590	8.140
15	59.300	64.100	32.000	31.500	13.900	5.880	3.940	3.680	2.940	3.170	4.340	8.500
16	54.800	53.900	31.100	26.400	13.300	6.240	3.810	3.380	2.840	3.220	5.240	8.020
17	50.400	48.400	30.600	24.700	13.200	6.020	3.820	3.340	3.180	3.420	6.390	8.430
18	46.500	48.400	25.500	21.100	13.200	5.980	3.700	2.850	2.500	4.2900	5.900	8.020
19	45.600	47.700	27.300	23.900	14.100	5.990	2.650	2.840	2.620	4.030	10.700	10.500
20	44.300	40.700	26.700	24.400	13.900	5.970	3.210	2.940	2.910	3.210	11.600	12.900
21	41.100	38.500	26.800	23.500	13.900	5.930	3.720	3.170	2.900	3.680	11.400	15.000
22	37.400	37.100	27.800	25.200	13.400	6.130	3.350	3.340	2.850	3.500	10.700	14.300
23	37.900	37.200	33.900	35.100	14.000	5.990	3.260	5.340	2.850	3.380	9.080	12.100
24	38.400	59.000	36.400	42.800	19.500	5.660	3.210	6.820	3.150	3.410	8.330	10.100
25	37.600	95.400	32.500	26.800	15.300	5.510	3.200	7.110	3.660	3.010	11.900	10.700
26	33.700	92.500	34.200	25.000	15.700	5.480	3.270	4.790	3.750	3.530	11.900	10.100
27	32.900	73.700	45.900	22.000	13.100	4.660	3.430	3.800	3.300	4.120	12.600	9.530
28	26.800	59.100	33.300	21.700	12.800	4.970	4.000	3.960	3.340	3.870	10.300	8.340
29	26.500	43.600	32.300	20.600	12.800	5.570	5.480	3.210	3.370	4.670	10.100	8.720
30	26.300		26.300	20.500	11.100	5.520	4.620	3.590	3.350	4.560	9.390	9.030
31	26.100		25.500		11.000		3.950	4.180		3.240		8.830
Average	54.400	49.750	33.470	25.390	14.750	7.646	4.339	3.851	3.078	3.539	7.496	10.270
Lowest	26.100	22.400	25.500	20.100	11.000	4.660	2.650	2.750	2.500	2.860	3.210	8.020
Highest	109.000	99.800	45.900	48.400	23.200	14.600	6.050	7.110	3.750	4.670	12.600	15.400
Monthly total (million cu m)	145.70	124.70	89.64	65.82	39.52	19.82	11.62	10.32	7.98	9.48	19.43	27.52
Runoff (mm)	42	36	26	19	11	6	3	3	2	3	6	8
Rainfall (mm)	42	63	36	49	36	21	34	53	22	50	79	29

Statistics of monthly data for previous record (October 1938 to December 1995)

Mean flows	Avg.	56.450	56.680	44.600	30.650	20.140	14.260	8.397	7.073	8.666	14.960	30.750	45.570
	Low	6.252	5.548	5.619	4.255	2.854	1.504	0.401	0.290	1.740	2.782	3.751	5.308
	(year)	1976	1976	1976	1976	1976	1976	1976	1976	1959	1959	1990	1975
	High	133.600	120.800	163.200	85.060	61.140	41.560	48.810	18.690	38.640	74.570	128.100	128.700
	(year)	1939	1977	1947	1951	1983	1955	1968	1977	1946	1960	1960	1960
Runoff	Avg.	44	40	35	23	16	11	7	5	7	12	23	35
	Low	5	4	4	3	2	1	0	0	1	2	3	4
	High	104	85	127	64	48	31	38	15	29	58	96	100
Rainfall	Avg.	68	47	53	47	58	54	53	64	62	64	70	73
	Low	13	3	5	4	7	5	5	3	5	6	8	16
	High	132	135	152	99	131	124	117	149	129	163	178	316

Summary statistics *Factors affecting runoff*

	For 1996		For record preceding 1996		1996 As % of pre-1996
Mean flow	18.070		28.050		64
Lowest yearly mean			10.100	1973	
Highest yearly mean			51.290	1960	
Lowest monthly mean	3.078	Sep	0.290	Aug 1976	
Highest monthly mean	54.400	Jan	163.200	Mar 1947	
Lowest daily mean	2.500	18 Sep	0.050	7 Jul 1976	
Highest daily mean	109.000	10 Jan	349.000	17 Mar 1947	
10% exceedance	44.770		67.810		66
50% exceedance	9.412		15.940		59
95% exceedance	2.959		3.181		93
Annual total (million cu m)	571.40		885.20		65
Annual runoff (mm)	166		257		65
Annual rainfall (mm)	514		713		72
{1961–1990 rainfall average (mm)			690}		

$ Abstraction for public water supplies
$ Flow reduced by industrial and/or agricultural abstractions
$ Augmentation from effluent returns

Figure I.10.8. Example of NRFA yearbook tabulations

39002 – Thames at Days Weir

Grid reference	41 (SU) 568 935
Operator	EA
Local number	1900
Catchment area	3 444.7 km^2
Level of station	45.8 m OD
Maximum altitude	330.0 m OD
Mean flow	28.38 m^3 s^{-1}
95% exceedance (Q95)	3.22 m^3 s^{-1}
10% exceedance (Q10)	69.2 m^3 s^{-1}
1961–1990 average annual rainfall	690 mm

Figure I.10.9. Station characteristics

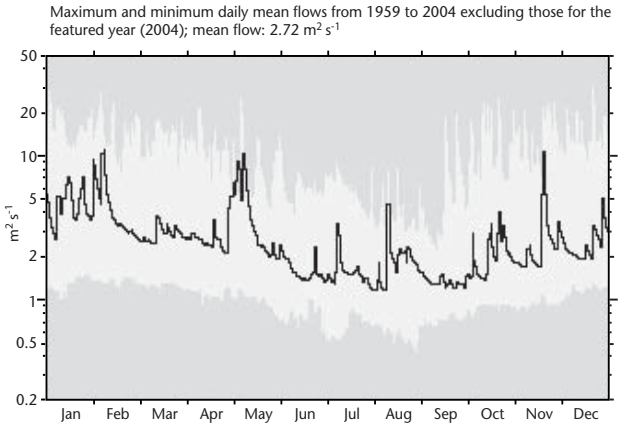


Figure I.10.10. Sample hydrograph of gauged daily flows

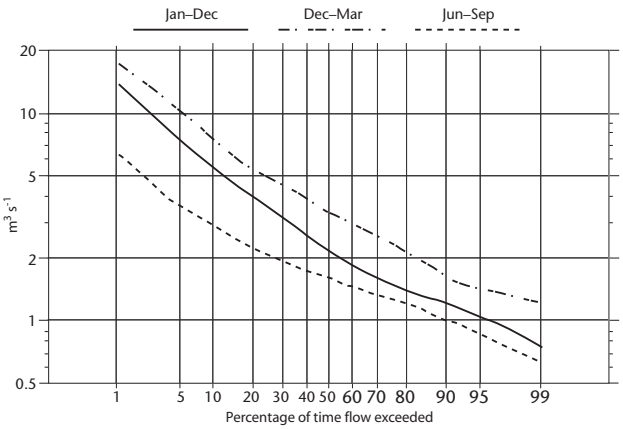


Figure I.10.11. Flow duration curve for gauged daily flows

Station description

Adjustable thin-plate weir (5.48 m wide) plus 15 radial gates, replaced a barrage of radial and buck gates in 1969. Rating formulae based on gaugings – tailwater calibration applies for flows > 70 cumecs; above 100 cumecs overspill occurs. Daily naturalized flows available for POR (equal to gauged flows up to 1973) – allow for Didcot power station losses only. Peak flows under review.

Catchment description

Mixed geology (oolitic limestone headwaters, oxford clay below). Predominately rural with development concentrated along the valley.

Factors affecting runoff

- Runoff reduced by public water supply abstraction
- Runoff increased by effluent returns
- Runoff reduced by industrial/agricultural abstraction

Figure I.10.12. Example of metadata

10.4.6 Data-exchange formats

There are currently no standards for data exchange formats for hydrological data. The only standards that exist are the de facto standard formats produced by the most common data loggers and database software systems. Current data exchange formats generally fall into two categories, as follows.

Text-based files

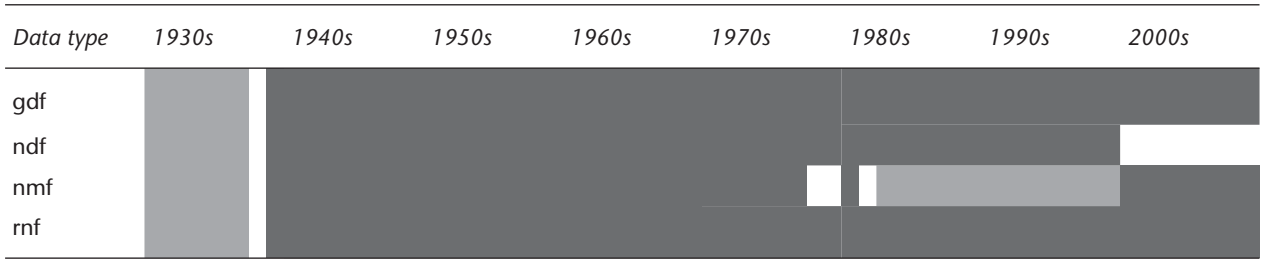
Text-based data files have the benefit of being easily readable by a user with the simplest of computer software. Data time series are often stored as columns of dates, times and values, with each value separated by a delimiter that could be a comma (resulting in a comma-separated value, or CSV, file), or other character or fixed number of spaces.

Proprietary format

The disadvantage of text-based formats is the size of the resulting file. Many software systems use proprietary formats that make far more efficient use of memory. This results in smaller files that take up less space on computer disks or transfer media and are faster to move around the Internet. The disadvantage is, of course, that specialist software is usually required in order to read the files.

XML

There is a movement towards data exchange format identification with the widespread uptake of XML, which stands for Extensible Markup Language (Word Wide Web Consortium, 2004). XML is itself a standardized format for data files, albeit at the most general level. With the increasing use of the Internet during the 1990s, particularly in the areas of publishing, manufacturing and retail, there was a need for a standard way for computer programs to be able to read data files and make sense of the content, without prior knowledge of the format. In 1998, XML was defined by the World Wide Web Consortium (W3C) as a platform and method for putting structured data into a text file. XML is similar to HyperText Markup Language (HTML), the most common format for files on the Internet, which allows many different software packages to display the content of files of this format in the same way. But whereas HTML describes how the content should look, XML says nothing about presentation and simply describes what the content is.



Gauged daily flows (gdf): 1938 to 2003
Naturalized daily flows (ndf): 1938 to 2002
Naturalized monthly flows (nmf): 1938 to 2002
Monthly catchment rainfall (rnf): 1938 to 2001

Figure I.10.13. River flow and catchment rainfall in NRFA

Table I.10.10. Gauging station's name and coordinates in XML

```
<gaugingstation>
    <name>River Thames at Wallingford
    </name>
    <coordinates>
        <easting unit=metres>461300
        </easting>
        <northing unit=metres>189900
        </northing>
    </coordinates>
</gaugingstation>
```

The content of the data file is written within tags, which describe what the data are about. For example, Table I.10.10 contains a gauging station's name and coordinates in XML.

A computer program reading this XML would know, without understanding anything about hydrology, that the file contains information about a 'gauging-station', that it has an attribute called 'name' that this attribute has a value 'River Thames at Wallingford', and that it has 'coordinates' with further attributes 'easting' of value 461300 and 'northing' with value 189900, both with units of 'metres'. The '<>' symbols are called tags, and pairs of tags enclose data values, while the text within the tags describes the data enclosed.

The advantages and disadvantages of XML are widely discussed, but can be summarized simply.

Advantages: Ability to separate form from content, and thus quickly apply different rules of display to a range of files of the same format. The data that can be stored in a file, as well as rules for these data, can be explicitly stated and software can use this to validate data files whilst reading. Files can be also searched efficiently.

Disadvantages: The uncompressed text file means that file sizes are large. XML was not invented for the purpose of describing time-series data that can increase file sizes by a factor of 10 over even uncompressed text-based formats.

One major advantage of XML is that it can be specialized in particular subjects. For example, libraries have defined an international format for describing the tags and rules for storing information about books in XML. These standards indicate that all libraries can provide data that can be read and understood by all other libraries. The same is gradually occurring in the more complex area of environmental science. Already there are emerging XML formats for a wide range of applications, including the description of molecules and the Climate Science Modelling Language. GIS data now have a comprehensive XML-based standard called the Geography Markup Language (GML) that will allow the interaction of digital maps from all sources, and could be used for the dissemination of spatial data. GML is the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features (Cox and others, 2004). GML serves as a modelling language for geographic systems as well as an open interchange format for geographic transactions on the Internet.

Many of the definitions of these XML specialisms (areas in which XML specializes) are still evolving and thus should be used with care. However, some successfully defined languages have achieved ISO standard recognition. An XML specialization in the field of hydrology has not yet been developed, although the United States National Weather Service has established a Hydrology XML consortium and produced a draft hydrology XML schema.

References and further reading

Cox, Simon, Paul Daisey, Ron Lake, Clemens Portele and Arliss Whiteside (eds.), 2004: *OpenGIS Geography Markup Language (GML) Implementation Specification Version 3.1.0*. Recommendation Paper, February 2004, Open GIS Consortium, Inc. and ISO Reference No. OGC 03-105r1.

Kilpatrick, Mary C., 1981: *WATSTORE: A WATer Data STOrage and REtrieval System*. United States Government Printing Office publication, 52, United States Department of the Interior, United States Geological Survey, Reston, Virginia, pp. 341–618.

Maurer, T., 2004: *Globally Agreed Standards for Metadata and Data on Variables describing Geophysical Processes: A Fundamental Prerequisite for an Integrated Global Data and Information Infrastructure and Thus Improved Management of the Earth System for Our All Future*. Global Runoff Data Centre, Report 31, October 2004.

Woolf A., B. Lawrence, R. Lowry, K. Kleese van Dam, R. Cramer, M. Gutierrez, S. Kondapalli, S. Latham, D. Lowe, K. O'Neill and A. Stephens, 2006: Data integration with the Climate Science Modelling Language. *Advances in Geosciences*, Volume 8, pp. 83–90 (http://www.copernicus.org/EGU/adgeo/8/adgeo-8-83.pdf).

World Meteorological Organization, 1981: *Case Studies of National Hydrological Data Banks: Planning*, Development and Organization. Operational Hydrology Report No. 17, WMO-No. 576, Geneva.

World Meteorological Organization, 1983: *Guide to Climatological Practices*. Second edition, WMO-No. 100, Geneva (http://www.wmo.int/pages/prog/wcp/ccl/guide/guide_climat_practices.html).

World Meteorological Organization, 2001: E*xchanging Hydrological Data and Information: WMO Policy and Practice*. WMO-No. 925, Geneva.

World Meteorological Organization and Food and Agriculture Organization of the United Nations, 1985: *Guidelines for Computerized Data Processing in Operational Hydrology and Land and Water Management*. WMO-No. 634, Geneva.

World Wide Web Consortium, 2004: *Extensible Markup Language (XML) 1.0*. Third edition, W3C Recommendation 4 February 2004 (http://www.w3.org/TR/2004/REC-xml-20040204).