FINAL

RFC Archive Database/Files System Requirements

August 29, 2001

" The hardest single part of building a software system is deciding precisely what to build. No other part of the conceptual work is as difficult as establishing the detailed technical requirements, including all the interfaces to people, to machines, and to other software systems. No other part of the work so cripples the resulting system if done wrong. No other part is more difficult to rectify later."

Frederick P. Brooks, Jr
"father of the IBM System/360 computer family"

Table Of Contents

Purpose

The purpose of the this document is to define requirements for the RFC Archive Database/Files System. This archival will enable the River Forecast Centers to do the following:

- Verification
- Studies to improve future products
- Provide an easier pathway to use data between the operational environment and the calibration environment (e.g. using OFS time series data in ICP)
- Channel routing development
- Unitgraph development
- Case studies
- Operational forecast assistance, such as displaying past events for comparison to current event and comparison of observed data to current forecast and “x” number of previously issued forecast for current event
- Applied Research
- Respond to Customer Inquiries

Classification of Requirements

The RFC Archive Database/Files System Project is currently using the following requirements categories that were originally developed by the European Space Agency :

1.0 Functional Requirements

Functional requirements are used to consider system behavior, redundancy, human aspects and trade-offs between issues, weighing the benefits of each.

Note: For the Archive Database/Files System, Functional Requirements are divided into 2 subcategories:

1.1 Structural - description of data structures to archive
1.2 Dynamic - description of dynamic manipulations of the archival system

2.0 Performance Requirements

All performance requirements must have a value which is measurable and quantitative, not a value which is perceptive. Performance requirements are stated in measurable values, such as rate, frequency, speeds and levels. The performance values are based either on values extracted from the system specification, or on an estimated value.

Note: For the Archive Database/Files System, Performance Requirements are currently focused on Query performance.

3.0 Interface Requirements

Interface requirements are handled separately, with hardware requirements being derived separately from the software requirements. Software interfaces include dealing with an existing software system, or any interface standard that has been requested. Hardware requirements, unlike software give room for trade-offs if they are not fully defined, however all assumptions should be defined and carefully documented.

Note: For the Archive Database/Files System, Interface Requirements are focused on interaction with existing file systems, RDBMS, and office automation tools.

4.0 Resource Requirements

Resource requirements divulge the design constraints relating to the utilization of the system hardware. Software restrictions may be placed on only using specific, certified, standard compilers and databases. Hardware restrictions include amount, percentage or mean use of the available memory and the amount of memory available.

Note: For the Archive Database/Files System, Resource Requirements consider use of COTS software, and identify specific media need for storage.

5.0 Verification Requirements

Verification requirements take into account how customer acceptance will be conducted at the completion of the project. Verification requirements specify how the functional and the performance requirements are to be measured and verified. The measurements taken may include simulation, emulation and live tests with real or simulated inputs.

Note: For the Archive Database/Files System, Verification Requirements identify some guidelines for storage of certain data values.

6.0 Acceptance Testing Requirements

Acceptance test requirements detail the types of tests which are to be performed prior to customer acceptance. These tests should be formalized in an acceptance test document.

Note: For the Archive Database/Files System, Acceptance Testing Requirements have not yet been identified in detail.

7.0 Documentation Requirements

Documentation requirements specify what documentation is to be supplied to the client, either through or at the end of the project. The documentation supplied to the client may include project specific documentation as well as user guides and any other relevant documentation.

Note: For the Archive Database/Files System, Documentation Requirements focus on type, aspects and media format of documentation that must be provided.

8.0 Quality Requirements

Quality requirements will specify any international as well as local standards which should be adhered to. These sections can include: quality factors, correctness, reliability, efficiency, integrity, usability, maintainability, test-ability, flexibility, portability, reusability, inter-operability and additional factors.

Note: For the Archive Database/Files System, the focus of the Quality Requirements is on QC flags.

9.0 Reliability Requirements

Reliability requirements are those which the software must meet in order to perform a specific function under certain stated conditions, for a given period of time.

Note: For the Archive Database/Files System, Reliability Requirements are focused on protection from either failure or obsolescence and recovery from disaster.

10.0 Maintainability Requirements

Maintainability requirements look at the long term life of the proposed system. Requirements should take into consideration any expected changes in the software system, and any changes to the computer hardware configuration.

Note: For the Archive Database/Files System, Maintainability Requirements are focused on hardware and software technical refreshment.

Archive Database/Files System Requirements

1.0 Functional Requirements

1.1 Structural

FS-1 Allow storage of the following data types[1]:

Items 1) - 3) will be stored in rdbms. Items 4) - 7) will be stored as flat files.

1) Point Data

Definition: Point data are observations or forecasts of a physical parameter (such as temperature, river stage, precipitation, etc.) taken at or valid at a single one-dimensional point such as at a weather station or a river gauge. The key parameters are where (station location by station ID or lat/lon/(elev)), what (physical parameter), and when (observation time or forecast valid time). The data are representative of/for a point on the earth's surface (or under the surface or in the atmosphere). Reference data for point data are the meta-data that describe~~s~~ this specific location, such as station name, county and state point is located in, stage-discharge relationship for this location.

a. Reference data
   - allow storage for current location information, location history and history of revisions to locations
   - ability to add "local data" specific to an RFC to the archive (e.g. at MBRFC, the rating shift information and date applied)
   - allow storage for rating curve history and history of revisions to rating curves.
   - slope profiles (APRFC)
   - longitude correctly stored with negative sign to be consistent with the international standard for latitudes and longitudes.
   - allow for storage of reservoir reference data such as storage-elevation curves.
b. Observations (see SHEF PE codes[2])
c. Forecasts (see SHEF PE codes, refer footnote 2)
d. Derived data (aggregations, summaries, statistics, etc.)
   - Accumulations (hourly, daily, monthly, yearly, seasonal, etc.)

---

[1] Some of these data types will require changes to SHEF code, see FS-9.

[2] Weather Service Handbook No.1, Standard Hydrometeorological Exchange Format, version 1.3

- Statistics (mean, median, standard deviations, departure from normals)
- selected reservoir statistics such as end-of-month reservoir content as percent of usable capacity

e. Profile data (slope measurements, soil moisture at depths, soil temperature at depths, snow temperature at depths, air temperature at height)

2) Areal Data

Definition: Areal data are processed observations or forecasts of a physical parameter that is characteristic of an area on the earth's surface such as a river basin, a county, a forecast zone, etc. The data represent a two-dimensional area rather than a single point. Most often, point observations are processed into areal values through some kind of averaging scheme such as the creation of Mean Areal Precipitation from point observations of precipitation at rain gauges.

a. Reference data (including OFS area and basin definitions)
b. Observations (see SHEF PE codes, refer footnote 2)
c. Forecasts (see SHEF PE codes, refer footnote 2)
d. Model states
e. FFG (see SHEF PE codes, refer footnote 2)
f. Derived data (aggregations, summaries, statistics, etc.)

3) Model Data
a. Reference data (ForecastGroup, CarryoverGroup)
b. Segment definitions
c. Mods

4) Text Products, including but not limited to:
a. river forecasts (RVF)
b. flash flood guidance (FFG, FFH)
c. coordination messages (HCM, HMD, ADM)
d. web products
e. AHPS products
f. water supply products & drought contingency discussions
g. snowmelt & spring outlook products
h. other (contingencies, QPF, etc.)
i. FLDWAV products

5) Grids, including but not limited to:
a. Observations (MPE)
b. Forecasts (HAS-QPF)
c. FFG

6) Graphics, including but not limited to:
a. river forecasts

b. web products
c. AHPS products
d. water supply products
e. snowmelt & spring outlook products
a. other (contingencies, QPF, etc.)
b. FLDWAV products

7) Images, including but not limited to:
   a. river forecasts
   b. flash flood guidance (FFG,FFH)
   c. web products
   d. AHPS products
   e. water supply products
   f. snowmelt & spring outlook products
   g. GIFs and GIS maps
   h. digital photos

8) The following products were taken into consideration but are outside the scope of this endeavor.
   a. Satellite precipitation estimates
   a. Numerical weather model grids
   a. HPC-QPF (handled by NPVU)

FS-2 For the data described in FS-1 provide for storage of observations as frequently as they are recorded and allow for changes to frequency of reporting.

FS-3 Allow additional and/or new observation data and/or types to be added at any time (e.g., if a new data sensor is added during the year, or USGS daily flow data for the previous year becomes available).

FS-4 Provide ability to store archived observed comments/remarks/notes with the archived data.

FS-5 Allow for observation data values at different QC levels for the same location/area.

FS-6 Allow storage for all NWSRFS developed time series (e.g., runoff components) as currently found in the Processed Database.

FS-7 Allow storage for ensemble (ESPADP) information and time series.

FS-8 Allow storage for SWS[3] pure and adjusted model output data.

[3] Statistical Water Supply System developed by CBRFC and used by several RFCs.

FS-9 While beyond the scope of this project, SHEF will need to accommodate the following:

1. area vs. station data
2. allow multiple sensors of same data type at a station
3. snow temperatures at various depths
4. soil temperature at various depths
5. soil moisture at various depths
6. slope measurements
7. air temperature at various heights
8. levels of data quality
9. number of days
10. SHEF codes for ESPADP time series output[4]

1.2 Dynamic

FD-1 Input

1. Parse/post SHEF
2. Parse/post DATACARD
3. Manual data entry for any time period

FD-2 Editing

1. Edit any data set and provide capability of retaining any corrected value(s)
2. Edit and retain remarks, comments and notes about any data set
3. Edit an extracted subset of data from a prescribed set of search criteria

FD-3 Transformation

1. Ability to add/subtract/multiply/divide all or selected data by a constant
2. Store/calculate river stage from slope measurements (note this will require dated slope profile records)
3. Store/calculate river flow from stage measurements (note this will require dated rating table records)
4. Store/calculate incremental precip (e.g. hourly, six-hourly, daily) from accumulated precip
5. Store/calculate snow density from snow depth and snow water equivalent data
6. Store/transform selected time series from smaller time step to larger time

[4] CBRFC has a proposal.

step.

FD-4 Output (reporting/querying)

1. List selected meta-data (location information & history, area, segment and operation definitions)
2. List data in various forms with a user specified period
   a) Text/Tabular
   b) SHEF
   c) DATACARD
   d) by water year or calendar year
   e) One line per record, delimited (unload format, for ingest into other databases or spreadsheets)
3. Graphical
   a) Line graph (e.g. hydrograph, snow water equivalent)
   b) Bar chart (e.g. incremental precip)
   C) Multi-composite graphs (e.g. Precip Bar Graph and Stage/Discharge Hydrograph v/s. Time in hours)
4. GIS
   a) Points
   b) Grids
   c) Mean-areal

5. List data deficiencies (missing obs or “holes” in the data)
6. List comments/notes about data, optionally list with the data
7. List data-set characterization terms (mean, median, max, min, std dev, etc.)
8. Allow printing of any display, graphical or tabular.
9. Allow for display of products including web products, RVF’s, AHPS products, and water supply products.
10. Allow for display/reporting of all OFS output data including adjusted observed flow, simulated flow, forecast flow, adjusted forecast flow, model states and mods

FD-5 Accumulation/aggregation

Store/calculate data aggregations for user selected time period including hourly, 3-hourly, 6-hourly, daily, monthly, seasonal, water year, calendar year, and other user specified periods.

FD-6 Characterization

1. Store/calculate statistics for period of record, current climatic normal period (e.g. 1971-2000), and user selected period for the following metrics: Mean, Median, Std. Deviation, Maximum and Minimum.
2. OFS
   a) Store/calculate model state statistics for season or user specified period.

b) Store/calculate OFS mods frequency analysis grouped by: season, type of year, forecaster, and basin.

FD-7 Analysis

1. Store/calculate data for OFS forecast analysis (bias, sensitivity)
   a) Contingencies (different QPF, QTF, freezing level)
   b) Operational (OFS) network vs. calibration (MCP) network
      i) Data (map, mat) (how are the two networks different?)
      ii) Forecasts (what effect do the different networks have on the forecast?)

2. Store/calculate data for ESP forecast analysis (bias, sensitivity)
   a) Weighting schemes
   b) Sensitivity to different model states (e.g. with/without mods)
   c) Comparison of skill to Statistical Water Supply (SWS)

3. Pair forecast stage values with corresponding observations for verification???.

2.0 Performance Requirements

PE-1 Generally, the Archive Database/Files System should be structured to allow efficient retrieval of data. The expected performance/speed of queries for data will vary according the different data categories, including: observations, models gridded/graphical, etc.

PE-2 Response times for various types of to be determined at a future date. Expectations of RFCs is that response times should take from a few seconds for simple tasks, a few minutes for moderate tasks and several minutes for difficult tasks.

3.0 Interface Requirements

Duplications of some items here with items in section 1.2 are noted but allowed because of different intent.

IN-1 Allow for the access to both archive and operational data (NWSRFS, IHFS-DB) in a manner seamless in display and transparent in retrieval.

IN-2 Allow OFS data to be exported from the archive data sets.

IN-3 Allow the import of any back-end parsing and posting of OFS outputs, state variables, etc. to the archive.

IN-4 Allow for accessing/converting to-from/saving different formats, including but not limited to: GRIB, netCDF, ASCII, SHEF, Shape[5], DATACARD, etc.

IN-5 Allow for retrieval of non-observational data from Informix files (lat, lon, elev, etc.)

IN-6 Allow for data to be imported into other applications, such as spreadsheets, word processors, relational DBMS and GIS package.

IN-7 Allow for access to all storage media used for archived data

4.0 Resource Requirements

RS-1 A dedicated Archive Database/Files System Server is needed at each RFC to store local archives and must operate in conjunction with all AWIPS components.

RS-2 Ability to easily add storage devices to the Archive Database/Files System Server should be taken into the design consideration.

RS-3 Ability to move data sets to off-line storage should be available through other storage media such as CD, tape, etc.

RS-4 Although beyond the scope of this project recognize the need for archiving of selected “global” data. Refer to FS-1, #8 for more information.

5.0 Verification Requirements

VE-1 The data in its different forms tied to a date-time stamp for consistent and easier access.

6.0 Acceptance Testing Requirements

AT-1 The Archive Database/Files System must successfully run through a test suite yet to be developed.

AT-2 The Archive Database/Files System Server must be accessible to all NWS AWIPS servers.

[5] widely used format used by gis applications for storing spatial data

7.0 Documentation Requirements

DO-1 Documentation for the Archive Database/Files System must address the needs of hydrologic modelers, application developers, and the end users.

DO-2 Documentation must be supplied for any COTS software which is not used in the basic AWIPS system. This documentation must include general information on running the COTS software, and specific documentation of how it is used in the Archive Database/Files System.

DO-3 Documentation of the complete database structure must be provided including data dictionary and data relationships. This documentation must include required and optional (if any) directory structures, files, formats, and tables.

DO-4 Documentation of procedures to move data from AWIPS and/or incoming data stream into the Archive Database/Files System must be provided. This documentation will also cover procedures for removing data from the Archive Database/Files System.

DO-5 Documentation of "internal" quality control operations must be provided.

DO-6 Documentation of "accessory" programs (e.g., to retrieve data in various formats such as SHEF or DATA CARD) must be provided.

DO-7 Documentation must be current and available through electronic and paper media.

8.0 Quality Requirements

QU-1 Allow storage of QC flags with the data.

QU-2 Allow for multiple levels of data quality within the system (e.g., data as received and latest corrected data)

QU-3 Allow for comparison of the distributions and statistics of the clean published data with the data at various levels of internal and external (e.g. manual edits) processing.

QU-4 Allow QC flagging of data to reject based on reasonable ranges (e.g., keep the value, but QC flag to indicate rejection due to out-of-range condition).

QU-5 An adequate backup and recovery system must be provided to minimize disruptions to RFC operations.

9.0 Reliability Requirements

RL-1 Each Archive Database/Files Server (and Client, if separate) will need a continuous source of power via a Universal Power Supply to prevent outages in the event of utility failures.

RL-2 The operating platform used to host the Archive Database/Files Server (and Client, if separate) should always be on the vendor's list of currently maintained equipment.

RL-3 The operating system used by the Archive Database/Files System should always be a currently supported version.

RL-4 It must be possible to restore a failed system to minimal operations within *X* hours, and full operation within *XX* hours/days. (via copying of required files from a backup, or replacing equipment and restoring from backup.)

RL-5 If the system is designed with a fail-over backup, the primary system must be restored to full operations within *XXX* hours/days.

Note: The number of hours/days (*X, XX, & XXX*) referred to in RL-4 and RL-5 are yet to be determined and will be dependent upon implementation method/system chosen.

10.0 Maintainability Requirements

MA-1 The system will allow for installing additional storage capacity and/or replacing storage capacity as the need arises for each server platform.

MA-2 The system will allow for frequent technical refreshment of the operating platform, operating system, and COTS software. Hardware refreshments should be aimed to take advantage of price/performance improvements in the storage and retrieval of data. Operating system and COTS software refreshments should be aimed to take advantage of new capabilities and vendor support.

Methodology

Background Information

In the Fall of 2000 a team was formed by the Donna Page, RFC Development Manager at OHD. This team is comprised of personnel from the RFCs, OHD and OS/HSD. The team was formally defined in early January 2001. The vision of this team is to create an RFC archive database design that will be used by all RFCs and the team must deliver to the RDM a database design document by September 30, 2001. For more information about the team go to *www.nws.noaa.gov/er/nerfc/archive/archivedb* . This site is maintained by team leader Victor Hom (NERFC).

How This Requirements Document was Created

Donna Page sent a survey to all the RFCs in the Fall of 2000. Draft #1 was put together by Dan Urban, contractor at OHD based on the replies by the RFCs. Eight RFCs provided input, the offices were: APRFC, CBRFC, LMRFC, MBRFC, MARFC, SERFC, ABRFC and NERFC. Identifiable requirements were grouped according to European Space Agency categories, first preserving traceability to original source document. Each requirements category was edited for content, grouping similar requirements, building additional sub-categories where appropriate, and eliminating duplication where possible. During this step, some editorial license was exercised at the expense of preserving traceability. The resulting requirements for each category were numbered for unique identification.

Draft #1 of the requirements document was released to the team for review and feedback. Conference calls were held in January thru March and minutes taken during these calls were used to develop Draft #2 of the requirements document. This was rewrite was done by Dan Urban.

A team meeting and workshop was held April 18-19, 2001 in Salt Lake City, UT at CBRFC. Draft #2 was again review and a major rewrite of the Functional Requirements (Structural) was completed. Unreviewed parts of the document were assigned out to team members and Draft #3 was completed and redistributed to the team for finalization. Draft #3 will be reviewed one more time and any needed changes made before distributing to all the RFCs for one last round of comments before moving on the design phase. Team members Steve Shumate (CBRFC), James Paul (ABRFC), and Arleen Lunsfund (APRFC) each were responsible for review and rewrite of several sections of the requirements document. Draft #3 compilation was done by Julie Meyer (MBRFC).

Draft#4 incorporates changes suggested by team members prior to the May conference call.

Draft#5 incorporates changes suggested by the team during the May 21st conference call and final review prior to distribution to the field.

The document labeled FINAL encompasses all changes discussed following the review of this document by the RFCs, the Regions and OHD/HL.