Government of India & Government of The Netherlands

Technical Assistance
HYDROLOGY PROJECT

DHV CONSULTANTS & DELFT HYDRAULICS with HALCROW, TAHAL, CES, ORG & JPS

VOLUME 9
DATA TRANSFER, STORAGE AND DISSEMINATION

DESIGN MANUAL

Table of Contents

Glossary

Authenticated data	fully processed data ready for archiving
BLOB	Binary Large OBject
Catalogue	meta-data with search and selection tool
CD-R	Recordable CD-ROM
CD-R/W	Rewritable CD-ROM
CD-ROM	Compact Disk type Read Only Memory for computer use
CGWB	Central Ground Water Board
CGWBDPC	CGWB data processing centre
CSV	Text based data file format with Comma Separated Values
CWC	Central Water Commission
CWCDPC	CWC data processing centre
DAT	Digital Audio Tape
DLT	Digital Linear Tape
DPC	Data Processing Centre
DRF	Data Request File
DSC	Data Storage Centre
DVD	Digital Versatile Disk
Encryption	Coding data to make it accessible to key owners only
FAQ	Frequently Asked Questions
Field data	Observed data entered in GW/SWDES passing data entry tests
FTP	File Transfer Protocol
GIS	Geographical Information System
GW	Groundwater
H/W	Hardware
HDS	Hydrological Data Supplier
HDU	Hydrological Data User
HIS	Hydrological Information System
HLTG	High Level Technical Group
HP	Hydrology Project
IMD	Indian Meteorological Department
ISDN	Integrated Services Digital Network
LAN	Local Area Network
Observed data	Data collected in the field
PDF	Portable Document Format
RAID	Redundant Array of Independent Disks
S/W	Software
S/W	Software
SCSI	Small Computer System Interface
SGWD	State Groundwater Department
SGWDPC	State Groundwater Data Processing Centre
SI	International System of Units
SQL	Structured Query Language
SW	Surface Water
VPN	Virtual Private Network
WAN	Wide Area Network

1 INTRODUCTION

1.1 GENERAL

The prime objective of the Hydrology Project is to develop a sustainable Hydrological Information System for 9 states in Peninsular India, set up by the state Surface Water and Groundwater Departments and by the central agencies (CWC and CGWB) with the following characteristics:

- Demand driven, i.e. output is tuned to the user needs
- Use of standardised equipment and adequate procedures for data collection and processing
- Computerised, comprehensive and easily accessible database
- Proper infrastructure to ensure sustainability.

This Hydrological Information System provides information on the spatial and temporal characteristics of water quantity and quality variables/parameters describing the water resources/water use system in Peninsular India. The information needs to be tuned and regularly be re-tuned to the requirements of the decision/policy makers, designers and researchers to be able to take decisions for long term planning, to design or to study the water resources system at large or its components.

This manual describes the procedures to be used to arrive at a sound operation of the Hydrological Information System as far as hydro-meteorological and surface water quantity and quality data are concerned. A similar manual is available for geo-hydrological data. This manual is divided into three parts:

a) Design Manual, which provides information for the design activities to be carried out for the further development of the HIS
b) Reference Manual, including references and additional information on certain topics dealt with in the Design Manual
c) Field/Operation Manual, which is an instruction book describing in detail the activities to be carried out at various levels in the HIS, in the field and at the data processing and data storage centres.

The manual consists of ten volumes, covering:

1. Hydrological Information System, its structure and data user needs assessment
2. Sampling Principles
3. Hydro-meteorology
4. Hydrometry
5. Sediment transport measurements
6. Water Quality sampling
7. Water Quality analysis
8. Data processing
9. Data transfer, storage and dissemination, and
10. Surface water and groundwater protocols.

This Volume 9 deals with data transfer, storage and dissemination procedures and consists of a Manual Design, an Operational Manual and a Reference Manual.

This Design Manual deals with the design aspects of the data storage system.

1.2 STRUCTURE OF THE HYDROLOGICAL INFORMATION SYSTEM

Hydrological, geo-hydrological and hydro-meteorological data are being collected by the State Surface Water and State Ground Water Departments, the Central Water Commission, the Central Ground Water Board, and India Meteorological Department. The data collected by the State Agencies are stored in a database at the State Data Storage Centre available in every State, whereas the data collected by the Central Agencies are stored at their Regional and National Data Storage Centres. Each is periodically updated to reflect the new data collected from the field stations and wells.

At each Centre, a catalogue is available about the information stored in that Centre itself and in all other Data Storage Centres. Data transport links have been established between the various Centres for exchange of information.

The databases enable integration and consolidation of all types of hydrological data as text data, tables, time-series, spatial data, maps and related data for **efficient, reliable** and **secured** use.

The databases in the Data Storage Centres make use of a uniform data structure in SQL Server database environment. Application software has been developed to carry out all required database functions in all State, Regional and national Data Storage Centres. The software is **standard, uniform** and **generic** in nature.

The structure of the HIS in which the databases have to function has been described in Volume 1 of the manual.

2 HYDROLOGICAL DATA USERS GROUP

2.1 GENERAL

The Data Storage Centre (DSC) maintains a database system for permanent storage of hydrological data pertaining to its service area. It receives data from the data producing offices, i.e. the owner Data Processing Centres (DPCs), and makes the data available to authorised hydrological data users.

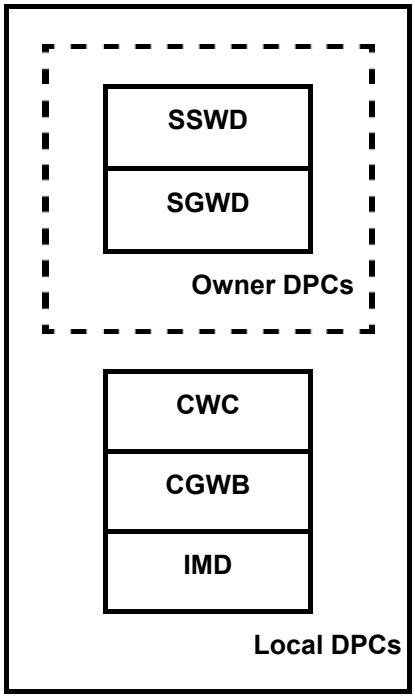
The following categories of data are discerned:

- field data
- authenticated data
- objects, and
- temporary data.

The **authenticated data** constitutes the deliverable product of validated data, the **field data** is the raw data. The **objects** can be virtually any data item, in this category fall useful products that do not belong to the field/authenticated data. Examples are scanned maps / toposheets, yearbooks, well logs, reports, special time-series, etc. Many of the objects consist of files. The **temporary data** is a category that is used for data exchange between local agencies during data validation.

The data are supplied by:

- owner DPCs
- local DPCs, and
- other sources



Local DPCs are related to the service area of a DSC, e.g. for a state data storage centre all DPCs that are active within the state boundaries are local DPCs.

Owner DPCs form a subset of the local DPCs, they are the suppliers of the hydrological data in the DSC. Each owner DPC has a direct link to the DSC. At state level, owner DPCs are SGWDPC and SSWDPC; they are linked to the state DSC. At national level, a regional SW/GW DPC may be linked to its own DSC.

The users of the data include:

- owner DPCs
- local DPCs, and
- HDUs

HDU is short for hydrological data user.

Figure 2.1: Local and Owner DPCs

An overview of suppliers and users of the data stored in the DSC distinguished by categories of data is presented in Table 1.

Category	Supplier	User
Field data	**Owner DPC**	**Local DPC**
Authenticated data	**Owner DPC**	**HDU**
Objects	**Any source**	**Owner DPC**
Temporary data	**Local DPC**	**Local DPC**

Table 2.1: Data suppliers and users per data category

In the following Sections an overview is given of suppliers and users. Reference is made to Section 3.2 for definitions.

2.2 HYDROLOGICAL DATA SUPPLY

The major data supplying agencies are SSWD, CWC, SGWD, CGWB and IMD. Usually the agencies supply data through their DPCs.

The DSC receives field data and authenticated data from the owner DPC for storage in the databases for field data and for authenticated data. The local DPCs deliver field and / or authenticated data to assist the data validation at the other local DPCs. These data are loaded in the temporary repository and are discarded after finishing the data validation / processing. The validating DPC keeps track of which data are used for the data validation, in such a way that all steps in a validation process can be traced back.

The stored data do not only comprise purely hydrological data but also meta-data. One part of the meta-data, associated with the measurement, is received from the data supplier (owner DPC), together with the data values, i.e. the pure hydrological values. During processing, the DPC adds data related to the processing steps and methods and performance indicators to the meta-data. The DSC adds meta-data associated with the database and data amounts like time span of a data set, date when it was received, information about the data owner.

The intervals between successive data exchanges are relatively long. For field data a monthly interval is applied. The authenticated data are updated annually. The intervals for temporary and object data are irregular and fully demand driven.

In the chain from data entry to data validation, the major data handling software packages are: SWDES, WQDES_SW, HYMOS, WQDES_GW, GWIS.

Communication with the owner DPC is by LAN, for communication with local DPCs any convenient means of data transport can be used, e.g. LAN, email, FTP and physical media.

2.3 HYDROLOGICAL DATA USERS

Hydrological data users can request the DSC to supply specific hydrological data. The diversity among HDUs is large. It includes government and state decision-makers and hydrology professionals. HDUs can be found in organisations like: Ministry of Water Resources, (state) departments for Surface water (CWC, SSW), Groundwater (CGWB, SGW) and Meteorology (IMD). Several of these departments are suppliers of the data as well as data users.

Other data users include governmental offices and departments like: Irrigation Department, Pollution Boards, Geology and Mines Department, Water Supply bodies, Public Health, Agricultural, Industry, Fishery, Roads and Railway departments, and many other governmental and private organisations. This diversity of users has implications for the services of the DSC, which should accommodate all users.

The Data Storage Centre services following major user categories:

- **Data validating Agencies related to the service area of the DSC**

  These agencies are the owner DPCs and the local DPCs participating in the HIS, viz. SSW, SGW, CWC, CGWB and IMD. During the validation process the DPCs make use of the services of the DSC to exchange field and authenticated data to enhance the validation process. The DSC collects/receives the data from the local DPCs and temporarily stores the same for use by its owner DPCs. Further, the DSC makes the data of the DPCs available to the interested local DPCs to enhance the data validation process. The DSC keeps received data (which does not belong to the service area of the owner DPCs) separated from the field data and authenticated hydrological data under its administration. To store such data, a temporary repository is maintained.

- **Hydrological Data Users**

  This diverse category of hydrological data users has only access to authenticated hydrological data and not to the field data and the temporary data. For search and selection of the required data a Catalogue is made available. The Catalogue comprises the meta-data and a search and selection engine (browser). The result of a data selection session is request for data, which is generated in a computer file. This file is dubbed Data Request File (DRF). The DRF can be submitted to the specific data custodian DSC(s) via telecommunication, e.g. email or FTP (if available) or by post/hand carried on physical media. The DSC verifies the HDU's authorisation level to establish if the HDU is entitled to receive the requested data. For authorised HDUs the

data request may be run on the database and the results are made available to the HDU. To support all this, adequate procedures and tools have been implemented.

The data can be made available in various formats, e.g. optimised for

- HYMOS
- GWIS

and in generic formats like:

- plain ASCII
- structured CSV
- Others, yet to be defined

- **Meta-data users**

  The only direct users of the meta-data proper are the other DSCs participating in the HIS. They receive each other's meta-data on the availability of the hydrological data in the DSCs.

  The HDUs get indirect access to the meta-data, i.e. through the Catalogue. Also for the DPCs, the Catalogue is the preferred channel to access the hydrological data.

  Proper procedures and a supervising body are put in place to organise the exchange of meta-data between the DSCs.

3 DATABASES

3.1 GENERAL

In this section, technical aspects of the databases are described from the user's perspective and in generic terms only.

The databases have been founded on data structures closely related to those used by SW DES / HYMOS and GWWIS.

The system has been set-up such that the data interchange between DSCs and owner / local DPCs is easy and effective. Both the implemented data structures and the DBMS easily accommodate future expansion and adaptation to cater for the changing user needs as well as to adopt new technologies.

All data are maintained in well-defined computerised databases using the SQL Server database management system. To ensure long-term sustainability of the data. All data are properly stored and archived in compliance with specified standards so that there is no loss of information. All DSCs have implemented the same standard procedures for import and export, dissemination and backup.

3.2 MAJOR DATA GROUPS

The DSC supports and maintains separate databases for the following data groups, viz.:

- Hydrological data
  - Field data
  - Authenticated data
  - Object data
  - Temporary data

- Meta-data (derived from the hydro(geo)logical databases)
- MIS data

In each of the DSCs, the data structures are identical. This applies to all aspects of the hydro(geo)logical data and the related meta-data.

Any data value (**what**) is well defined in space (**where**) and time (**when**), but also with respect to source, measurement conditions, relationship with other data, owner, status of processing etc. This also applies to static data and semi-static data. All data have a time label to define its validity in time.

3.2.1 FIELD DATA

The field data are measurement results, which have been corrected for administrative and self evident errors only. The field data can be of any non-processed type, including:

- Geographical and space oriented data, i.e. static or semi-static data on catchment and hydrogeological features and hydraulic infrastructure
- Location oriented data, including static or semi-static data of the observation stations, wells and laboratories
- Time oriented data, covering equidistant and non-equidistant time series for all types of meteorological, climatic, surface water and groundwater quantity and quality data.

The field data has meta-data linked to it.

3.2.2 AUTHENTICATED DATA

Authenticated or processed data have successfully passed the thorough validation processes and are accepted for hydrological use. The authenticated data comprise the following types:

- Geographical and space oriented data, i.e. static or semi-static data on catchment and hydrogeological features and hydraulic infrastructure
- Location oriented data, including static or semi-static data of the observation stations and wells and laboratories
- Time oriented data, covering equidistant and non-equidistant time series for all types of meteorological, climatic, surface and groundwater quantity and quality data
- Relation oriented (derived) data on two or more variables/parameters used with respect to meteorological, climatic, water quantity, quality data
- Derived data and processing results, and
- Other results like aggregated data and extremes

3.2.3 TEMPORARY DATA

The temporary data can be of any type or category, i.e. any temporary stored data, e.g. as part of data exchange between DPCs during data validation and processing. The temporary data group is defined to distinguish from permanently stored data.

For reasons on quality assurance and traceability of processing steps, a log is kept about external data, i.e. data pertaining to owner or local DPCs, used for data validation and processing.

3.2.4 OBJECT DATA

The storage of object data in an administered repository is a service to the owner DPCs. The DSC caters for proper administration, the accessibility, archiving and backup of the objects. Only with the consent of the object proprietor, the existence / availability of objects may be publicised via the Catalogue. Maps, GIS layers, scanned maps, manuals, files, reports and other document types can be accommodated in the object repository.

3.2.5 META-DATA

The meta-data are derived from the data stored in the database and comprise information on the available field and authenticated data, object data and other categories in the DSC. Meta-data primarily identify the data in space and time, the data type(s) and the covered time period(s).

One integrating element of the HIS is the exchange of the meta-data between the DSCs. Each DSC receives the meta-data of the other DSCs and based on that, each DSC and its users can search for data that belong to any HIS station in the project area. However, users that are not involved in the data validation processes, i.e. the HDUs, only get access to the meta-data related to the authenticated data.

The meta-data is to be kept up to date. Each time new field/hydrological data is stored in the hydrological database the meta-data are updated as part of the storage process. As the meta-database also contains meta-data of the associated DSCs, a regular exchange of meta-data between the DSCs is taking place to synchronise and update the mutual meta-databases.

Depending on his authentication level, a HDU may get access to all meta-data or a subset thereof. In particular the field data and the object data have usage restriction, mostly limited to the operational area of the DSC.

3.2.6 MIS DATA

A limited number of key performance indicators are monitored by the DSC for MIS use. Incoming data flow (sources, amounts, types, dates of receipt), available data (aggregated) and meta-database status, user interactions (requests, supply destination, amounts, types) are monitored automatically.

3.3 DATA CATEGORIES

The database holds data of different categories and for different purposes. This section describes the data categories that are supported by the DSC. Addressed are the major data categories, viz.: Geographical or Space Oriented Data (the spatial aspects), Location Oriented Data (site specific aspects), Time Oriented Data (time series), Relation oriented data (derived data)

The static and semi static data describe the origin of the hydrological data in their full context, and how the data were collected, when, by which agency and what data collection/measurement method was used. Hence, the (semi)static data are part of the input data. The data sources are state and national agencies that are active in the service area of the DSC. For SW these are primarily the State Surface Water Department, CWC and IMD. For GW these are State Groundwater Department, CGWB and IMD.

The various data categories, i.e. geographical or space oriented data, location oriented data, time oriented data and relation oriented data are described below.

3.3.1 GEOGRAPHICAL OR SPACE ORIENTED DATA

Geographical or space oriented data comprise static or semi-static data of the following kind:

- Administrative maps,
- Physical maps
- Geological maps,
- Hydro-geological maps,
- basin descriptive data, and
- hydraulic infrastructure

Maps and GIS layers

Basin and hydro-geological features are stored in the form of maps related to geography, topography (contours), rock types, land use, layout of hydraulic infrastructure, catchment boundaries, major administrative and political boundaries, measurement locations, location of structures, industries, etc. Thematic map data are held in layers or shapes displayable either in single or multiple layers. Other varieties are contour maps, sections, diagrams, images, basins and others, DTMs, DEMs, and scanned maps.

Maps are geographically referenced, and geographically referenced measurement locations added to the database are automatically added to the map base.

Maps and GIS layers are stored as objects, including the descriptive data required for their interpretation and importing into the related application (GIS) packages. Processing of maps is not a function of the DSC.

Basin descriptive data

Basin descriptive data include:

- tables of fixed sets of parameters describing linear, areal and relief aspects of the drainage basin and/or channel network, geology, hydro-geology, and
- text files comprising free text on relevant basin features.

Hydraulic infrastructure

Data on the hydraulic infrastructure comprise:

- historical records of survey data on longitudinal and cross-sectional profiles,
- longitudinal profiles of a fixed set of hydraulic and geomorphologic characteristics of river bed and banks.

3.3.2 LOCATION ORIENTED DATA

These data comprise a wide range of static and semi-static information at point locations related to:

- observation stations,
- observation/piezometer wells, and
- hydraulic structures.

Observation stations / wells

In this context, station is interpreted as the place of measurement, e.g. surface water gauging stations, observation wells, piezometers and meteorological observation stations. The software has the facility to hold data of the following general types.

- Identification name and code - by which the station is identified in the Catalogue.
- Summary station data - on location (e.g. latitude, longitude, altitude, river name, basin name, aquifer identification, administrative and political regions and responsible agency), for which a facility is available to group stations for reports.
- Station description - information that comprises all relevant data about the site and its surroundings and may include:
  - site, channel and control (hydraulic and morphological conditions) description
  - station access
  - benchmarks locations and levels
  - facilities and equipment in use
  - record of repair, maintenance and replacement of equipment
- Station log - information related to the interpretation, reliability and processing of a record over specified periods of time. The record can be held in a structured database file.
- Survey records - reference to survey data on longitudinal profiles and cross sectional profiles at and adjacent to the gauging station available under hydraulic infrastructure/space oriented data.
- Well logs - well logs of groundwater observations wells and piezometers are stored allowing preparation of detailed well log reports and diagrams.
- List of station series - a listing of the time series available at a station including a descriptor, units, reference, time interval of measurement, acceptable minimum and maximum values, acceptable rate of rise and rate of fall, identification codes for faulty and missing data and start and end dates. The last dates are automatically updated.

Hydraulic structures

Features of hydraulic structures (barrages, dams, inlet structures, outfalls, culverts, bridges, etc.) are stored in a fixed set of parameters such as location, type, geometry, discharge characteristics, etc.

3.3.3 TIME ORIENTED DATA

Time oriented data covers equidistant and non-equidistant time series for all types of meteorological, climatic, surface and groundwater data, including water quality.

The DSC has the facility to store time oriented data of the following types:

- equidistant time series, and
- non-equidistant time series.

A provision is available to flag the data with respect to whether it is original or estimated and with respect to its quality.

Equidistant time series

Records with a fixed time interval (or with a group of unequal time intervals, which repeats in a perfect cyclic manner) are automatically time-labelled and facilities are available to hold records for a measurement interval of 1 minute to 1 year. Such records include all possible periodically measured instantaneous observations (e.g. stage observations), accumulative observations over the time

interval (e.g. daily rainfall) and average of observations (e.g. daily mean flow). A facility is available to distinguish such records. The user is free to choose an appropriate indicator for missing values.

When imported in a Relational Database System, the overhead resulting from storage of date and time, which is largely redundant information, can grow to enormous proportions. To avoid that waste a time series vector (TSV) is defined as a container to hold equidistant time series data. That TSV contains a fair amount of simple values, covering a standard time period of a day, week, 30 days or multiples thereof, in any case a period that is related to daily life. For proper interpretation of the TSV, it is associated with a reference time (date/time of first sample), the value of the time increment and the number of readings/values in the vector. Further, a packing / unpacking method is associated with the TSV. The TSV object is stored in a compressed format. The TSVs are held in BLOBs (Binary Large OBject). The use of BLOB improves the storage performance at the cost of increased processor load. The result is a much smaller and faster database. The use of BLOB technology has no benefit when only individual items out of TSV have to be selected. A SQL query on the contents of the BLOBs is not supported.

Non-equidistant time series

Records taken with an unequally spaced time interval are stored with an entered time label and include:

- event-based observations (e.g. occurrence of a 1 mm tip of a tipping bucket raingauge. It should be noted that the HIS standard specifies storage of rainfall data as equidistant time series.)
- threshold-based observations (e.g. occurrence of the change in a variable exceeding a specified magnitude).
- constant observations (e.g. gate levels with constant level or fixed number of pumps operating for specified time).

3.3.4 RELATION ORIENTED DATA

Relation oriented data on two or more variables/parameters are used with respect to meteorological, climatic, water quantity, quality data. The software has the ability to store relation oriented data of the following kind:

- profile measurement data,
- sets of two or more quantities observed concurrently, and
- the parameters of the relationships between two or more quantities.

Profile measurement data

Profile measurement data include streamflow and sediment transport measurement data obtained by measuring at a number of points in the cross-section. The results of streamflow measurements are stored according to the velocity area and float method. Similarly, the software accommodates for suspended and bed load transport measurement data.

Concurrent observations

Records of this type consist of a time label with two or more concurrently observed/computed quantities at a station, like current meter measurement summary data along with a flag for qualification.

Relationship parameters

These records store the parameters of a relation e.g. stage-relation curves, current meter ratings, stage-discharge rating curves. Not less than four equations are available for a relation to cover the full range of the independent variable. The records include a station name/instrument code, validity period of the relation, type of equation, equation boundaries, parameter values and summary error statistics for each equation. Facility for flagging a relationship for making a distinction between a good quality and a doubtful relationship is available.

3.4 PRIMARY ENTITIES

Next lists of primary data entities are given.

The entities are organised per hydrological discipline, viz.: surface water, groundwater hydro-meteorological data, water quality data and map data. The water quality data is stored as category of surface water data and groundwater data.

3.4.1 SURFACE WATER DATA

- Catchment (basin) data
- Hydraulic infrastructure
- Station location and site description
- Station technical description
- Equipment description and data
- Equidistant and non-equidistant time series data for:
  - Water levels
  - Flow measurements
  - Discharges
  - Water quality variables
  - Sediment samples and sediment concentrations
  - Relation oriented data
- Statistical and aggregated data.

3.4.2 GROUND WATER DATA

- Well location and site description
- Well technical description
- Well assembly and performance data
- Aquifer data
- Lithology
- Well hydraulics
- Equipment description and data
- Equidistant and non equidistant time series data
  - Water level
  - Well performance
  - Water quality parameters
- Statistical and aggregated data
- Groundwater resource assessment

.

3.4.3 HYDRO-METEOROLOGY DATA

- Station location and site description
- Station technical description
- Equipment description and data.
- Time series data for meteorological related data as:
  - precipitation
  - temperature
  - atmospheric pressure
  - wind
  - radiation
  - humidity
  - evaporation
  - relation oriented data
- Statistical and aggregated hydro-meteorological data.

3.4.4 WATER QUALITY DATA

- Laboratory location and particulars
- Collection procedures and guidelines
- Methods of analysis
- Physical Analysis
- Chemical and Biological Analysis organic and inorganic (level 1, level 2, level 2+).

3.4.5 MAP DATA

- Contour maps
- Groundwater availability maps
- Hydrogeological maps
- Water quality maps, and
- Other thematic maps.

4 ARCHITECTURE OF THE DATA STORAGE CENTRE

This Section addresses the main functions and data flows in the DSC in relation to the data providers (DPCs), data users (HDUs) and other DSCs. A concise description of the DSC, incl. functional diagrams indicating the main functions and interfaces with closely related organisations and users are presented.

4.1 DATA OWNERSHIP

The DSC stores and administers the storage of all field and authenticated (processed) hydrological data collected in its service area. The owner DPC(s) supply the data. The DSC stores and administers that data and makes the same available to authorised Hydrological Data Users. It also maintains a meta-database of all data stored in its own databases and the data stored in the databases of the other DSCs participating in the HIS.

The data storage under HIS is essentially based on a distributed network of databases. The databases reside in DSCs. Each DSC has a specific service area essentially the operational areas of the owner DPCs and the data collection stations belonging to these areas. The DSC stores the data

originating form a unique set of data sources, i.e. data collecting stations and derived data pertaining to the same stations and combinations of stations. Several agencies may operate in the same basin or aquifer; but each agency has the responsibility for the data collection stations owned and operated by it.

There is no duplication of hydrological data, in the sense that only the custodian of a data object is entitled to distribute that object.

Any DSC can take part in the exchange of data between data providers and data users

4.2 INPUT DATA FLOW

The DSC imports the following categories of data:

- **Field data**

  The field data are measurement results, after correction for administrative and evident errors. The field data are supplied by the owner DPC(s)

- **Temporary data**

  The temporary data are supplied by the local DPCs for validation purposes only. After use, the temporary data are discarded. The temporary data do not belong to the owner DPCs but to local DPCs. In a DSC, data that are denominated temporary data if they do not belong to that DSC or the owner DPC. That very same data may be registered as field / authenticated in its originating DPC / DSC.

- **Authenticated (validated and processed) data**

  This data has successfully passed the thorough validation processes and is accepted for hydrological use. The authenticated data is supplied by the owner DPC(s)

- **Object data**

  The object data is delivered by the owner DPC for storage, the data may comprise any data object that is not included in the hydrological databases (field and authenticated), but is of enough importance to store and administer it properly for later use and reference. The object data can be made available to selected HDUs only.

- **Meta-data from other DSCs**

  The meta-data identifies all available data in the DSC. Each DSC receives the meta-data of the other DSCs and based on that, each DSC and its users can search for data that belong to any HIS station in the project area. Each DSC generates the meta-data related to the stored hydrological data.

- **Data requests**

  Requests for data are received form Hydrological Data Users. The requests accurately specify which hydrological data the HDU has selected for retrieval. The requests are generated by the search and selection tool, which accompanies the Catalogue on the hydrological data.

4.3 OUTPUT DATA FLOW

The following data categories are exported or produced:

- **Requested data**

  The requested data are delivered to authenticated HDUs only. In support of this proper authentication, and where needed, encryption procedures are implemented. The HDU is given a choice out of a series of supported formats.

- **Meta-data**

  The meta-data is made available to the other DSCs that participate in the HIS. Since all DSCs apply identical structures and tools, the meta-data can be exchanged in a convenient and effective format.

- **Reports**

  The DSC generates standard reports on its operations and the available data. In particular the data flow (in and out) and the amounts of stored data are reported in a categorised manner. The reports also include the activity and access logs.

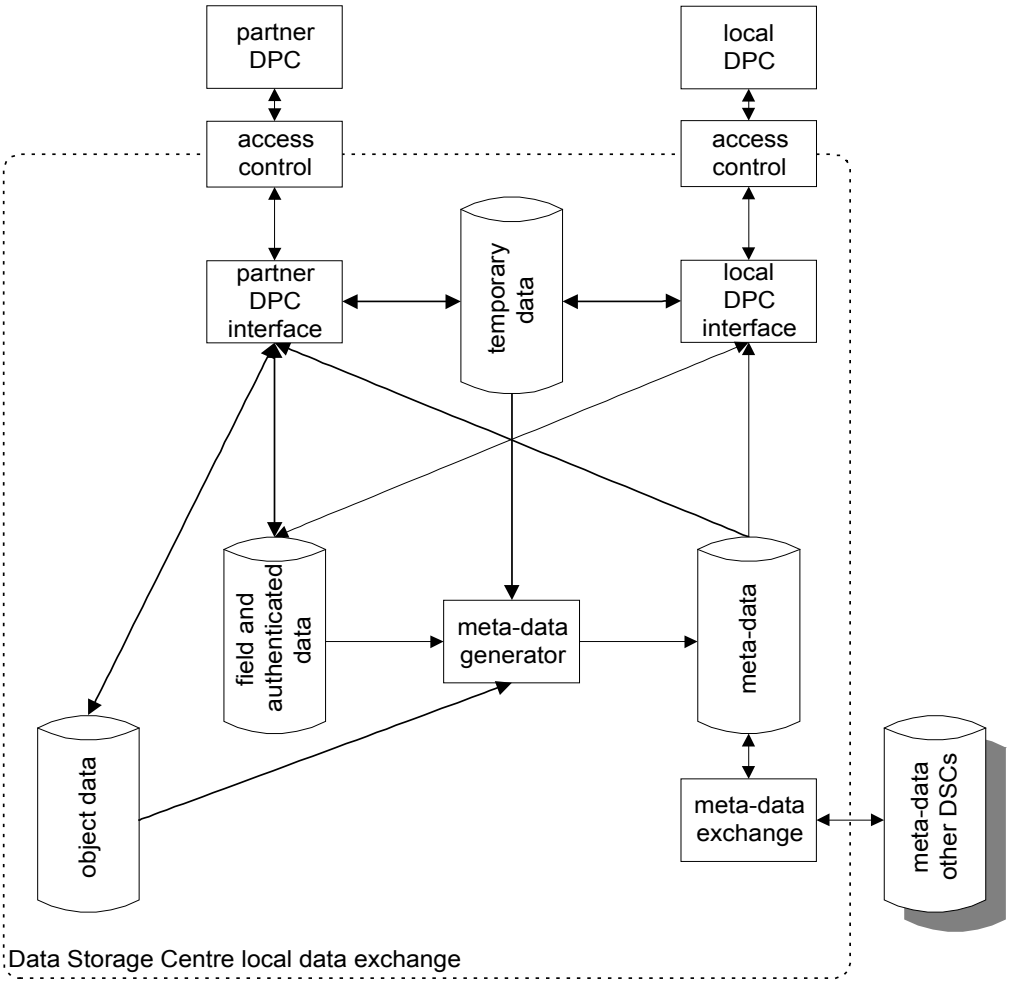


Figure 4.1: DSC local data exchange

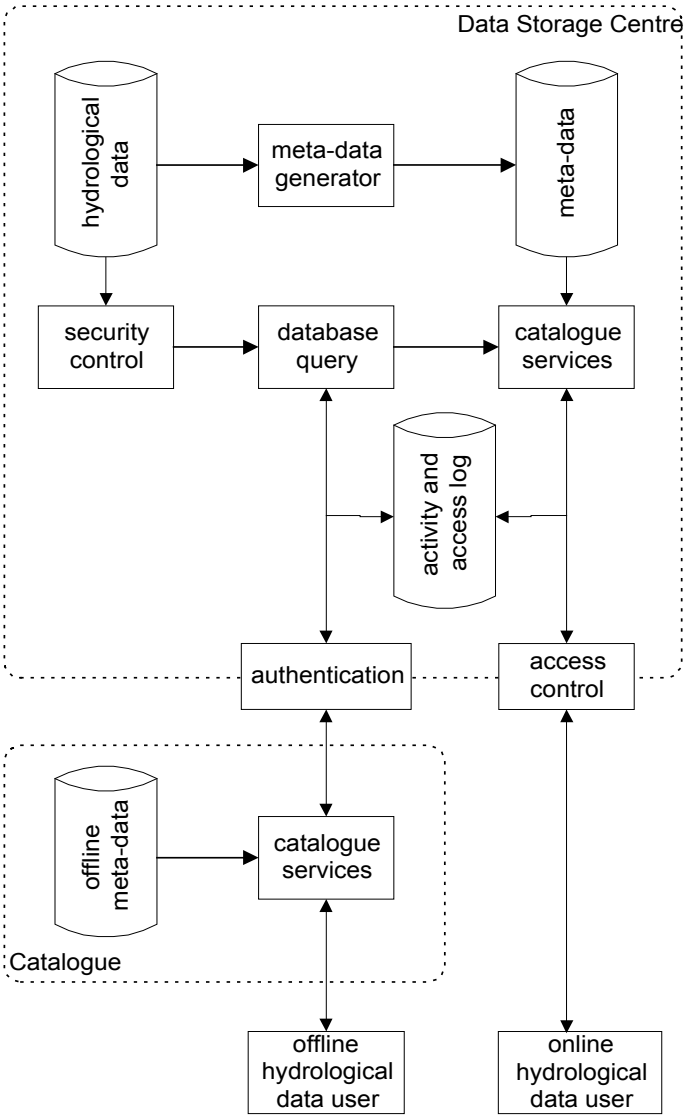


Figure 4.1: DSC local data exchange

4.4 META-DATA AND THE CATALOGUE

The information about the stored data is maintained in a meta-database, in combination with a search and selection tool that makes the Catalogue. Each DSC maintains the meta-data on the locally stored data and shares meta-data with all other DSCs. Consequently, the combined meta-data sets of all DSCs contain the information on all stored data.

The DSC should be perceived as a library of computerised hydrological data. Processing of the hydrological data is not a responsibility of the DSC, that is, the DSC is entirely dedicated to data storage.

4.5 VALIDATION SERVICES

During validation and processing of field data, DPCs may need to exchange data with other DPCs that operate in the same basin/aquifer. That data exchange is executed on a demand basis and is routed via the DSCs, i.e. the DSC acts as a hub in the data exchange network.

The DSC maintains a separate storage section, the temporary repository, for this service and the data will be discarded after finishing the data validation processing. The temporary data may not be mixed with the field data and the authenticated data in the hydrological database.

The hydrological data is the data featuring in the Catalogue of the DSC.

The temporary data is received from other DSCs / DPCs and is used to enhance the data validation and processing. In the supplying DSC the data may have the status of field data or authenticated data. In the receiving DSC that status is retained, however, since the receiving DSC is not the custodian of the temporary data, the data may be used but not added to the groups of data available for HDUs. As a result, the data may not be included in the Catalogue as data that is maintained by the receiving DSCs, but only in the Catalogue of the originating DSC.

Owners of the DSCs

There is a variety of DSCs 28 of which service one or more owner DPCs. The two DSCs located at the NIH and the NWA are for imparting training and validating the software. The DSC at the NIC is provides on-line catalogue for the central web server.

Owner DPCs	DSC Nos.
SSWD	2
SGWD	2
SSWD and SGWD	7
CWC	6
CGWB	10
IMD	1
NIH / NIC / NWA	**3**
Total	**31**

Table 4.1: Data Storage Centres and numbers

Need to incorporate the three other DSC's which do not actually store data (except for public domain data)

From Table 4.1 it follows that several types of DSCs have been implemented, each with specific owner DPCs. The main difference between the DSCs is that in each DSC another set of data is stored. In fact, this also applies to DSCs of similar type to cater for the hydrogeological features in the DSC's operational area. As a result, each DSC essentially supports identical database structures but of different implementation.

In most states, a single DSC caters both SSWDPC and SGWDPC, however, in some states these DPCs are physically too far apart to combine the data storage needs in a single centre, they have an individual DSC instead.

The national agencies maintain DSCs at two levels, i.e. at regional level and at national level. The national DSCs are quite comparable with the similar state DSCs; the amounts of stored data are larger. The national DSC may not have original data, but copies of data originating form the regional DSCs that store the data for the (regional) owner DPCs.

The data in the national DSCs may be included in the Catalogue as available from the national DSC. However, such data should be clearly marked indicating that the original data is in the custody of a regional DSC. This because any data item can be owned by only one DSC, i.e. the DSC associated with the originating DPC. That regional DSC should also be identified to allow the HDU to verify if that

particular DSC keeps an updated version of the (part of) the required data. This information should be available from the Catalogue, in any case of the online Catalogue version.

4.6 DATA STORAGE CENTRE RESPONSIBILITIES

The Data Storage Centre maintains and manages the databases for the data emanating from the participating DPCs. In Sections 7.1 and 7.2 the DSC responsibilities are formulated in some detail.

4.7 DSC FUNCTIONALITY

The main objective of the HIS Database is to be a Data Store for hydrological, geo-hydrological and hydro-meteorological data, as to enable data retrieval by State/Government decision makers and other authorised professional users, which require such data. In order to meet this objective, the HIS includes all functions needed to create, maintain, distribute and integrate data into a sustainable system as well as to provide comprehensive services to the users.

HIS DSC main functions:

- Database periodic updating
- House keeping modules for database maintenance and publishing
- Meta-data updating and viewing
- Integration of Catalogue data from all the other participating states into the system
- Production of periodic reports (pre-designed)
- Special Queries for Information (pre-designed)
- Import/Export of data files upon request
- Conversion of standard files into HIS data structure
- Conversion of HIS data structure into standard ASCII and CSV data files
- Basic modules for Object data retrieval
- Data security module for users authentication and authorisation check
- Activity Log function
- Accounting module (could be used for commercial purposes if so decided)
- Internet connectivity function
- Query and Report Generation tools and functions
- Audit module.

5 DSC FUNCTIONS AND INTERFACES

The DSC executes external functions for which interaction with other centres and users is required. An important function is data exchange, including data transfer, security control and data format conversion.

The internal functions are more related to the day to day operation and administration of the DSC. A common element is maintenance of the database operations at a proper level and safeguarding of the data.

The most important external and internal functions and the related interfaces are described in this Section.

5.1 EXTERNAL FUNCTIONS

The external functions cater for the interfacing with the Data Processing Centres, Hydrological Data Users and the other Data Storage Centres. The following major functions are addressed: import, conversion, formatting, hub service, meta-data exchange, dissemination of the Catalogue, handling of data requests, data export, security and user support. These functions are optimised for best performance with the DPCs and HDUs.

Data Import

The owner DPCs deliver field data and authenticated data for storage by the DSC. The field data is received from the owner (originating) DPCs, which obtain this data from their monitoring networks. Like the field data, the authenticated data are received from the owner DPCs, after thorough processing. The DSC is not in the position to verify the data validity. It is the responsibility of the agencies to exercise proper scrutiny and validation procedures as agreed under the Hydrology Project.

The DSC, however, is responsible for storage of the data in the correct format and with unique labels linked to it. Upon reception of data, the DSC assesses the data integrity. The integrity rules cover the administrative discrepancies and apparent errors. The rules have to be finalised in the tailoring phase. Any inconsistencies or errors have to be communicated with the data owner. Only when the data are error free, properly formatted and consistently identified the storage process may proceed and the HIS data base updated. For this adequate functions have been implemented in the DSC software. If existing data is replaced, then this should be reported to the DPCs, possibly also the authenticated data have to be recalculated / replaced.

Data in several formats can be accepted, e.g. SWDES / GWDES native format, SWDPC / GWIS export format, IMD Transfer format and as ASCII files. Other convenient file formats may be defined during the design and implementation phase.

Data conversion

The owner DPCs, which are the originators of the data, produce data in compliance with SI standards only. This also applies to historical data, which is converted to SI units during data entry. Hence, there is no reason for conversion to other units. In exceptional cases, data may be scaled to more convenient units.

Data formatting

In support of the import and export services, adequate formatting procedures are applied on the data. Formatting is related to the way the data are stored by the DSC and by the DPC / HDU respectively. The DSC is a service provider to the DPCs and the HDUs. In that capacity, the DSC provides the conversion procedures aiming at minimum effort for the DPCs and the HDUs.

The DPCs export data in a convenient format, e.g. native, ASCII / CSV, various file formats, etc. Where needed, the DSC converts the data before importing them in the database. Also, when delivering data to DPCs or HDUs, the data are converted to a format that is convenient for the recipient. The HDUs are offered a choice out of supported data formats.

The supported data and file formats were decided during the system design / tailoring. Tools are available to design new formats / templates, both for import and export of data.

Hub service (exchange of temporary data)

The hub service is provided to the owner DPCs and the local DPCs in the operational area of the DSC. The hub service implies collecting/receiving data from other DSCs/DPCs and storing the same in temporary storage for later dispatch to the requesters. Also the reverse function is supported, i.e. requested data are made available to other DSCs, where needed and feasible, data are delivered in the required formats. A separate repository is kept available in support of this service.

Maintenance and exchange of the meta-database

The meta-database is of paramount importance for the (potential) data user and accurately reflects the availability of data. The meta-database is accompanied with easy to use tools for maintenance, exchange with other DSCs and integration of meta-data. The on-line version of the meta-data accurately reflects the state of the databases residing in each of the DSCs. All DSCs maintain the same structures for the meta-database. The meta-database can be exchanged between DSC's and the latest version for each DSC has to be retained.

Dissemination of the Catalogue

The Catalogue is distributed on request to the State and Central Organisations and other authorised HDUs. It contains the combined meta-data for each of the DSCs. And on top of that, a search and selection tool which integrates the meta-data to a single source of information for the HDU.

The DSC exchanges meta-data with the other DSC and verifies the integrity of the meta-data. Subsequently, the meta-data are set on-line, together with the search and selection engine, for HDU access via Intranet or Internet.

The DSC makes new versions of the meta-database available via Internet / FTP. The size of updates is similar to the size of new versions (compressed ASCII), therefore only new versions are made available. However, in case it is required at some stage that incremental updates be provided, the same should be possible.

Processing of requests for data

Based on the Catalogue services, the HDU generates a data request file (DRF), for processing by the DSC. If the DRF is valid (integrity check) and the HDU authorised, then a database run is executed. The retrieved data, if any, are assembled for delivery.

Export of hydrological data

The DSCs monitor and control the data retrieval by the State and Central Organisations as well as the HDUs. The retrieved data may be supplied in a way, which most suits the HDU, e.g., by file transfer or on physical media. Encryption to be applied where needed.

Security on data access

The DSC implements a data security system to protect the data against tampering, damage, loss and unwanted use.

DPC information

Owner DPCs are kept updated about the data availability and interactions with the DSC. For this the DSC generates adequate reports.

User support

HDUs may run into difficulties while using the Catalogue services, receiving the data or loading the data into their processing system. Problems can have their origin in the DSC, in the data transfer and on the HDU's end. To properly assist the HDU, the DSC maintains a properly equipped helpdesk. The helpdesk staff has access to relevant manuals and support information.

The DSC also has facilities to assist HDUs with data formatting difficulties. In the rare case that a HDU has no way to handle the available data formats, the DSC can mould the data in the required format.

In order to continuously improve the performance of the DSC and to detect shifting user's requirements, the DSC monitors and analyses the interactions with the users and, whenever needed adapts procedures, introduces new functions, new data types etc. The developments in the data communication industry also affect the user's requirements and adaptation of the communication channels is occasionally needed. The design of the DSC allows for easy adaptation. All DSCs participate in the ongoing adaptation to the user's needs.

Data processing

The DSC administers, stores and distributes data but does not take part in data processing and validation. The DSC is not authorised to alter any data in its administration.

However, in case some data has been imported incorrectly, data deletion with adequate security and audit checks is required. This applies to masters also.

Communication DSCs

The DSCs adapt the access channels and the changing user's requirements in a concerted manner.

5.2 INTERNAL FUNCTIONS

The internal functions are basically the day to day operations related to data storage, backup and archiving processes. Further, the DSC staff maintains the software and hardware systems, this also includes replacement of outdated components, installation of new releases of the database structure and the DBMS software with supporting tools.

Management

Management of the database is one of the functions of the DSC, it includes:

- Updating of the data availability records.
- Addition of authorised new users to the database and maintenance of an Authorised users list (for both internal users and HDU's)
- Provision of service to authorised users upon request, extracting data files from the HIS database and translating the data to standard ASCII files and defined formats

- Communication with other Data Storage Centres in the HIS.

Creation and maintenance of the hydrological databases

After commissioning of the DSC, there will be an ongoing need for adaptation of the structure of the databases, addition of new tables (e.g. new data sources) addition of new data types, tuning of the databases, etc. The DSC implementation has ample flexibility to allow these adaptations.

Loading of the data

The loading of the data is executed in compliance with strict rules to safeguard the data integrity. A major function is the translation of data files generated by the dedicated software into HIS standard database structure. However, data validation and processing are not part of the DSC's mandate.

Maintenance and safeguarding of the stored data

All stored data are protected against loss due to any hazard (fire, earthquake, explosion, inundation, theft, power supply problems, fungi, moisture, etc.). After a failure or disaster, it is possible to restore the latest backup to bring the databases to a controlled state. It is possible to regenerate lost index tables.

Operation of an archive and a backup system

Archiving of the contents of the databases at scheduled intervals and ad-hoc is supported. An accurate archive administration system has been implemented. Archives are kept at safe places, away from the DSC compound.

The ArcServe backup system allows for a friendly user interface to take backups, both complete and incremental, scheduled and as-and-when on any backup device in the system. The restore process can also be easily controlled for proper recovery in case of corruption or other disasters.

The system has also incorporated the feature of combining the existing data with the restored database to create a single database.

Creation and maintenance of a local meta-database

The meta-database, reflects the actual state of the various hydrological databases. Any changes in the hydrological database are to be reflected in a meta-database up-date. The up-date is manually triggered.

The on-line version of the meta-database on LAN in the DSC is up-to-date, the on-line Catalogue will only be up-to-date after synchronising. Meta change database should also incorporate the feature of tracking what data has been archived.

Monitoring of the data flow and data amounts in the DSC for MIS

For internal / external use, some key performance indicators are collected and stored. Of particular interest are quantities of received, stored data and supplied data. These quantities are not expressed in number of bytes but in period of time they cover. It cannot be seen when the data were delivered at the DSC, nor was it on time or too late. Other indicators related to the user's interaction are also monitored and logged.

Preparation of reports

Normal database related reporting has been implemented. For the DPCs and HDUs these also include data availability reports, data supply reports, data access reports and data retrieval reports, e.g. on monthly and annual basis.

5.3 INTERFACING ORGANISATIONS

The DSC maintains interfaces with the owner DPC(s), the other DSCs and the HDUs. Via these interfaces, data are exchanged via various communication channels, each with its peculiarities. The data exchange requirements also vary from event to event. The DSC supports a range of data communication channels, in hardware, software and associated protocols, to cater for the needs of the data users and other related organisations. Proper conversion / formatting tools are implemented to meet the requirements of the DSC's owners and clients, in particular for importing and exporting of the data. All data exchange is supported by intuitive user interfaces without bothering the user with the technicalities of data storage and retrieval.

The access rights to the DSC that belong to the interfacing organisations are listed in the Table 5.1 in relation to the data storage facilities. This Section addresses a number of interfacing aspects.

interfaces with:	owner DPC	local DPC	other DPC	DSC	national DSC	HDU
field data	access all write own	access all	none	exchange on request	none	none
authenticated data	access all write own	access all	access if permitted	exchange on request	exchange on request	access if permitted
public domain data	freely distributed	freely distributed	freely distributed	freely distributed	freely distributed	freely distributed
meta-data	access all	access all	access all	up-date and access all scheduled exchange	access all scheduled exchange	access all
object database	access all write own	access if permitted	access if permitted	exchange on request	exchange on request	access if permitted
temporary storage	exchange	exchange	none	none	none	none
communication channel	LAN	telephone physical media	telephone physical media	telephone physical media	telephone physical media	telephone physical media

Note: the national DSCs will have access rights on all data of their regional DPCs.

Table 5.1: Data access by interfacing organisations

Each DSC has one or more **owner DPCs**. Owner DPCs have a direct LAN connection with the DSC. Usually they reside in the same building. At state level, owner DPCs are SGWDPC and SSWDPC, which are linked to the state DSC. At national level, a regional SW/GW DPC may be linked to its own DSC.

Nowhere in the HIS two DSCs can be owner of the same data. However, it is possible that a DSC keeps a copy of a certain data set, that data it may have available for its owner DPC, in that case the owner DPC is a HDU. The National Central Agencies have the right to disseminate data pertaining to their local offices.

Local DPCs are related to the service area of a DSC, e.g. a national agency that has a data collection net work in the service are of the state DSC (hence within the state or regional boundaries) is a local DPC. As a result, it gets rights to exchange data for validation purposes.

Write own data: Data can be written/altered by the owners only. No data item has more than one owner. The 'write own' permission is the prerogative of the owner DPCs.

Access implies read and retrieve of certain data, addition/change/removal of data is not permitted.

Access if permitted is supported for most of the other HDUs/DPCs. Depending on the authorisation of such users, they may be faced with restrictions and get limited read / retrieval rights, possibly only for a part of the data like a basin or a district and / or certain periods.

Owner DPCs and local DPC may freely **exchange** data via the DSC hub function. The Catalogue does not support such data exchange.

LAN: local area networks, i.e. the owner DPCs have LAN connection to the DSC.

Telephone: this can be any convenient communication channel, e.g. STD, ISDN, leased line and others. Data transport can be via email, FTP or Internet services.

Physical **media**: data transported on physical media. Transport on physical media is very cost effective when data volumes are not very small. For very small and/or urgent exchange of data, the telephone channel is most appropriate.

5.3.1 OTHER DATA STORAGE CENTRES

At fixed intervals, each DSC dispatches the meta-database belonging to its own local data (data that is owned by the DSC owner DPCs) to the other DSCs for their perusal. For this a co-ordination and control system is to be established and maintained.

As the DSCs also act as service providers to other HDUs, they may make data available to other DSCs, when requested to do so. E.g. when a DSC receives a request for data belonging to multiple DSCs, one centre may take charge and collect all the required data from the various DSCs. Security procedures have been implemented. In a distant future, all data dissemination may be channelled through a single supervising hydrological organisation.

The HDU's request for data will be the result of search and selection run on the Catalogue. The exchange of meta-data takes place between the DSCs. None of these data exchanges needs to be real time.

A convenient data transport mechanism has to be set-up. This not necessarily has to be a costly high-speed communication system.

The Data communication system can change with time as technology evolves. The system should be able, as far as possible, to use any communication medium effectively.

5.3.2 DATA PROCESSING CENTRES

The **owner DPC** has direct access to the DSC via a LAN connection. The response times are short, virtually real time, in particular during Catalogue use.

For data storage, rigorous integrity checking procedures have been implemented in the DSC. Flaws are reported to the delivering DPC.

User interfaces have been implemented on the DPC computers for: data delivery / storage, data search and selection (the Catalogue) and data retrieval. Reporting systems have been implemented to keep the owner DPC fully updated.

The **local DPCs** may communicate via any convenient channel. For exchange of short messages telephone connection (whatever the variety) is supported, in hardware and software. Response times on service request are kept as short as possible. The local DPC may deliver (field / authenticated) data for use by the other (owner / local) DPCs for data validation purposes.

For both owner and local DPC categories, the Catalogue supports all data categories they have permission to.

The other DPCs are regarded as HDUs. The other DPCs primarily use the Catalogue functions. In the HDU data category, they have high permission levels for most of the data.

5.3.3 HYDROLOGICAL DATA USERS

Hydrological Data User is a generic denomination for any organisation / person that uses hydrological data. The HIS DPCs can be HDUs and request for data from DSCs.

The primary tool for the HDU is the Catalogue. The Catalogue services support search and selection of needed data in the meta-data sets of all the participating DSCs. Further the Catalogue comprises tools for data exchange with DSCs.

Major supported data exchange functions are: updates of the meta-data, dispatch of a data request file and reception of delivered data. In Section 8, the Catalogue is described in detail.

Any privileged HDU may submit a request for data, usually by dispatch of a DRF. The Database Administrator verifies the HDU's privileges for the requested data. A HDU may only receive data in compliance with his level of authorisation. To maintain sufficient security levels, the HDU identification may be encrypted.

The DRF is routed directly to the concerned DSC who is then asked to service the HDU direct. The data can never come to the controlling DSC. The information that the DRF has been received is, available with the DSC, though.

5.4 DATA TRANSPORT CHANNELS

The principal external data exchange is based on file transfer. A major advantage of the file transfer system is that it is virtually independent of the communication infrastructure, i.e. it can be based on transport by physical media but also by high-speed data communication or VSAT.

Data is periodically transported from the DPCs to the DSCs, and, upon request, from the DSC to the HDU as per procedures laid down in Section 7.5. The transport of data supports the specified data structures and formats. Each data transfer is checked for transfer errors and completion through a data transfer protocol. The data transfer protocol includes identification of the sender, the receiver, record count, and data characteristics, identifying data type. The data transport is based on standard commercial protocols, without the use of proprietary software or protocols.

In any of the modes of data transfer, adequate error control have been implemented.

The DSC is equipped with one or more external telephone lines to allow multiple communications simultaneously. The number of lines and the types have been decided by

NIC role needs to be defined in detail. The catalogue will be generated at the NDCs located at CWC, Delhi and CGWB, Faridabad and uploaded onto the NIC web server.

LAN

Owner DPCs get access via LAN, this is in particular useful for the SSW DPC and the SGW DPC. Although the owner DPCs are the owners of the data, security measures and access control have been implemented on the network(s), also including online virus protection.

Data requests by the DPCs, e.g. generated using the Catalogue, are processed on line through immediate data servicing and data view tools.

Leased Line

In some cases data transfer via leased line would be most effective to link DPCs with the DSC. The DSCs software and hardware (where applicable) supports dial-up communication.

Dial-up connection

The DSC software supports dial-up (PSTN / STD) connection with other DSCs and DPCs. Data transfer via FTP and email is supported. The commonly used tools are included with the system. In addition, the software supports dial-up via ISDN using a digital modem. For more demanding dial-up data communication, a gradual shift form STD to ISDN is to be expected in the Capitals. The DSC may need to be adapted for that.

CD-ROM

Since the CD-ROM is widely accepted as a physical data carrier and has very low recurring costs, it is used for distribution of the Catalogue to (potential) HDUs. The CD-ROM contains the meta-data and a search and selection engine. Next to that, several tools for effective data exchange and security control have been included on the CD-ROM.

The CD-R / CD-R/W technology is also used for transfer of large amounts of data.

A signature/checksum accompanies data on physical media. This allows verification of the authenticity of the data, i.e. if no errors were introduced or tampering took place.

File Transfer Protocol

With all participating agencies powerful and user friendly FTP software has been implemented. For non-HIS HDUs, the FTP software has also been included in the Catalogue CD-ROM.

E-mail

With all participating agencies a user friendly POP3 type email server system has been implemented. For non-HIS HDUs, the POP3 client software is also included in the Catalogue CD-ROM.

Web interface

For remote users a web browser has been included on the CD-ROM. The web browser has the capability to launch the email and/or FTP functions when needed. At each DSC appropriate web pages have been implemented. The web pages are used to update the HDUs on latest developments, give access to downloadable software and meta-data updates. Further the Catalogue may be accessed online. The online Catalogue interface has the same look and feel as the CD-ROM based Catalogue interface.

5.5 SECURITY

The DSC implements various levels of a user authorisation system, access control by firewall and virus protection. HDU requests for data (DRF) are checked against a User Authorisation Table and access to data are only permitted after positive identification and authentication of the HDU and his specific privileges. The database design supports multiple security / access levels. HDUs holding a certain authentication level may only get data belonging to that level and data of lower levels. The data owners specify the data security levels and the user authentication. The access to the metadata is free. Data is disseminated only after authentication.

It is recommended that the web server at the DSC, if implemented, not be on the same network as the database server.

The following security checks have been implemented:

Users Authorisation - User login to the system are checked, identified and authenticated, by user-ID (user-id) and password.

The **user ID and password** is compliant to basic ISO Data Security Standard (describing the password structure, length, revoking time, etc.). Each DSC has its local Authorised User Table, and thus is able to decide about user's privileges.

User Registration – The system supports registration of the user's data requests and the deliveries.

Data access privileges are follows:

- Database use
  - Search and selection (Catalogue access only)
  - data retrieval (read and retrieve privilege)
  - database update (read, write, update and delete privilege)

- Database management
  - Performing of data management and administration tasks (special supervisory privilege)
  - Privilege for using the standard SQL query generator and report generator option for dynamic production of queries and reports (pre-designed queries and reports are used by all users)

Auditing function: Displaying, logging and reporting of all activities by: date, time, user-id, updates and deletes from the database.

Gateway: As part of a security scheme the DSCs may communicate with each other via a central gateway. However, this does not impair the data exchange with the DPCs and the DSCs. The data exchange should not be entirely dependant on that gateway, always alternative methods of communication should be accessible.

Firewall: A firewall further limits the access to the DSC functions. The firewall functions at IP-level (verifying IP and TCP addresses) and at application level. The user may get to certain services only, e.g. Catalogue and FTP services.

Virus protection: A virus protection shield has been implemented on all in coming and out going data. On the LAN and connected stations, a continuous rigorous virus monitoring and scanning system is operated. All computers, both standalone and connected to the LAN, are protected against virus.

Encryption: Where needed and feasible, data may be delivered in encrypted format. A standard encryption system, such as SSL (secure Sockets Layer) or IPSec (IP security protocol), has been implemented. A system to manage distribution of public and private keys was set-up.

5.6 FUTURE

The design of the DSC is of a generic concept, allowing easy adaptation to future requirements and technological evolution. The HDU's requirements will also change to accommodate new applications of hydrological data and to increase the efficiency of the HIS. In this context, the choice of database allows adaptation and scaling to the ever-changing needs and technology.

A development, which will certainly affect the concept of the DSC, is the ongoing penetration of web technology. Eventually, any HDU may have access to sufficient communication bandwidth to use the Catalogue on-line. The DSC design permits further integration of web technology.

Authentication and access control is an integral part of the DSC. Efficient, standard data encryption is also included.

6 THE CATALOGUE

The Catalogue is the combination of meta-data and supporting software, e.g. search and selection tools, communication tools, updating tools and security control. The final product, after using the Catalogue, is a data request file with query instructions, which have to be processed by the DSC.

The HDU perceives the Catalogue as a web type interface, which guides him during his search for hydrological data pertaining to the HIS. The Catalogue makes use of context sensitive menus. The meta-data contains information on data availability only and no real data. Data visualisation by remote users is not supported for security and performance reasons. However, trusted clients, like the owner DPCs, get online access, which allows them to visualise the data in tabular and graphical format. Visiting HDUs with adequate authentication may be offered similar facilities in the owner DPC.

The product of the search process is a request for data stored in a computer file. That computer file, dubbed Data Request File (DRF), can be submitted to the custodian DSC of the required data. The DRF not only contains the data selection lists but also HDU identification data. To maintain sufficient security levels, the HDU identification may be encrypted. The DSC processes received DRFs and executes the queries on the databases. Data will only be made available to authenticated HDUs who have adequate authorisation to receive the data. Security sensitive data may be encrypted using a public and private key combination.

In this Section, the architecture and the processes of the Catalogue are described.

6.1 CATALOGUE ARCHITECTURE

Figure 6.1 shows the various components of the Catalogue in their context.

6.1.1 META-DATA

The meta-data comprises all the entities that are required to locate and select any data item or set of items existing in the supported databases. The Catalogue only contains data on data-availability, the hydrological data proper can only be accessed via the DSCs.

6.1.2 THESAURUS

A thesaurus containing all the searchable keys has been implemented.

6.1.3 MAP LAYERS

To support spatial search and selection and to visualise the spatial distribution of stations the Catalogue contains map / GIS layers. Supported categories are administrative boundaries, hydrogeological features and location data such as the positions and categories of data collection stations.

Maps layers are displayed with north up; north and east scales are identical. Map layer(s) can be printed showing the spatial distribution of the selected stations in the context of rivers, aquifers, population centres, etc.

6.1.4 WEB INTERFACE

A web interface provides the Catalogue functions and supports various modes of access, viz.:

- LAN / WAN
- Point to point communication via dial-up (STD), ISDN and leased line
- Virtual Private Network (VPN), e.g. through Internet.
- Offline implementation on CD-ROM, the request may be submitted via email, Internet, FTP, Internet, physical media.

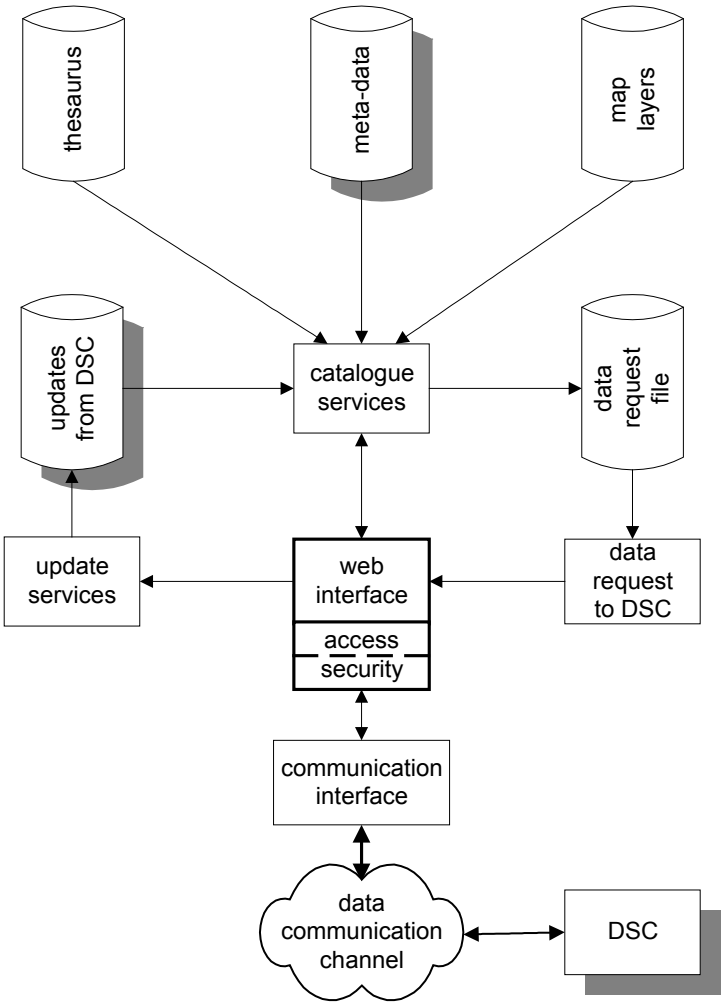


Figure 6.1: Catalogue data flow

All those communication channels are fully supported. Whatever the communication channel, when using the Catalogue services, the user experiences the same (identical) look and feel. The latest version of popular browsers like Internet Explorer and Netscape are supported. Internet Explorer is part of the Catalogue on CD-ROM.

The web interface has two ends, one at the DSC, the other with the user. The interface has been implemented with state of the technology and is easy to maintain. For maximum performance, in particular when using dial-up lines, the data exchange is kept to a necessary minimum, avoiding any redundant data transfer. In this context, the user interface is kept sober but effective.

The web interface based on common mark-up language standards like HTML, SGML or XML/XSL. For the implementations Java(script) was used. The Catalogue is supported for the PC environment, without forcing the HDU to use another OS than he is used to. Initially, this will be Windows98.

The web server is based on Win2000 advance server.

The qmail has been implemented under Linux

6.1.5 ACCESS/SECURITY

The DSC supports security/access control on all communication channels. A username / password combination is part of the security control system. For different user categories, different permission levels have been implemented. Levels of distinction are:

Hydrological Data Users	Data categories	Permissions
owner DPC	all data	Read, retrieve, write
local DPC	all data	Read and retrieve
Sister organisation SSW, SGW, CWC, CGWB, IMD	all authenticated data	Read and retrieve
Other HDUs	authenticated data	Read and retrieve

Table 6.1: Hydrological Data Users and related access permissions

Encryption may be implemented on sensitive data.

6.1.6 CATALOGUE SERVICES

The Catalogue services execute queries on the meta-data. The results are passed to the web interface. The final search results are put in a 'data request file' (DRF) The data request file contains all the information that the DSC needs to run the queries on the hydrological databases. The catalogue services support all meta-databases belonging to any number of DSCs. At start-up, it traces the available meta-databases and the area covered by each. The meta-databases may have overlapping service areas.

6.1.7 UPDATE SERVICES

During the data search process, the user gets the opportunity to request the DSC for an update of the meta-data. This is only necessary when the CD-ROM is used as meta-data source. The updates are stored on hard disk. The update is a complete set of meta data.

6.1.8 COMMUNICATION INTERFACE

The communication interface comprises the hardware and the communication protocols. The latter support error free data exchange. File signatures to allow change detection of the data accompany data on CD-ROM and other physical media. Changes can be the result of many causes, e.g. corruption of the media, tampering, replacement/removal of data etc.

6.1.9 DATA REQUEST FILE

Eventually, after selection of the required data, the catalogue services generate the DRF. This file has a format that can be interpreted by the DSC query tools.

6.2 SEARCH AND SELECTION

The concept of the search and selection process is depicted in next Sections. The process is almost entirely mouse-driven by point-and-click method. By simply clicking on the desired categories / items, the user can make selections, in compliance with the Windows standard. To support spatial entities and to give the user an overview of selected stations / objects, a number of simple GIS layers is included in the Catalogue services. Any GIS layer can be active or passive. Search and selection at data level is based on lists, which can be accessed by the mouse. Single and multiple items can be

activated or deactivated. The user may navigate through a tree structure each lower branch going into more detail. To make a selection, it is not required to go into the finest detail, unless the user wants other than the default settings for that branch level. The default settings are user adjustable. After finishing the search and selection, the DRF can be generated.

Apart from the point and click method, search and selection by running (complex) direct queries on the meta-database are be supported. This is particularly useful for experienced users. It also limits the amount of data transfer when directly accessing the catalogue services via the Internet. Not all the changes of the selected lists have to be communicated then, only the query results.

During the selection process new selections can be appended to or deleted from the present selection. A multilevel undo / redo option supports recovery from errors and / or unsuccessful search steps.

All options can also be selected from tables by point and pick selection, except for the rectangle option, which supports graphical entry of an area of interest. Each selection can be appended to the aggregate of previous selections or as opposed to that, removed.

Spatial selection

For most applications, the search and selection process would start by zooming in on the area of interest. However, the user is not limited in the combinations that he can make. It is possible to apply another selection order, e.g. first select all stations meeting certain requirements like all water level stations established in 1998 in India and then visualise their spatial distribution on a map.

Date established is not implemented.

Selection inversion, within a layer, is supported. Example: a few stations out of all in a basin are selected, subsequently the station selection for that basin is inverted to select all stations except the previously selected ones. Subsequently, areas may be switched off, to select the stations in the areas of interest only.

One or multiple station categories can be active simultaneously, e.g. water quality stations together with ground water observation wells.

Parameter selection

To complete the selection, criteria related to parameter values, e.g. water levels, and ranges thereof may be added to the selection. Since the Catalogue does not contain any hydrological data but meta-data only, the effect of the value/range criteria on the selection will only become known after running the search at the database(s) of the DSC.

The value criteria comprise a lower and an upper bound. The separation between the bounds may range from nil to wide apart. The default bounds for any entity are its minimum and maximum value as defined in the data directory.

Period selection

The period of interest is defined by starting date/time and ending date time. Multiple periods can be selected. Special periods like calendar year, hydrological year and season can also be supported. Further, it is possible to select sub-annual series, e.g. the months May and June for all years from 1955 to 1965.

The period of interest is selected by entering a start and an end date, all data available of the active stations / objects between those two data become selected. Entry of date is in the following ways:

- in an entry box, by typing the date according to the current date format
- in the same entry box but by scrolling year, month or day entry up or down
- graphically by stretching, shrinking or moving a time selection bar in a time bar graph. In this case, both start and stop date can be adapted. The active time selection is displayed on the selection window, default is all data, i.e. oldest up to present. The user may adjust the displayed window to period selected by the selection time bar.

The data availability for individual parameters of individual stations can be displayed in a bar graph showing time bars (horizontal) for each selected parameter. Not-selected parameters can be shown in half tone and activated by clicking on an associated box. The show/no show option can be selected in the view/options menu.

Also static and semi static data have a time label and can be selected for a specific time frame. If during that time frame data were changed, then more instances of that parameter will be retrieved from the database.

Combinations

Search and selection results of different Catalogue runs may be combined (and, or, not) to further focus on the required data or to extend the selection.

6.3 TECHNOLOGY

6.3.1 GRAPHICS (SELECTION / DISPLAY)

The GIS part of the Catalogue interface supports various GIS layers. Any selectable spatial objects in an active layer can be selected by mouse point and click. Complex selections are possible by selecting multiple objects in one or multiple layers. Layers may overlap partly or fully, e.g. a number of districts selected in combination with a geological object and a basin.

To inform the user, the GIS can display the spatial distribution of selected stations/objects in the context of other GIS layers. Stations can also be selected by point and click on the map or by encircling groups of them in a mouse drawn rectangle. A right-click on a station pops-up a list of meta-data pertaining to that station.

Only a limited number of layers and features is made available for search and selection. The most important are:

- State boundaries
- District boundaries
- Major population centres
- Organisation boundaries
- User made boundaries (e.g. project area polygon)
- Rail
- Roads
- Basin boundaries
- Major rivers
- Reservoirs
- Data collection stations (per category)

Each map layer can be active or not active

While running the GIS, the user may pan, zoom in and zoom out at wish on the displayed map. At start up a map of the service area is displayed, showing the states participating in the HP.

6.3.2 LIST BASED SELECTION

All search and selection can be fully based on lists, also the spatial selections, except selection by encircling groups of stations.

Each list has a top field, which selects all the entities in that list, or deselects all fields when any field was in selected state at the time of clicking. Selections can be inverted. The contents of lists can be sorted along various keys, e.g. alphanumeric and / or certain attributes, e.g. alphanumeric, District wise. Multiple undo/redo levels are supported. The items that are displayed in a list can be filtered by keyboard entry of text, mixed with wild cards. Consecutive selections in a list can be added to or subtracted from a larger set. This allows the preparation of very large lists in smaller steps.

The navigation within the lists and sub-lists has a tree structure. Selections are only valid for the active branch and all lower branches. Within lists, the standard windows list selection methods have been implemented (incl. shift, alt, and ctrl keys where applicable)

A thesaurus of search terms is maintained and made available to the HDU.

Response times of the search and selection process are short. For simple activities on Intranet and CD-ROM/PC response times should not exceed 5 seconds, complicated searches on the meta-data may take tens of seconds. Response times using the online version of the Catalogue should be reasonable and as short as possible. To achieve this, the Catalogue has been optimised for Internet use, i.e. limiting data exchange.

Data Type

Data types, are selected from point and click tables. The tables are sorted according to class and individual properties. Main classes of hydrological data are:

- Field Data
- Authenticated Data

Constituents to these classes are:

- Meteorological Data
- Surface Water Data
- SW Water Quality
- Ground Water Data
- GW Water Quality

All classes can be activated or de-activated in any combination. Each class has a subset of parameters, which also can be, individually or in any combination, activated or de-activated by the user.

6.4 DATA DISTRIBUTION

6.4.1 DISSEMINATION OF REQUESTED DATA

Retrieved data can be made available by data communication or on data media like diskette or CD-ROM. In certain cases, for security reasons it could be considered to have all data encrypted before it is transported by mail or data communication.

The choice between data communication and the use of physical media (physical transport) largely depends on the availability of communication facilities and the amount of data to be transported. If the HDU has access to a LAN connection to the Data Centre that LAN has a preference. For HDUs at more remote sites, modem based data transport might be considered. For limited amounts of data, say less than 1 Mbytes, a dial-up telephone connection could do the job. However, the effective speed of the line and the noise level can limit the effectiveness.

File transport - by FTP or e-mail - might not be feasible in case large amounts of data have to be transported or the data transport time will be too long. Then the data have to be recorded on a physical medium. Preferred media are Diskette and CD-ROM, as they are properly standardised and can be used on all PC's specified under HP.

Whatever data communication system is applied, it is fitted with a proper data transfer protocol supporting accurate error checking, and re-transmission facilities in case of detected errors.

6.4.2 EXCHANGE OF META-DATA

A separate activity is exchange of meta-data between the DSCs, this to keep the Catalogues up to date. There is no immediate need to keep the meta-data tightly synchronised, and as a result the update process can be rather simple. The meta-data can be exchanged on CD-ROM, e.g. annually. Updates can then be obtained from the custodians via file transfer, say on a weekly / monthly basis or at the time of access to the DSC, using physical media or any convenient data communication technology. The implementation of the DSC supports those modes of file exchange.

Preferably, the gathering of meta-data prior to preparation of a new Catalogue CD-ROM is organised by a specially appointed organisation.

6.5 INTERACTIONS AT DSC LEVEL

This overview addresses the main data categories and the communication channels used by the various users of the DSCs. The formulated concepts, requirements and specifications are to be integral parts of the DSC specifications.

6.5.1 INTRODUCTION

Each DSC is directly associated with and connected to its Owner DPC(s). The Owner DPCs belong to the same Agency or State as the DSC. Further it has a rather close relationship with the Local DPCs/Local DSCs. The local centres have a spatial aspect in common with each other; e.g. they operate within the same administrative boundary or basin boundary. Also with the other DSCs there is a relationship, i.e. all DSCs co-operate with each other by sharing the same software, exchanging Meta-Data, taking part in the delivery of data to each other's clients and by participating in the same system development Task Group.

The DSC is a safe storage place for data pertaining to its Owner DPC(s). The DSC has four main repositories containing Field Data, Authenticated Data, Temporary Data and Objects. The first three repositories contain Hydrological Data. The most recent years of this data are kept on-line, the remainder of the historical data is kept offline but can be instantly loaded.

6.5.2 CATALOGUE AND META-DATA

The availability of any data item is published in a Catalogue. Duplicates of data, wherever they reside, are not published in the Catalogue. The Catalogue comprises search and selection software, Internet tools and the Meta-Data. The DSC generates Meta-data for the data in its custody. The Meta-data contains all information required to retrieve any individual data element or assemblies of data like time series.

The Meta-Data of different DSCs are kept in separate files, a new DSC is added to the Catalogue by addition of the associated Meta-Data file. The Catalogue detects which Meta-Data files are available and makes the Meta-data accessible for the search and selection tools in an integrated manner. The Catalogue user does not notice that more than one Meta-Data files are processed concurrently. If in future more DSCs are established, then no change to the Catalogue is needed, and the simple addition of the Meta-Data files is sufficient. This approach makes the system flexible to changes. For the Field Data separate Meta-Data files are used, the Field Data is only visible to the Owner DPC and Local DPCs and does not need to be delivered to other DSCs or the HDUs. It should be noted that the entire contents (Hydrological Data and the Objects) of the DSC, both online and offline, feature in the online Meta-Data and can be searched and selected.

The Catalogue does not contain confidential information; after all it only contains data about data availability and not the actual data. Therefore the catalogue data is made freely available to anyone. HDUs may obtain / receive catalogue data from individual DSCs. The Catalogue contains tools to print all its contents or a user selected subset thereof in a formatted way. The subset could be basin wise, state wise, project wise etc. The printout allows the HDU to use the Catalogue on paper. One of the supported printouts is a formatted and user readable overview of the contents of the DRF.

Security – contains LAT / LONG, under discussion between CWC and MoD/SoI

6.5.3 DELIVERY OF FIELD AND AUTHENTICATED DATA AND OBJECTS TO THE DSC

Owner DPCs deliver data to the associated DSC in a formal way. The DSC verifies the data integrity, consistency and formats against explicit guidelines. The Owner DPC can deliver any of the supported data types. Upon acceptance, after checking the data integrity by the DSC, the data are loaded in the appropriate databases and the existence of the same is reflected in the Meta-Data.

Stringent scrutiny.

6.5.4 TEMPORARY DATA

The Temporary Data constitute a special class in the sense that after use they are discarded. A DPC may request Local DPCs to deliver Field Data and/or Authenticated Data to enhance the hydrological data processing. The Authenticated Data and the Field Data are published in the Catalogue. However, the Field Data is only visibly to Owner and Local DPCs. All Temporary Data belonging to the Local DPCs is used for data processing only and discarded thereafter, in compliance with the philosophy that only the Owners keep data in official custody. Meta-Data is also generated for the Temporary Data, as with any data class. The standard Catalogue functions can be used to execute search and selection on the Meta-Data associated with Temporary Data. In this way the same Catalogue tools approach all data repositories.

6.5.5 CENTRAL WEB SITE

A Central Web Site, e.g. at www.india-water.com, has been established to act as a central node for the HIS. Prospective HDUs may execute the search and selection process on the Catalogue at the Central Web Site whereupon that server distributes the DRF to the respective DSCs. The DSCs deliver the requested hydrological data to the HDUs, provided that the HDUs are properly authenticated and are entitled to receive the requested data. The delivery may have some throughput time due to manual security checks. It may occur that only part of the requested data is sanctioned for distribution.

Actually, the Central Web Site has the same functionality as the DSC Web Servers of the individual agencies; however, the Central Web Site will not have hydrological data available. Its main services are the centralised collection and maintenance of the catalogue data and provision of a convenient access point for the HDUs. The Central Web Site also acts as a portal to the Web Servers of the agencies.

The Central Web Site also acts as an HIS resource centre. The home page has a link to this resource centre, which provides users with information on software updates and software upgrade downloads. It also allows viewing and downloading of standard HIS documents including HIS manuals. A simple query and search engine coupled with a simple tree structure is created for this purpose.

6.5.6 META-DATA EXCHANGE

Each DSC keeps its Meta-Data available on its FTP site. The Central Web Site regularly collects the Meta-Data from all DSC FTP sites and make the same available in the Central Catalogue. At intervals of about a year a CD-ROM version of the entire Catalogue might be issued. Other DSCs and also the HDUs may collect the Meta-Data files from any of the DSCs' FTP sites or directly from the Central Web Site. It should be noted that any Catalogue can support any number of Meta-Data files, hence, any number of DSCs.

6.5.7 OWNER AND LOCAL DPCS

Owner DPCs have direct access to the Catalogue via LAN, the DPC's own Meta-Data may be put in a special position for maximum performance. Any selected data is retrieved and made available online without delay, no intermediate action and personal authentication by the DBA is required since the Owner DPC requests its own data. If a DPC requests data from another DSC then authentication procedures have to be executed as implemented on the addressed DSC. Local and other DPCs have access to the Catalogue in a similar way as the HDUs.

Retrieval of Temporary Data is executed via the Catalogue. The Temporary Data are only visible to the Owner DPC (i.e. the DPC that borrowed the data and loaded the same in the Temporary Data Repository) and also to the Local DPC to which the data belongs. Or in other words for each DPC it is possible to see in the Catalogue where its data resides in Temporary Data.

6.5.8 HDU ACCESS

To HDUs, only Authenticated Data and published Objects are available, hence visible in the Catalogue. For HDUs a range of communication channels is available to access the Catalogue, viz.:

1. in any DSC via LAN
2. from any place via Internet on a individual DSC Web Server
3. from any place via Internet on the Central Web Site
4. offline on CD-ROM

The Catalogue generates DSC specific DRFs. The HDU may submit the DRFs to the specific DSCs or all of the DRFs to the Central Web Site for dispatch to the respective DSCs. DRFs are only run on the databases when the prospective HDU has submitted proper authentication and has passed all access control procedures. DRF(s) may be delivered to a DSC via e-mail, FTP or on physical media, e.g. diskette or CD-ROM. One of the supported printouts is a user readable formatted version of the contents of the DRF. For the HDU this acts as a report of what data was selected. The DSC uses the printout in its authorisation procedures.

Catalogue access	DRF submission modes			
	LAN	local WEB	central WEB	physical media
LAN	✓			
individual DSC WEB		✓		
central WEB			✓	
CD-ROM		✓	✓	✓

Table 6.2: DRF delivery to DSC in relation to mode of Catalogue access

Catalogue access	data transport modes			
	LAN	e-mail	FTP	physical media
LAN	✓	✓	✓	✓
individual DSC WEB		✓	✓	✓
central WEB		✓	✓	✓
CD-ROM		✓	✓	✓

Table 6.3: Data delivery in relation to mode of Catalogue access

The requesting HDU will only receive data after authentication and only data that he is entitled to. Data delivery is encrypted, unless the DSCs policy permits otherwise. The implementation of authentication and the levels thereof are the prerogative of the Owner DPCs; the implementation is by the DSCs. The same applies to encryption, if any is required.

6.5.9 CONNECTIVITY

As mentioned above, it is envisaged, that the DSCs under the Hydrology Project will exchange data with data providers and data users by various data communication technologies. Further, it is anticipated that the development of Web technology and the communication infrastructure will proceed with its present rapid pace.

The data providers, which are primarily the DPCs, have a direct connection with the associated DSC. These DPCs and the DSC are connected by Intranet technology over LAN. In a few cases, the distance between related Data Centres may be too large for a normal UTP LAN, in such cases a dial-up ISDN connection (64 / 128 Kbps), ADSL, leased line, RF connection or similar, whichever is most appropriate under the local conditions, would be required.

Most of the HDUs have no direct connection to the DSC, for them dial-up access based on Internet technology is most appropriate.

The search and selection functions of the Catalogue are primarily based on Web technology. Each DSC should offer access via the LAN, dial-up, Internet and NICNET. To support the Web based access, the Catalogue functions have been implemented on a Web server with appropriate software. The Web server is to be equipped with software interfaces and various communication hardware devices to support the specified functionality and the diverse communication channels.

Some Agencies may opt to link DSCs in a higher level network or a centralised WAN. In such cases the Agencies would have to obtain additional hardware and communication services to implement this

option. The Catalogue functions offer sufficient functionality to implement centralised search and selection like through the Central Web Site. However, it should be noted that the DPCs remain the owner and custodian of the data. Under the Hydrology Project it is not planned to mirror data at centralised DSCs.

The Catalogue functions have the capability to handle catalogue data from multiple DSCs, i.e. from a single DSC to all associated DSCs. Actually, the Catalogue makes use of all the catalogue data that it has access to. For a local HDU the catalogue data of the local DSC would suffice, however, for multilateral or national projects catalogue data of several DSCs would be needed. The Catalogue does not contain confidential information, after all it only contains data about data availability and not the actual data. Therefore the catalogue data is made freely available to anyone. HDUs may obtain / receive catalogue data from individual DSCs.

6.5.10 SPECIFICATIONS FOR DSC WEB SERVER

The Web Server is a major interface between the HDU and the DSC; the Catalogue functions are implemented on the Web Server. The Authentication, Authorisation and Accounting functions related to HDU access are also implemented in the Web Server. The Accounting function is in particular useful for Agencies that want to implement a billing system to charge for the delivered data. Further, the Web Server handles user requests for data, based on the DRF as generated by the Catalogue at the Web Server or the stand-alone Catalogue (on CD-ROM) with the HDU. The same server caters for data delivery, possibly in encrypted format. Obviously, the Web Server has been equipped with a firewall and a virus protection package. The Web Server is implemented on a standard PC; the operating system is Linux.

The Database Server resides on a separate hardware system. Hence, the Catalogue functions on the Web Server require only a limited amount of storage capacity. This makes it feasible to establish the Web Server at a convenient place, which is not necessarily at the DSC. One option is to make use of a Web Service Provider. In any case the Web Server should have access to an Internet Node.

The Web server implementation supports the following functions and features:

Accounting package with Web Interface containing:

- Monitor Who's Online, Errors, Account Summary, Invoices and others
- HDU Service include Account Summary, View Invoice, Change Password
- Invoices ?
- Invoice not sent to web. Proforma invoice available
- Monitor all Services; e.g. DNS, FTP, WWW, SMTP, POP3
- Logging of accounting data in a scaleable database with ODBC and SQL support

Access features

- Support of IP allocation and routing
- Refuse access for non-payment, time-of-day, hours used or MBytes delivered
- DRF will be made

The accounting and billing function has the following features:

HDU Accounts

- Account reporting
- Proforma invoice, receipt and invoice.
- HDU email addresses
- HDU information instantly accessible
- Problem reports; generic or by type
- Send problem resolution by email to HDU

Pricing for delivered data

- Pricing is by station by year for kind of data, by record and by station by year by series

Access Limits

- Time-of-day
- Hours/month
- Megabytes delivered

Invoicing Features

- Invoice preparation / report
- Send Invoices via Post or Email
- Invoice includes number of sessions and hours used plus charges for delivered data

The invoicing should reference to access events the quantities of delivered data to allow the HDU to verify the invoice. The user always pays in advance. Therefore this is not applicable and is not being captured.

Management Reporting

- Web Server log report on demand.
- Account ageing analysis (detection of unused accounts)
- Payment history and money flow
- All advances
- Subscriber usage statistics.
- Usage stats only by IP address and not by user.

Resource Centre

- New software releases and features
- Upgrade downloads
- Search engine for HIS documents and manuals
- Download documents and manuals

General features

- Contact us
- New user registration request
- Acceptance of software upgrades and letters
- Modification of user details

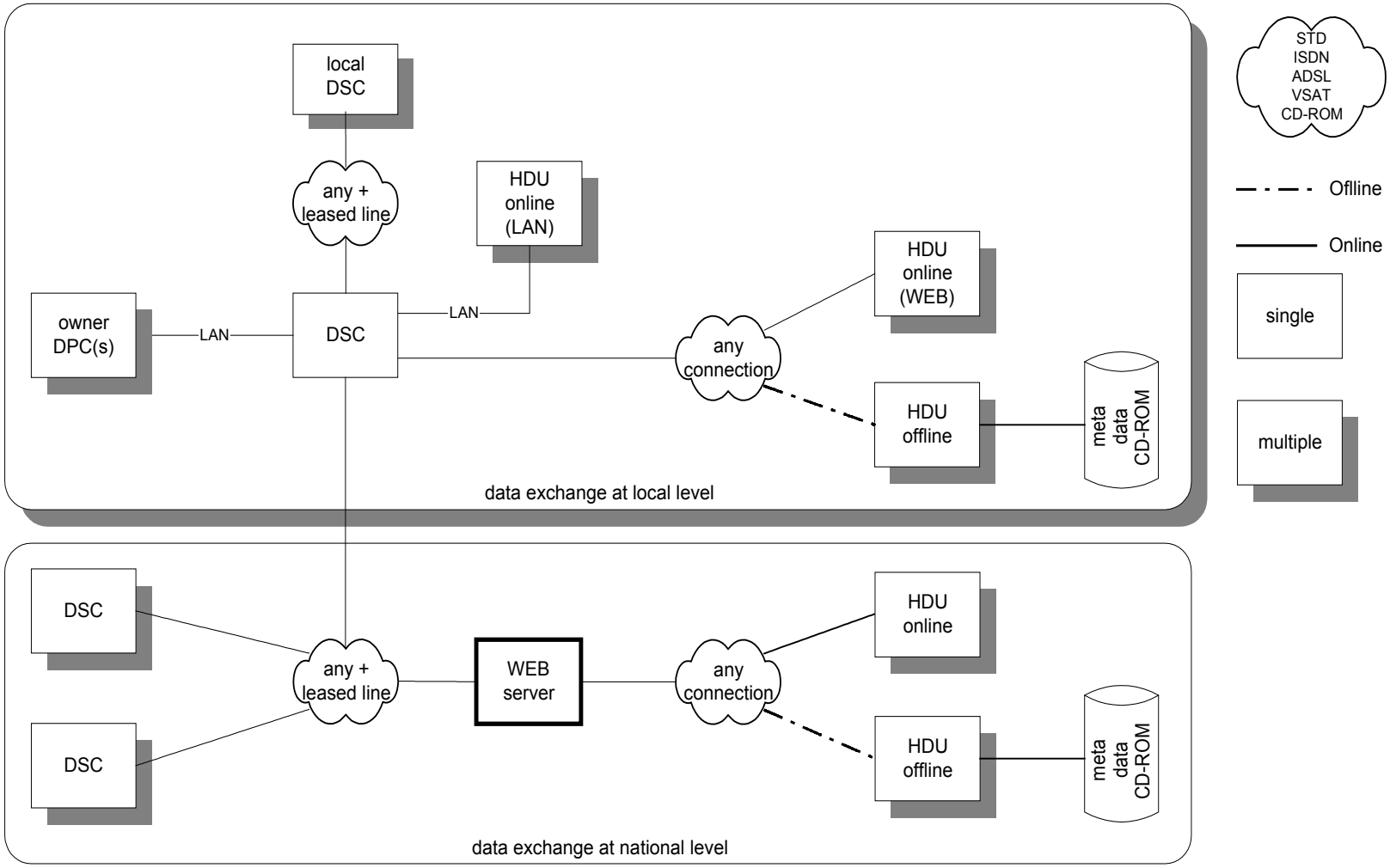


Figure 6.2: Data exchange overview

7 DSC SOFTWARE SPECIFICATIONS

Following are specifications for the HIS database software system to be used in the DSC. Most important in this application are the use of a standard DBMS and related tools, the high quality, and the uniformity. These qualities enable the sharing of data produced by the different agencies, and facilitate a stable and profound working environment.

The specifications are grouped according to specific subject.

7.1 GENERAL FEATURES

This section describes the general properties of the software, the working environment, method of work, hardware and software, systems and languages needed for the database.

- **Hardware Platforms** - The hardware platform is based on standard PC technology.
- **Working Environment** - The software system of the DSC is based the Windows-NT environment. The working environment of the HDU will be Windows-98 or later.
- **Client/Server** - The application system is based upon Client/Server architecture so as to enable intelligent and most cost/effective utilisation of hardware and software tools.
- **Drivers for Peripherals** - Standard drivers for peripherals such as: printers and data storage media recorders have been integrated in the Software System.
- **Addition of User Modules** - The Software System supports the addition of user (in this context DSC staff) developed modules (in particular for import and export of data). The modules are made accessible by adding related menu items. Source code of the menu system is available.
- **Flexible and Modular system -** the software is built of modules which are easy to adapt to changing working environment (OS update), flexible to changes in user requirements, new data types / objects, changing communication services, etc.
- **User Interface** - The Database System has an intuitive user interface, which is based on GUI presentation standards for Windows. The user interface and presentation are compliant with industry standards and capable of being web-enabled wherever relevant, especially in all areas pertaining to data dissemination.
- **Screen Presentation -** The software screen menus are easy to work with, self explanatory and of uniform format
- **Ease of Use** - the system is easy to learn, easy to customise and adjustable to changing needs.
- **Open and Portable** (**database and hardware platform independent) -** The system software is portable between different hardware and software platforms and databases. The DBMS is ODBC compliant, however, the BLOBs cannot be searched, they are manipulated as objects.
- **Scalable** – The system is scalable to adapt to growing storage and workloads.
- **Reliability, Availability and Integrity -** The software is expected to be reliable and to perform smoothly (no “hang-up”) .
- **Performance** - Satisfactory and adequate response time for users, at low resources use - for batch and on-line processing.

7.2 DATABASE FEATURES AND TOOLS

The database and tools (SQL Server) a universally accepted industry standard. The same applies to the development tools and programming languages.

- **Database software** - the Database software of the DSC, including the CD-ROM version of the Catalogue, is based on SQL Server database management software and tools. The Catalogue is be based on MS-Access.

- **Development tools -** The Application development tools are based on standard tools and languages of the leading vendors in the market and industry standard such as: C, C++, Visual Basic, Java or other much used and well established programming languages. 4GL tools.
- **Database queries, report generation and presentation tools:**
  - Standard report generation tools
  - Standard SQL query generator tools
  - Standard Graphic engine
  - Generation and editing tools for the data tables and structures
- **Data integrity** - The Software System includes an inherent data integrity checking/monitoring function, to avoid data corruption and losses
- **Database spatial feature** - the database has a feature to accommodate any data object
- **Administration and Management tools**
  - database backup and recovery
  - creating and managing database tables
  - system configuration and administration
  - creating and managing users' authority tables.

7.3 DATABASE MANAGEMENT AND ADMINISTRATION FUNCTIONS

In order to maintain system stability, it is of great importance to have adequate and easy to use service and management tools for the DSC staff.

- **Extracting data** - function to extract any of the stored data from tables. Next to standard data types, also special structures like time series and map objects are supported efficiently.
- **Database service menus** - the software supports on-line and batch service menus and procedures for the use of the database administrator. These menus have been customised (and flexible to changes) to implement the specific needs of the DSC. One of the batch services is data extraction based on the Data Request File, which is generated as the result of catalogue search and selection.
- **On-line services** - the software supports on-line services menus for authorised user's, like owner DPCs
- **Loading data** – the software supports loading of data in the database. Part of the loading is a rigorous integrity checking facility.
- **Periodic production of database meta-data** – tools to generate the DSC meta-data is part of the software. At regular intervals, a snapshot of the DSC meta-data is set apart for inclusion in the Catalogue. Subsequent changes to the meta-data are set apart for distribution via file transfer to any HDU that requests to update his Catalogue.
- **Activity log** - activity log function which includes the information about users login / logout date and time, data import / export / removal, Catalogue update, user requests, response time on user requests and similar performance indicators.
- **Users accounting** - accounting function module for future commercial purposes (such as: private users billing), enabling data dissemination and recording of data retrieved volume.omplete accounting information does not need to be tracked, especially connection time etc.

7.4 FILE IMPORT AND EXPORT

Most of the data exchange with DPCs and HDUs is through file transfer. Hence, file handling is an important function of the DSC. The file handling software is designed to be effective. All transfers support forward error control or similar technology. The DSC supports all file types associated with HIS for import and export. Therefore, the DPCs may deliver their data in a convenient format, most

likely the DPC's native format. When requested, the DSC delivers the data in the same format again. To accommodate this, the DSC has proper data conversion tools in place.

Thorough integrity control is implemented on all data I/O.

7.5 PRE-DESIGNED REPORTS

Following are the database reports to be included in and supported by the software. All the reports produced by the software are flexible to changes (addition or adaptations) and comprehensive.

- **Catalogue contents report** - Catalogue contents report are produced to user request
- **Periodic database reports** - monthly/annually reports are produced and distributed. These reports include DSC contents
- **Annual reports** - annually the DSC will assemble reports for management use. The annual report includes:
  - database contents
  - last year updates information
  - aggregated information for groundwater data
  - aggregated information for surface water data
  - aggregated information for meteorological data
  - aggregated information for water quality data

Note that the owner DPCs produces the hydrological yearbooks

- **Requests for Information - report** - a report of user requests for information is produced daily (night batch processing) and monthly
- **Activity log report** - activity log report is produced periodically it includes the information about users login/ logout date and time, data updating, deleting of data, user requests for data inter-states and intra-state

7.6 WEB SERVER / ISP TOOLS

Each DSC operates and maintains a web server. For that proper hardware and software has been implemented. Maintenance and development tools are also be part of the delivery. The web server has been set-up to support the Catalogue services and the data exchange with the DPCs, DSCs and the HDUs. Provisions have been made to implement a data pricing system supporting cost differentiation based on data type, amounts, HDU and contracts.

7.7 USER INTERFACE TOOLS

The DSC has powerful and state of the art 'world wide web' technology built-in, which allows the system to provide Intranet and Internet services to the DPCs and the HDUs. The web technology allows the development and support of sophisticated user interfaces to which the Catalogue belongs. These user interfaces support a combination of GIS and database technology, data communication technology. The user interfaces are used off-line and on-line, in both cases they have the same look-and-feel. Both the Netscape and Internet Explorer browsers are supported. The user's end of the security system is linked to the web interface.

7.8 SECURITY COMPONENTS

The software and hardware based security systems, prevent data corruption, hacking, illegal access (read / write) and loading of the services. Users should be properly registered before they may receive any data. User access to the DSC's systems is limited, depending to their authentication.

Access control: Standard access control based on username and password. On data supply stringent multilevel authentication procedures are implemented.

Encryption: PGP or similar where needed on external data supply.

Firewall: A further protection based on firewall technology has been implemented on the DSC. The DSC has been segmented to optimise security.

Anti-virus: A standard virus protection system (McAfee, Norton, or Solomon) has been implemented, on the network, servers, clients and other computers. An Internet based update service is included. A pre-paid service contract is part of the delivery.

8 HARDWARE SPECIFICATIONS

Following are specifications for the hardware at the DSC to be able run the database system and to execute the tasks assigned to the DSC. In Table 8.1 an overview is presented of hardware items with quantities as are available for running an average DSC.

Item	Number
Database server	1
Web Access computer	1
Catalogue access PC	1
Staff PC	2
Smart UPS	1
UPS	2
Laser printer	1
Inkjet printer	1
CD ROM/DVD reader	1
CD-R / CD R/W recorder	2
DAT tape recorder	2
Dial-up modem	2
ISDN modem (Optional)	1*
Leased line modem (Optional)	1*
LAN / WAN	1

*: optional, only if needed

Table 8.1: Hardware quantities for an average DSC.

8.1 COMPUTER

8.1.1 DATABASE SERVER

The database server conforms to normal server standards defined by servers of manufacturers like Compaq, IBM and HP. It supports two processors (Dual Intel Pentium III ≥ 500MHz) and Integrated-wide Ultra SCSI 3 controller. Main memory of 256MB expandable to 1GB, cache memory in excess of 512KB and bus speed of ≥100MHz. Two hard disks of 18GB (U2W SCSI : transfer rate > 80MB/s) each and CD-ROM / DVD with CD access speed of ≥ 32X and DVD ≥ 6X have been configured. The

server has two network cards – 10/100 Mbps autosensing with 10 Base T compatibility. Three PCI slots are free and available. One enhanced parallel (ECC/EPP) bi-directional, 2 fast serial, 1 USB and 1 PS/2 bus mouse port are available along with one 1.44MB floppy drive and Microsoft / Logitech bus mouse with mouse pad. Additional 50 pin (or compatible) SCSI external port to support the RAID system described in 10.8 below and DAT drive as specified in section 10.6. Keyboard compatible with Operating system is available. Video Controller with 2MB VRAM and 17" colour monitor with 1600x1200 resolution; 0.26mm or less dot-pitch.

The server supports features like remote wakeup, remote security, remote shutdown, remote systems management, error logging etc. It supports high performance fault tolerant with Wake on LAN facility.

The Operating System is Windows 2000 Advance Server (latest release) with ≥10 user license conforming to International standards with utilities and extensions Windows NT (latest release) can also be used. Virus protection software and Microsoft Office are included with the server.

The RDBMS supported is SQL Server 2000 with 5 user licence; along with standard drivers (including ODBC) and support programs are available.

8.1.2 WEB ACCESS COMPUTER

The Operating System is Windows NT (latest release). WebServer software (like IIS - Internet Information Server - available with Option Pack 4) has been installed for enabling web-hosting. A transaction server and commerce server is available to provide other functionality. Front Page or equivalent web publishing / development software is also provided, along with development tools like Dream Weaver, Fireworks, Cold Fusion etc.

8.1.3 PC FOR LOCAL CATALOGUE ACCESS

The PC for local catalogue access is based on the Intel Pentium III chip with a minimum clock speed of 500Mhz. The main memory ≥128MB with provision for expansion to 384MB and with a bus speed in excess of 100MHz. L2 cache memory is ≥512KB and >=256KB for Coppermine CPUs. Hard disk capacity is ≥10GB. The CD-ROM / DVD supports CD reading at 48 X or more and DVD at 6X or more. The system has to be network ready with standard ports, floppy drive and mouse. Two PCI slots are free and available. The video controller (AGP) has ≥8MB VRAM and the monitor is 17" with 1600x1200 pixels colour resolution and less than 0.26mm dot pitch.

The Operating System is either Windows 98 or Windows NT for workstation (latest release) available on original sealed media upgradable from Microsoft. Compatible virus protection software (in case network virus protection software is not available on the server) and MS-Office Professional to be included with the system.

8.1.4 PC'S FOR DSC OPERATION

The PCs for Data Storage Centre operation are based on the Intel Pentium III chip with a minimum clock speed of 500Mhz. The main memory is ≥128MB with provision for expansion to 384MB and with a bus speed in excess of 100MHz. L2 cache memory is ≥512KB and >=256KB for Coppermine CPUs.Hard disk capacity is ≥10GB. The CD_ROM supports CD reading at 48 X or more and DVD at 6X or more. The system has to be network ready with standard ports, floppy drive and mouse. Two PCI slots are free and available and the keyboard is standard. The video controller (AGP) has ≥8MB VRAM and the monitor is 17" with 1600x1200 pixels colour resolution and less than 0.26mm dot pitch.

The Operating System could either be Windows 98 or Windows NT for workstation available on original sealed media upgradable from Microsoft. Compatible virus protection software (in case

network virus protection software is not available on the server) and MS-Office Professional to be included with the system.

8.2 UNINTERRUPTIBLE POWER SUPPLY

The Data Storage Centre has centralised UPS systems for all computer equipment and peripherals but the database server and the laser printer. However, the database server has its own smart UPS which is capable of scheduling the functioning of the server. The web server and some peripherals are connected to one UPS; while the balance systems are supported by the larger centralised UPS.

8.2.1 3KVA UPS & 1KVA UPS

The 5 kVA, 3kVA and 1kVA line interactive UPSs were configured with maintenance free batteries with ≥120 minutes backup time on 50% load. Efficiency on battery operation is better than 90%. The backup time is upgradable to 4 hrs with maintenance free batteries with the same charger. The output waveform is pure-sine wave with a Total Harmonic Distortion of less than 10% on mains and on battery. The intelligent UPS can be interfaced with a computer for warning in case a shutdown in imminent. The UPS supports other features like fast correction for load changes, surge protection, low noise operation, trickle and boost charge for batteries, protection against battery reverse polarity, battery over-charge protection, battery discharge protection and protection against short circuit, over and under voltage. Battery recharge time to 90% of its capacity after discharge is less than four hours. Audible noise at 1 metre distance is less than 55 dB (A).

8.3 PRINTER

8.3.1 LASER PRINTER

The B/W laser printer supports a minimum resolution of 600x600 dpi and a printing speed in excess of 12PPM. 8MB Memory is expandable to 16MB. The printer language is MS-Windows compatible and it supports enhanced PCL5 / PCL6 and has Post Script support. A minimum of 20 scaleable fonts, Parallel interface and Ethernet connectivity (10/100 Base-TX card in 1 open EIO slot) are standard.

The 2 input trays have ≥200 sheets capacity. The printer is capable of printing on various media like plain paper, envelopes and transparencies. Paper sizes supported include A4, executive and legal.

An additional low-noise CVT with adequate capacity for peak load is also provided with the printer. This allows usage of the printer without installing it on the UPS.

8.3.2 INKJET PRINTER

The Colour Inkjet Printer supports a minimum resolution of 600x600 dpi (B/W) and 300x300dpi (Colour); a printing speed in excess of 4PPM (B/W) and 2ppm (Colour). 8MB Memory. The printer language is MS-Windows compatible and supports enhanced PCL5 and Post Script. A minimum of 20 scaleable fonts and Parallel interface are standard.

The tray has ≥200 sheets capacity. The printer is capable of printing on various media like plain paper, envelopes and transparencies. Paper sizes supported include A4, executive and legal.

8.4 CD-R/W

The CD-Writer has a speed of ≥ 4X in Write once mode, ≥ 2X in rewrite mode and ≥ 20X in Read Mode. It supports both CD-R and CD-RW disks. In some cases the CD-RW is of external type with independent power supply and casing; the Interface is standard parallel or USB with full software and connection cable.

The interface cables and software are totally aligned to the CD-Writers' specifications and configuration so as to allow for reliable and interruption free recording of data.

8.5 DAT / DLT TAPE DRIVE

Internal DAT drive with support for DDS-1, DDS-2, DDS-3 and DDS-4 cartridges. Data capacity of 20GB native (40GB compressed). Sustained transfer rate in excess of 3MB / sec in native (uncompressed) mode. Mean file search time of less than 50 sec on DDS-4. Recording format to support ANSI / ISO / ECMA DDS, DDS-DC, DDS-2, DDS-3 and DDS-4. MTBF is in excess of 200,000 Hours on 12% duty cycle. SCSI interface.

Standard accessories like interfacing and power cables, cartridges and effective software for file selection and to control the data recording on tape.

Alternatively, higher end DLT drives can be configured in place of the DAT drive with similar or better specifications.

8.6 DATA COMMUNICATION

8.6.1 ANALOGUE MODEM

External modem 56kbps v.90 modem with connectors to telephone line. The modem is capable to operate both in pulse and tone mode.

8.6.2 ISDN MODEM (SINGLE AND DOUBLE CHANNEL 64 AND 128 KB/S)

Single channel speed of 64kbps and double channel speed up to 128 kbps is supported.

8.6.3 LEASED LINE MODEM

The modem specifications are in line with the ISDN modem specifications and the capacity and throughput of leased line to be made available at the centre.

8.6.4 LAN

The switch is an Ethernet Network 100BT/UTP with 16 ports. The switch is intelligent and is capable of stacking for upgradation at a later date by adding additional switchs. Plug and play capabilities are supported.

The cables are routed from point to point using conduits of standard sizes and specifications. The cable routing is such that a minimum distance is maintained from power lines and other electromagnetic radiating devices. Wherever the cables are to cross floors, suitable ramps and

protective metal conduits have been provided over normal protection so that moisture ingress and damage to the cable is prevented.

The Cables are AT&T Cat V or equivalent.

An elementary level network management software is included and the network is capable of expanding to accommodate future growth (NICNET, Internet processing enhancement, new nodes, enhanced connectivity).

8.7 RAID DISCS

The database server supports RAID level 5. Provision is made for up to 5 hot swap Ultra 2 wide SCSI drives (supporting a mixture of low-profile and half-height drives), plus two common tray bays. For the controller Net RAID is implemented.

8.8 HARD DISK

External SCSI hot-swap hard disk of ≥17GB compatible with the RAID System and the SCSI Interface available in the database server. The hard disk is of Ultra2-SCSI WIDE type supporting data transfer rates in excess of 160MB/s.

8.9 ARCHIVING BOXES AND CABINETS

Boxes, cabinets and racks are available for storage of hard copy outputs, magnetic media (like floppies and tapes) and CDs. These guard against dust, moisture and are lockable for security reasons.

9 RELATED ASPECTS

In this Section, related aspects are accessed. Documentation topics are covered by paragraphs on Documentation and Manuals. Support topics are covered by the paragraphs on Service and maintenance, support, training and warranty. The last paragraphs address the demonstration of a working demonstration version and the final implementation of the entire system at all the DSCs.

9.1 SOFTWARE SYSTEM'S DOCUMENTATION

The software system was supplied with all the documents and manuals, required to operate, maintain and configure the system at the user's specific environment. The documents include:

- Software system general description and diagram
- List of modules included, and a short description for each module
- User's Guide - containing all the screen and menus, explaining each function and the use of it. The user's guide is in hypertext with printer support.
- Error handling - message list and error handling for each message
- System Maintenance Guide - containing all the information required to operate and maintain the system by the data Centre staff
- On line Help is to be supplied as integral part of the software system, to assist the users during operations
- A FAQ list is available to the Catalogue users
- A database maintenance guide.

9.2 MANUALS

The software system includes all the specific manuals, required to effectively use the DBMS and tools, which are part of the system. These include:

- Software Manual and User Guides for the databases and tools used by the software system
- Software Manual and User Guide to the GIS tools used by the software
  - Software Manual and User Guide for development tools and programming languages which are required to operate the system
  - Messages and Codes and error handling manuals for all above software tools.

9.3 SERVICE, SUPPORT AND MAINTENANCE

Software House and System Support

As was stated before, the uniformity of system is of most importance. Thus the software system should be centrally maintained. The supporting firm should be capable to support and maintain the application software in all participating states and agencies.

- **Local support** - the system software is supported by an efficient, locally based and well established Indian company (vendor may not be direct supplier). The system is designed to be **sustainable** in the **Indian** environment, both hardware and software wise.
- **Software House** - the software supplier M/s Rolta.

The M/S Rolta supplies service and support for all the deliverables. The support and service for the software includes:

- Installation and customisation of the system (all items supplied, or agreed on) in all the participating states
- Software support by phone and other off-line support
- Software support in-house when needed, at the DSCs
- Installation and customisation of new versions of database and software tools (as supplied, or agreed on)
- Development of new features and installation of new software versions in the DSCs
- Adaptation of the software for new versions of software tools and hardware platforms when required
- Advising the users about working environment and performance issues
- Removal of all ‘bugs’ and glitches in the software as may surface from time to time
- Designing, implementation and support on the mechanism of replication, synchronisation and publication of the catalogue.