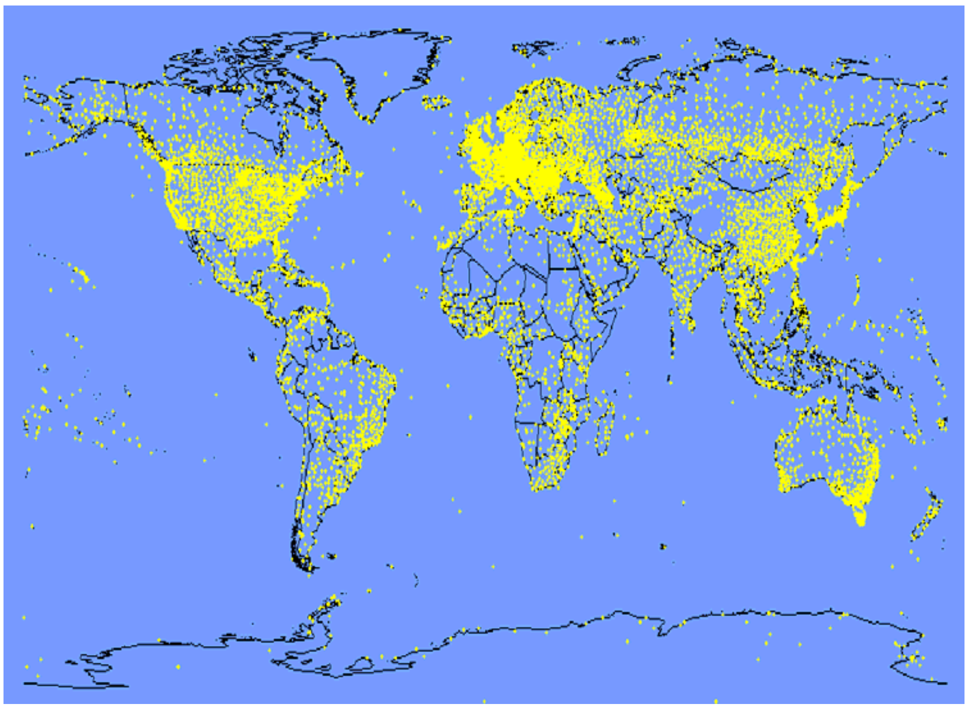
Technical Report 2001-01

==
==

National Climatic Data Center

The FCC Integrated Surface Hourly Database, A New Resource of Global Climate Data



US Department of Commerce
NOAA/ NESDIS
National Climatic Data Center
Asheville, NC 28801-5696
November, 2001

==
==

==
==

National Climatic Data Center Technical Report No. 2001-01

The FCC Integrated Surface Hourly Database, A New Resource of Global Climate Data

Neal Lott, Rich Baldwin, Pete Jones

November 2001
(Front cover depicts the location of stations in the database.)

U.S. Dept of Commerce
National Oceanic and Atmospheric Administration
National Environmental Satellite Data and Information Service
National Climatic Data Center
Asheville, NC 28801-5001

==
==

Table of Contents

Introduction

As a result of Environmental Services Data and Information Management (ESDIM) funding, Office of Global Programs (OGP) funding, and extensive contributions from member agencies in the Federal Climate Complex (FCC), the National Climatic Data Center (NCDC) has completed two phases of the Integrated Surface Hourly (ISH) database project:

1) The "database build" phase, producing ISH version 1 – The new database collects all of the NCDC and Navy surface hourly data (TD3280), NCDC hourly precipitation data (TD3240), and Air Force Datsav3 surface hourly data (TD9956), into one global database. The database totals approximately 350 gigabytes, for nearly 20,000 stations, with data from as early as 1900 to present. The building of the database involved extensive research, data format conversions, time-of-observation conversions, and development of extensive metadata to drive the processing and merging. This included the complex handling of input data stored in three different station-numbering/ID systems.

2) The first phase of quality control, producing ISH version 2 – This includes a) correction of errors identified after the "database build" phase (e.g., due to input data file problems); b) research, development, and programming of algorithms to correct random and systematic errors in the data, to improve the overall quality of the database; and c) data processing of the full period of record archive through these quality control algorithms (to be completed by early 2002).

The database has been archived on NCDC's Hierarchical Data Storage System (HDSS, tape-robotic system), and is in the process of being placed on-line. All surface hourly climatic elements are now stored in one consistent format for the full period of record. The database will be operationally updated on a weekly basis for 2001 and beyond.

Surface hourly is the most-used type of climatic data for NOAA customer-servicing and research, involving requests for the hourly data and for applications/products produced from the data. ISH will greatly simplify servicing and use of the data, in that users will not have to acquire portions of three datasets with differing formats, and will not have to deal with and program for the inconsistencies and overlaps between the three input datasets. Also, these accomplishments have created an end-to-end process for routine database updates. Finally, the database is being placed on-line for WWW access, and the more recent data are being placed onto a CDROM product.

Background

Participating Agencies

The Federal Climate Complex (FCC) is comprised of the Department of Commerce's National Climatic Data Center (NCDC), and two components of the Department of Defense (DOD)--the Air Force Combat Climatology Center (AFCCC) and the US Navy's Fleet Numerical Meteorological and Oceanographic Command Detachment (FNMOC Det). The FCC provides the nation's climatological support. The purpose of the FCC is to provide a single location for the long term stewardship of the nation's climatological data, and to provide the opportunity for customers to request any climatological data product from a single location.

The Problem

Surface-based meteorological data are by far the most-used, most-requested type of climatological data. However, a single integrated database of global hourly meteorological observations did not exist. Researchers requiring surface climatic data often acquired the data from several sources in a variety of formats, greatly complicating the design of their applications and increasing the cost of using the data. For example, when someone needed all available surface hourly data for a selected region (U.S. or worldwide) and time, they would receive data from three datasets which differed in format, units of storage, and levels of quality control. Alternately, the user would simply choose which one of the datasets might be able to meet their requirements, which often resulted in incomplete or inaccurate results. Many users complained about the problems this created in data usage, and in getting complete, accurate results.

An integrated upper air observational database (Comprehensive Aerological Reference Data Set--CARDS) was produced by NCDC to meet similar needs for upper air data. The ISH database is much larger in volume than CARDS, and will have an even greater frequency of usage. Additionally, the currently available datasets of monthly resolution (e.g., monthly precipitation totals) being used for climate change research do not provide the hourly and daily resolution required to describe regional climatologies and investigate extreme events, such as a 1-day rainfall event causing flash flooding. These monthly datasets also do not usually provide other needed parameters such as cloud data, visibility, wind speed, snow depth, etc. Therefore, this project was required to produce a single, integrated, quality-controlled, complete global climatic database of hourly data, for use in numerous applications, by private and public researchers, corporations, universities, and government agencies.

Development of ISH

Input Data Sources

The datasets used as input for the ISH database were:

1) NCDC and Navy surface hourly data, referred to as TD3280 (Steurer, Bodosky, 2000). (TD is an old reference to "tape deck.")
- Data Type: ASCII character data.
- Quality Control: Undergoes extensive automated and manual quality control.
- Data Origin: Mainly from Automated Surface Observing System (ASOS) stations currently, and previously from diskettes from the stations along with key-entered data.
- Content/Elements:

Comprises about 1400 National Weather Service (NWS) and U.S. Navy stations historically, generally for 1948 to present.

About 380 stations currently active.

Includes most surface elements observed in the U.S. (wind speed and direction, temperature, dew point, cloud data, sea level pressure, altimeter setting, station pressure, present weather, and visibility). Wind gust, precipitation amount, and snow depth are not included.

"Specials" are not included and only synoptic hours (every 3rd hour) are included for 1965-1981 (for most stations).

2) AFCCC Datsav3 surface hourly data, referred to as TD9956 (AFCCC, 1998).
- Data Type: ASCII character data.
- Quality Control: Extensive automated quality control; plus manual quality control for USAF stations.
- Data Origin: Mainly from Global Telecommunication System, various other sources, and keyed data prior to 1973. The keyed data include over 100 “tapedecks” which were laboriously converted to Datsav3 format by AFCCC, thus providing a great deal of data prior to 1973 for a number of international stations.
- Content/Elements:
  Comprises about 20,000 global stations historically, with many stations for 1973 to present, and some as far back as 1900.
  About 10,000 stations currently active.
  Includes all elements mentioned above for TD3280, along with wind gust, precipitation amount, snow depth, and other elements as reported by each station.
  Also includes "Specials."
  Observational practices vary by country.

3) NCDC hourly precipitation data, referred to as TD3240 (Steurer, Hammer, 2000).
- Data Type: ASCII character data.
- Quality Control: Undergoes automated and manual quality control.
- Data Origin: Various sources including ASOS and weighing rain gage data from stations.
- Content/Elements:
  Hour-by-hour precipitation amounts.
  Includes NWS stations and various cooperative observing sites.
  About 2800 stations currently active, with data generally for 1948 to present.
  ISH includes TD3240 stations which also have data available in either TD9956 or TD3280.

These three datasets were the most logical starting point for ISH, as they were the most-used hourly datasets available, and have also been subjected to considerable quality control and have adequate station history information available. If funding permits, other datasets and data sources will be added to ISH in the future.

The Process

All necessary metadata within the FCC were collected, coordinated, and loaded into a set of relational database tables. NCDC uses Oracle for this requirement. The metadata includes important information about the data such as station histories, dataset documentation, and inventories; along with critical information to control the process of merging the data. Since the three input data sources are archived in dissimilar station numbering/identification systems, the metadata must provide a cross-reference to identify data for the same location (i.e., same station with data in each of the three input datasets). This station history then controls the overall process flow and data merging, and also must account for station number changes over time.

A time conversion control file was used to convert the TD3280 and TD3240 data to Greenwich Mean Time (GMT), so that all input data were then in GMT time convention. The ISH data are therefore in the same time convention as CARDS upper air and many other global databases, model output, satellite data, etc. This is quite important for potential GIS applications. The creation of this time conversion control file was very cumbersome, involving research of several sources of information concerning time zones, time zone changes historically, etc. We had to account for time zone changes to properly merge the data.

We included a quality control check for TD9956 vs. TD3280 to ensure that the data were actually for the same location at the same time before performing the intra-observational merge, which creates a composite observation for that date-time. The quality control check was conducted on a daily basis (i.e., on each day's data) to determine if the data for that day should actually be merged into composite observations. Temperature, dew point, and wind direction were compared for each data value (e.g., temperature at a given station-date-time in TD9956 vs. temperature at same station-date-time in TD3280) to obtain a percent score for the day for coincident data. Criteria of 1-degree Celsius for temperature and dew point, and 10 degrees for wind direction were used as the pass/fail limits for each element, with an overall 80% score for the day required to perform the intra-observational merge for that day. In other words, 80% of the data values compared would have to meet the limit checks to "pass" for the day. Failure of these checks sometimes pointed to time conversion problems, updates to the control file, and re-processing of those stations. Subjective analysis of the data before and after processing proved this quality control check and the limits applied to be quite effective.

Procedures, algorithms, and then computer programs were written to merge the surface hourly datasets into one common database. More than one billion surface weather observations (covering 1900 to present) for approximately 20,000 global locations were accessed and merged during this process. Examples of input data types were: Automated Surface Observing System (ASOS), Automated Weather Observing System (AWOS), Synoptic, Airways, Metar, Coastal Marine (C-MAN), Buoy, and various others. Both military (e.g., USAF, Navy, etc.) and civilian stations, automated and manual observations, are included in the database.

The final version 1 database (see Lott, 2000) includes all observed elements at hourly intervals (every three hours for some stations), such as temperature, dew point, visibility, wind speed and direction, wind gust, cloud data, precipitation, snow depth, and many others. The database and subsequent products are available to all three organizations to support the FCC mission and its numerous customers.

To develop the version 2 database, we researched, developed, programmed, and processed the data through 54 quality control (QC) algorithms. This phase of quality control subjects each observation to a series of validity checks, extreme value checks, internal (within observation) consistency checks, and external (versus another observation for the same station) continuity checks. Therefore, it may be referred to as an inter/intra-observational quality control, and is entirely automated during the processing stage. However, it does not include any spatial quality control ("buddy" checks with nearby stations), which is planned for later development.

An example of just one of the algorithms performed is the continuity check for temperature, which does a "two-sided" continuity validation on each temperature value for periods ranging from 1 hour to 24 hours. An increase in the temperature of 8 degrees Celsius in one hour (e.g., from 10 C to 18 C) prompts a check on the next available (i.e., next reported) temperature–if that value then decreases by at least 8 degrees in an hour (e.g., 18 C to 9 C), then that indicates a very improbable "spike" in the data, and the erroneous value (e.g., 18 C) is changed to indicate "missing" for that observation. However, the original value is saved in a separate section of the data record for future use if needed. The same would apply for a downward "spike" in the data, and similar checks are performed for periods out to 24 hours, to allow for missing data and for part-time stations which do not report hourly or three-hourly data. The validation always checks the closest values available temporally (i.e., before and after the data point being checked), and the limit is automatically adjusted based on the elapsed time between values.

In addition, a selected number of systematic deficiencies are addressed with specific algorithms to correct those problems. Though all climatic elements are checked to some extent, the main elements validated are: wind data, temperature and dew point data, pressure data, cloud data, visibility and present weather data, precipitation amounts, snowfall and snow depth. As mentioned above, the input datasets had already been subjected to a great deal of quality control, so this additional quality control was designed to address problems which were less likely to have already been corrected. Detailed documentation on each of these QC algorithms is available (Lott, 2001).

See Appendix A for further details on the database design and processing.

Methodology and Milestones

Often, when creating a software system, there are numerous approaches that can be used. However, there are often specific reasons to go one route as opposed to another. Such is the case with the ISH population and management system.

When looking at a consolidated dataset that is to be accessed later from a relational database, and is the amalgamation of three datasets, the first idea is to consider doing the bulk of the work within the context of the Oracle database vs. performing the merging and preparation work beforehand. The latter approach was chosen due to the limited disk storage resources available to store all three parent datasets in Oracle. So the bulk of the prep work was done using Fortran and C programs.

Another historical note of interest is the number of people that worked on the system. During the course of several years there was a team effort of contractors and government employees working on the task. The responsibility for the contractors has been primarily the actual programming while the government employees, while doing some programming, have been primarily involved in testing the results. As is true with many software projects, the code evolved with time, and as necessary, work-arounds were developed to fix source code when problems developed.

Toward the end of the development phase, the original workstation for development and testing was replaced with a newer 64-bit workstation. Although the change should have been transparent to the ISH system, it was not. Many problems began to appear with code that had been working before the change. After much research, no cause could be identified, although evidence seemed to point to a system memory utilization problem. In any case, some work-arounds were put into place in order for some of the components to start working again.

One of the goals for ISH is to have the entire dataset available for query via the NNDC Climate Data Online (CDO) system (Lott, Anders, 2000). With the difficult and tedious task of blending the data from the three sources completed, end-users may then extract what is needed with relative ease and can focus on their research or studies, rather than getting bogged down in the merging process. A second goal is to continue to add to and improve this global baseline database for research and applications requiring data of this type, by adding additional datasets (i.e., merge into ISH), and developing/applying more extensive quality control checks.

As is the case with most software systems, ISH is designed to evolve, within the limitations of current funding and technology, to reach these goals. Here are the milestones for the overall effort:

- Develop and test the initial population system, including all required metadata and process control files. (Completed)
- Use this system to create the ISH database from the beginning of record to the present, and archive the database as ASCII files on the HDSS. (Completed)
- Store a portion of the database (as much of the more recent data as possible) in relational tables, and develop software for online access via the NNDC CDO system. (Completed)
- Modify this system to operate in a production environment on a routine (weekly to monthly) basis. (Completed)
- Develop, program, and process the database through a set of quality control algorithms, to improve the overall quality of ISH. (In progress, November 2001)
- Continue the routine updating of the database using the established procedures and software.
- As resources become available, make the entire database accessible on-line via NNDC CDO.
- As funding permits, add additional datasets to ISH, via the merging process.
- As funding permits, add additional station history/metadata to the database, to include as much instrumentation information as possible; thereby making the data more useful for climate change research.
- As funding permits, research, develop, and apply more sophisticated time series and spatial quality control checks to ISH; thereby making the data more robust and useful for all applications.

Funding proposals for FY 2002-2003 have been submitted to accomplish the latter three milestones.

Results and Conclusion

Project Deliverables

- ISH is an integrated surface hourly database constructed in a single standard format, and quality-controlled for the full period of record. Data quality flags and data source indicators are also included for each climatic element.

- ISH is a database which is useful for global, regional, and local-scale analysis and studies, and includes all observed parameters at each station. Approximately 20,000 worldwide stations are included, spanning time periods from as early as 1900 through to the present. Customers requiring either long-term or only the most recent data, either regional/global scale or simply one location's data, will access the same database.

- The overall system includes a complete on-line (WWW) inventory of the database for users to query, an on-line browse system for customers to browse and become familiar with the data, and direct on-line access to selected data from the integrated database.

Benefits

- The new, integrated database is a global baseline database and the first of its kind, and can be used by both U.S. and global customers.

- This effort contributes to the ESDIM goal of improving access for users. Researchers will be serviced more quickly and in a much more efficient manner at a lower cost to the FCC agencies (as compared to the personnel resources now required to access multiple datasets for customers). This in turn will reduce the overall operating costs to the government and the access and development costs for the users of the data.

- ISH provides a common data format for development of FCC applications software, for NOAA National Data Center applications, and for all researchers and users of the data to develop application programs easily and efficiently. Also, the database contains daily parameters (e.g., maximum/minimum temperature, daily precipitation) which can be used to generate daily summaries.

- The data can be used in important national and international research programs such as the World Climate Program; the US Weather Research Program; the Global Energy and Water Exchange Program; the International Geosphere Biosphere Program core projects in water, land, ocean, and biosphere; and NOAA's Office of Global Programs Climate and Global Change Program. Numerous environmental studies, coastal zone research, climatic extremes research, and blending with other data types (e.g., satellite, NEXRAD data) will utilize the integrated database.

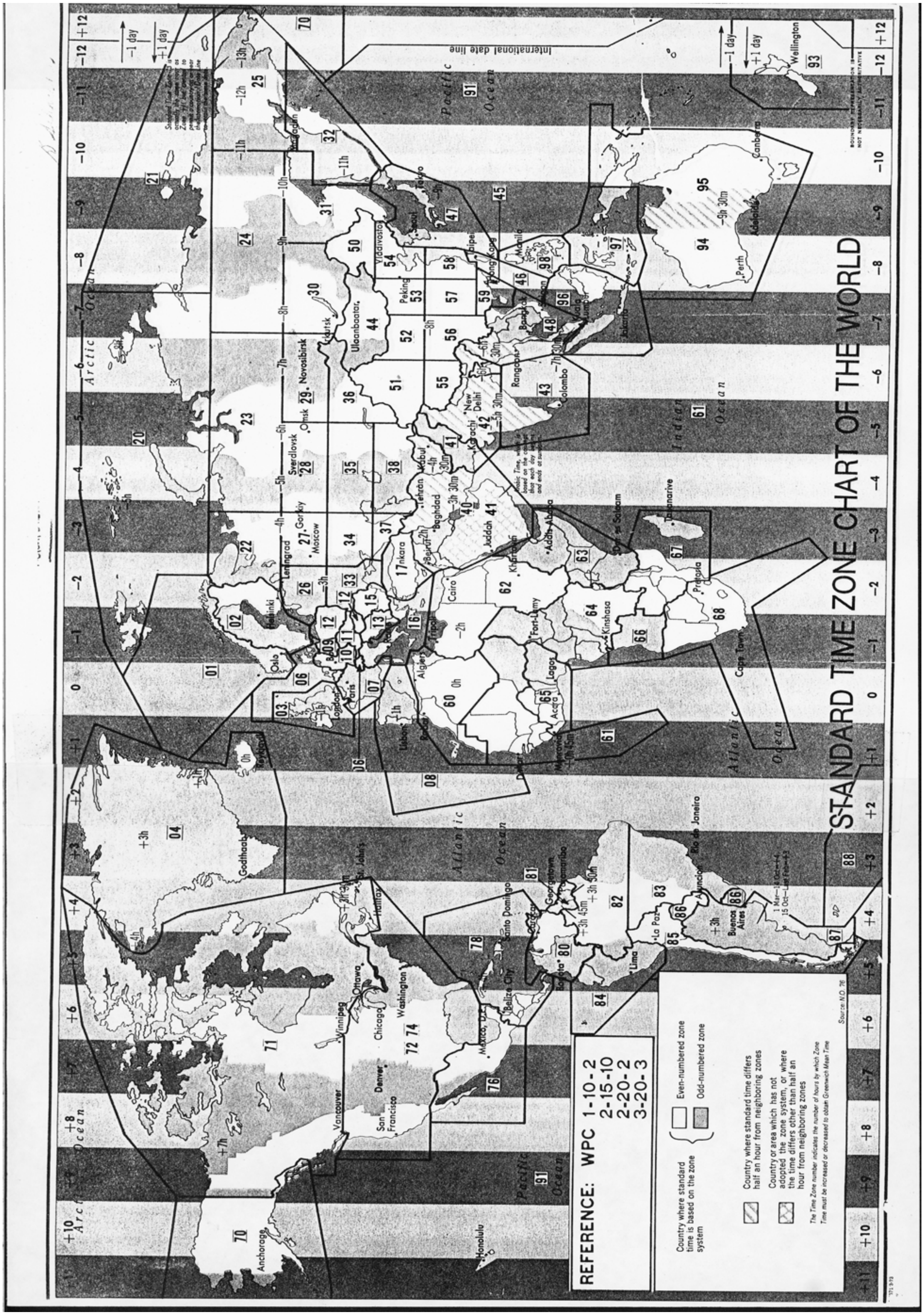
- This integrated database will be coordinated with and integrated into the goals and objectives of the National Environmental Data Index (NEDI), the NESDIS Virtual Data System (NVDS), the Intergovernmental Panel on Climate Change (IPCC), as well as other Climate and Global Change initiatives. The FCC agencies will continue to append to the database with future datastreams, thereby allowing for continued growth of the database into future years.

Database Inventory

The following table provides an inventory by WMO block of the number of stations in the ISH database having at least one year of data. See the map on next page for WMO block locations.

WMO Bk	Stations	WMO Bk	Stations	WMO Bk	Stations	WMO Bk	Stations
01	297	30	230	58	189	88	36
02	413	31	227	59	149	89	124
03	593	32	102	60	193	91	234
04	105	33	190	61	128	92	12
06	275	34	170	62	125	93	82
07	297	35	143	63	143	94	622
08	165	36	130	64	155	95	163
09	74	37	208	65	123	96	92
10	491	38	202	66	30	97	67
11	226	40	356	67	186	98	85
12	219	41	202	68	270	99	334
13	153	42	225	69	548		
14	71	43	151	70	223		
15	322	44	100	71	1085		
16	280	45	13	72	2113		
17	122	46	60	74	700		
20	59	47	521	76	261		
21	58	48	350	78	244		
22	180	50	52	80	130		
23	177	51	68	81	27		
24	116	52	107	82	144		
25	99	53	118	83	290		
26	220	54	178	84	101		
27	147	55	22	85	163		
28	237	56	163	86	57		
29	178	57	243	87	183		

Map of WMO Block Locations:



Acknowledgments

There were many people who contributed to this project's success. The key members of the team were: Rich Baldwin, NCDC; Vickie Wright, NCDC; Dee Dee Anders, NCDC; Danny Brinegar, NCDC; Neal Lott, NCDC; Pete Jones, Marada Corporation; and Fred Smith, Marada Corporation. As mentioned previously, the Air Force (AFCCC) and Navy (FNMOC Det) contributed to the effort. Of particular note is Bob Boreman (NCDC) who devoted a great deal of time and effort to ISH, especially in the research and development of the time conversion control file. Bob passed away in July 2001 and will be greatly missed.

References

AFCCC. Documentation for Datsav3 Surface Hourly Data. [Asheville, N.C.]: Air Force Combat Climatology Center, 1998.

Lott, Neal. Quality Control of USAF Datsav3 Surface Hourly Data–Versions 7 and 8. [Asheville, N.C.]: USAF Environmental Technical Applications Center, 1991.

Lott, Neal. Data Documentation for Federal Climate Complex Integrated Surface Hourly Data. [Asheville, N.C.]: National Climatic Data Center, 2000.

Lott, Neal. Quality Control Documentation for Federal Climate Complex Integrated Surface Hourly Data. [Asheville, N.C.]: National Climatic Data Center, 2001.

Lott, Neal and Dee Dee Anders. NNDC Climate Data Online for Use in Research, Applications, and Education. Twelfth Conference on Applied Climatology. Pages 36-39. American Meteorological Society, May 8-11, 2000, Asheville, NC.

Plantico, Marc and J. Neal Lott. Foreign Weather Data Servicing at NCDC. ASHRAE Transactions 1995, V. 101, Pt 1.

Steurer, Pete and Matt Bodosky. Data Documentation for Surface Hourly Airways Data, TD3280. [Asheville, N.C.]: National Climatic Data Center, 2000.

Steurer, Pete and Greg Hammer. Data Documentation for Hourly Precipitation Data, TD3240. [Asheville, N.C.]: National Climatic Data Center, 2000.

Appendix A – Processing Details and Design

This section describes in much more detail the design and processing of ISH.

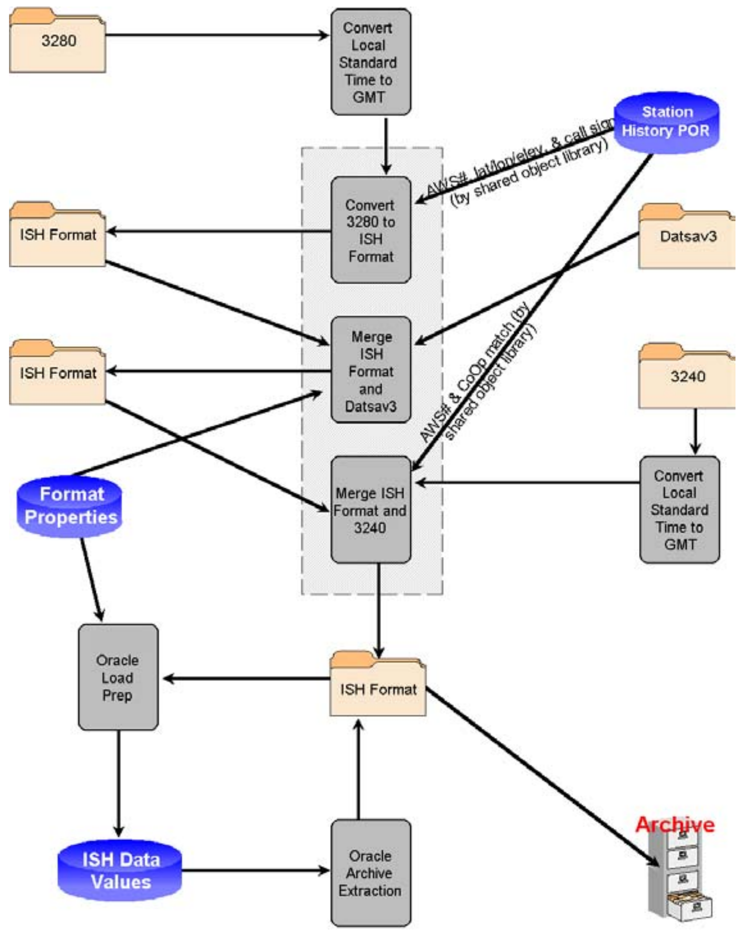


Figure 1 - ISH Initial Population Process Flow

Within the scope of the current project, a process flow has been designed and developed to meet the first goal of initial creation of the dataset from the beginning of record to the present. Figure 1 illustrates this process flow. Each of the file folders represents an ASCII file. The rounded corner boxes are processes implemented as programs. The drums are Oracle tables. Data are processed in WMO-block sequence, for WMO blocks 01 through 99.

Essentially, the process follows these steps in merging the three datasets together:

1. <u>TD3280 data are converted from local standard time (LST) to Greenwich Mean Time (GMT)</u>. This is done to then use GMT for all global stations.
2. <u>Convert any TD3280 data from its native element-oriented format to the ISH format</u>. This process also filters out several elements that are calculated from other base elements and transforms many of the observation values to fit the ISH scheme. Notice also that the Oracle database contributes the AWS ID, latitude, longitude, elevation, and call sign to be included in the ISH output file.
3. <u>Merge the TD3280 data and Datsav3 data where possible</u>. In this step, we are pushing both types of data through the program regardless of whether or not a merge can actually take place. The net result is that both datasets come of the process in the ISH format with the data source flag set appropriately.
4. <u>Convert any TD3240 data from LST to GMT</u> that will be merged into the TD3280/9956-ISH data.
5. Perform the <u>merge of the TD3240 data into the 3280/9956-ISH data</u>.
6. <u>Store the ISH data</u> in the HDSS archive.
7. <u>Store the data in the Oracle database</u>. The Oracle "prep program" splits out the given ISH file into three text files that exactly match the Oracle table structures. These temporary load files are then imported rapidly into Oracle through SQL Loader.

Data Flow Diagrams

The purpose of these diagrams is to map the paths that the data in the ISH production system take while undergoing the transformation and merging process. It is important to realize that they are not flow charts and are not necessary chronological. Nor are they intended to be comprehensive. Not all the process boxes represent one and only one program. Often, several processes are performed together in the same code.

The diagrams begin with level 1 which is an overview of the major flows of data within the system. The next three diagrams "explode" three of the processes (i.e., 1.1, 1.3, and 1.4) to show more detail. Process 1.2 was not exploded because it is descriptive enough at level 1.

Data Flow Diagram Level 1 - Overview

This diagram, shown in Figure 2, depicts the four main processes of the ISH production system. The first is "1.1 - Data Selection and Inventory," which is the selection of the TD9956, TD3280, and TD3240 data files and their inventory processes for quality control. The second is "1.2 - Difference Input and Output." This process compares the ISH inventory counts to the greatest of the TD9956 and TD3280 inventory counts for the same station, year, and month. The differences are reported to the I_Diff table where they can be queried.
Note: "Arid" refers to the name of a workstation.

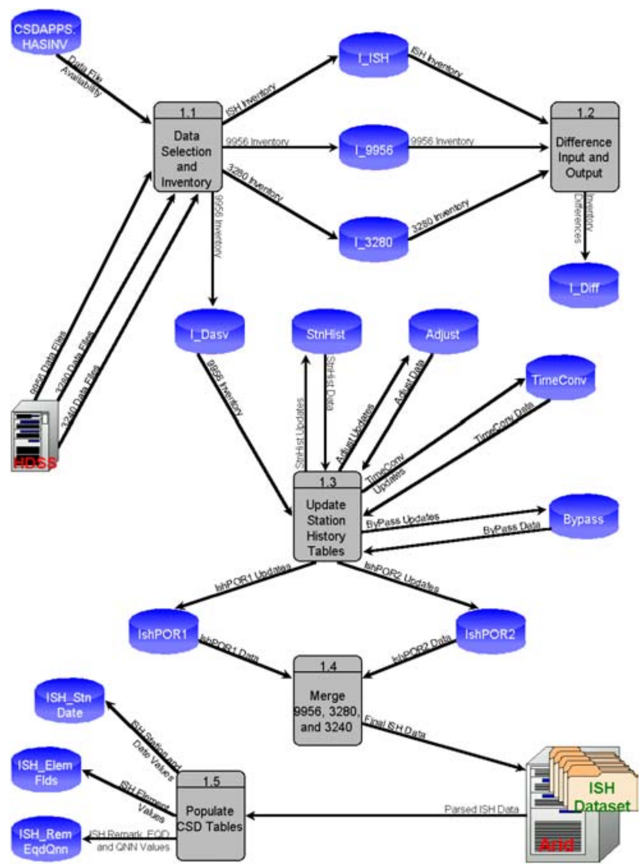


Figure 2 - Data Flow Diagram, Level 1

<u>Data Flow Diagram Level 1.1 - Data Selection and Inventory</u>

This diagram, shown in Figure 3, outlines the flow of data from the various components of the data selection and inventory process. Essentially, this is the code that the user employs to bring the desired TD3240, TD3280, and TD9956 data files to the workstation for conversion, merging, and inventory. If the requested files are already available, then it checks for date differences to determine if what's already available is the latest/best data to use. If not, the newest data files are transferred for processing.

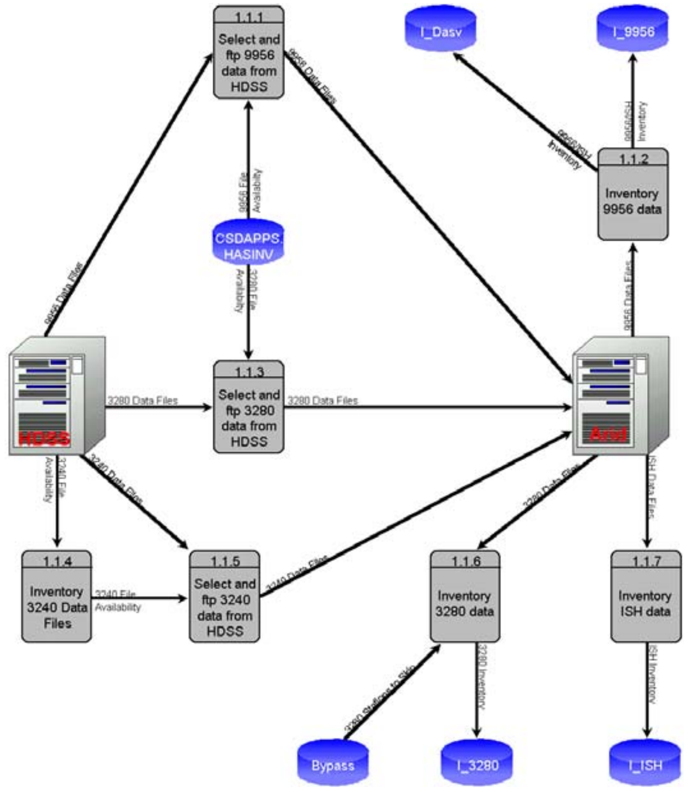


Figure 3 - Data Flow Diagram, Level 1.1 - Data Selection and Inventory

The inventory component is run in conjunction with the merging processes to provide primarily two things. First, they are used to be sure that after the transformations, eliminations, merging, and repackaging of the conversion and merging programs are finished, that none of the essential data values have been lost. Second, the inventories can be used to provide a thumbnail sketch of the dataset for users to quickly identify what is available.

Data Flow Diagram Level 1.3 - Update Station History Tables

This diagram (Figure 4) explodes the process involved in the update of the station history tables. In order for the system to have the necessary station history data, two tables were created to provide an inventory that can be used to match AWS, WBAN, and COOP station ID’s together for the given time frames–Ishpor1 and Ishpor2.

The Adjust, Bypass, and TimeConv tables are used as “control” tables to control various components of the data processing, while the StnHist table was created by a prototype program that ties the AWS, WBAN, and COOP together.

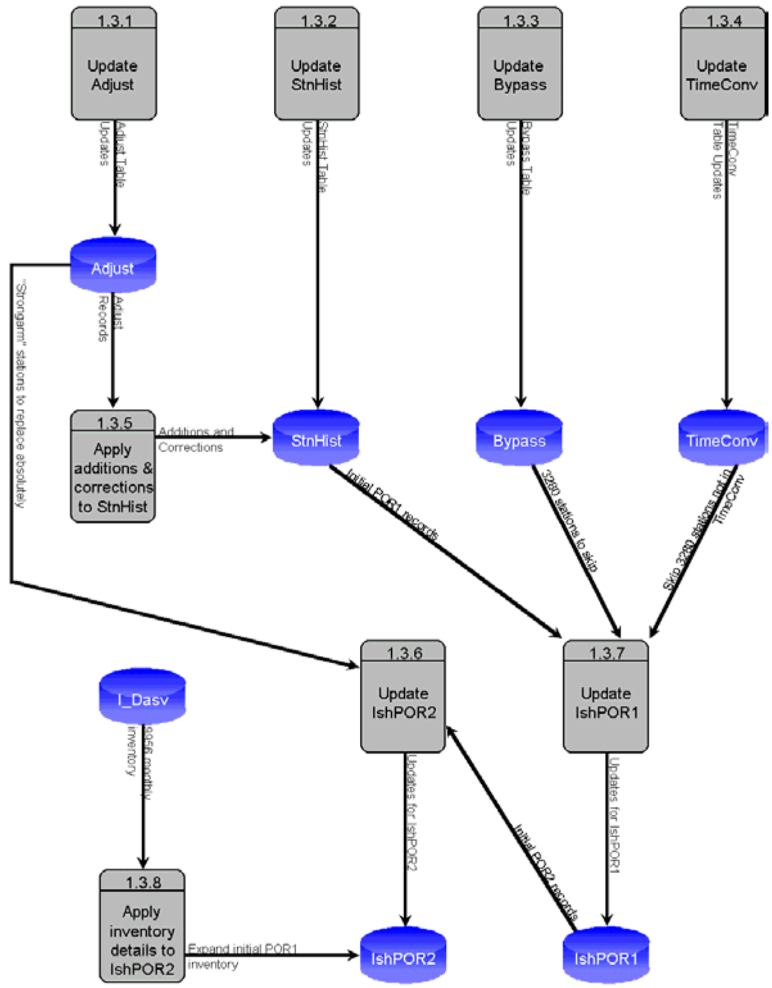


Figure 4 - Data Flow Diagram, Level 1.3 - Update Station History Tables

Data Flow Diagram Level 1.4 - Merge TD9956, 3280, and 3240

This final data flow diagram shows the data flows that pass between the various pieces of the merging process. Note that the file-level merge control (process 1.4.2) is implemented as a Perl script to call the other programs as needed based on the requested data to be merged. The IshPOR1 & IshPOR2 tables supply station data and the data necessary to match the TD3280, TD3240, and TD9956 data on a record by record basis.

Notice also that there are two levels to the merge: file and record. At the file level, a broad decision is reached as to which TD3280 and TD3240 files are available to select to blend with the requested TD9956 data. At the record level, the latest candidate TD3280 and TD3240 files have already been pulled from the HDSS and the decision level is narrowed down to each record (actually, each element in each record) where available TD3280 data generally take priority over the same TD9956 data for the same station at the same date.

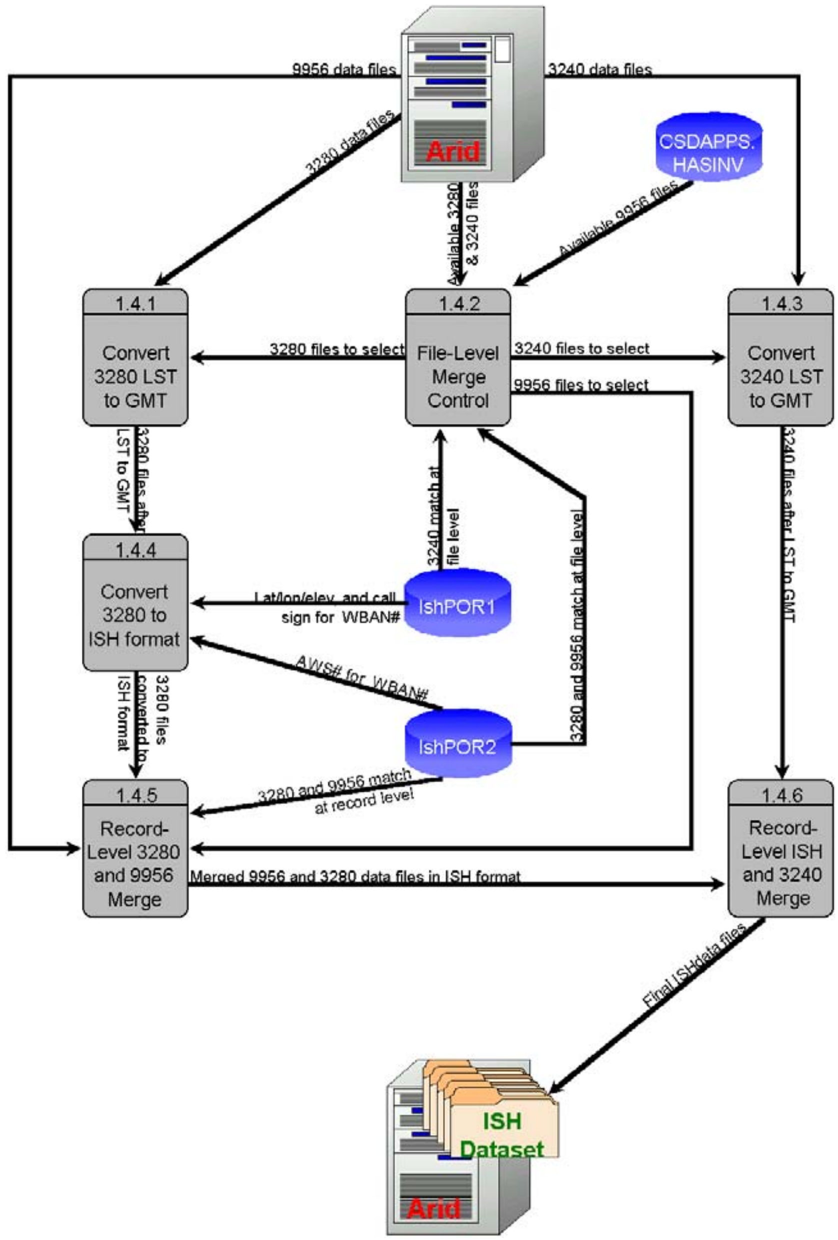


Figure 5 - Data Flow Diagram, Level 1.4 - Merge of 9956, 3280, and 3240 Datasets

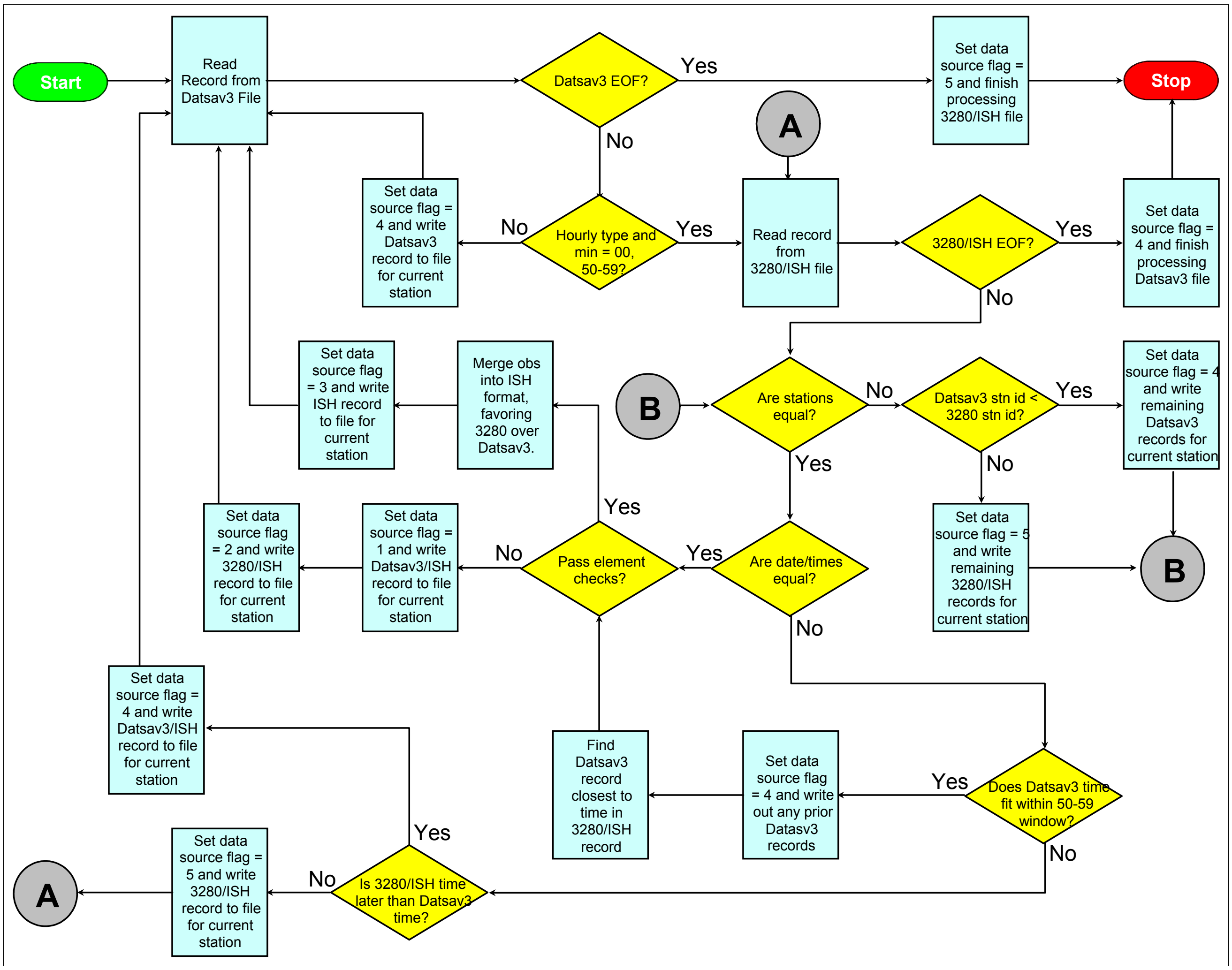


Figure 6 - General merge logic flowchart for TD3280 and TD9956.

<u>Merge of TD3280 (ISH Format) with TD9956 (Datsav3 Format)</u>

- Initialization
  - Reads and process command line input, trap any associated errors.
  - Opens input and output files, traps errors.
  - Call the daily merge routine for merging data by day.
  - Load include files (nodes.h, state.h, record.h, element.h, rules.h).
  - Build a linked list containing information from ISHPOR2 table.
  - Set processing state variables.
- Processing
  - Process TD9956 data.
    - Check of input TD9956 file unit is made and state variables set.
    - Read TD9956 data record and store in element data structures.
    - Blank lines are trapped and any data stored in prior records is written out.
    - State variables are set to control conditions for TD9956 input.

        - Date and time variables and AWS and WBAN station identifiers are set.
        - Checks for specials and hourly type records (00 min or 50-59 min) are performed.
        - Specials and non-hourly records are written out with source code 4.
        - Processing continues by returning to item (1a) above to prepare to read another TD9956 record.
    - Process TD3280 data
        - File unit conditions are checked and state variables set for the TD3280 (ISH).
        - Read TD3280 data record into a element data structure.
        - Blank lines are skipped in the TD3280 file.
    - Merging
        - Comparison of AWS station identifiers from the TD9956 and TD3280. Check for various station/date ordering conditions, reset invalid station identifiers (999999), reset state variables, synchronize station identifiers by writing out records and linked-list buffers.
        - Comparison of current record date versus current processing date.
            - Records with matching dates are added to linked-lists for the current processing date (00 hours and 50-59 minute).
            - Records not matching the current processing date signify a new day or data gap. Continue merge process evaluation; set state variables.
        - Compute merge criteria values for temperature, dew point and wind direction.
            - If the merge criteria pass set percentage, then merge all records in each of the TD3280 and TD9956 data linked lists. Write out records with source code 3.
            - If the merge criteria fails, the TD9956 record will have a source code of 1 and the TD3280 record will have a source code of 2. If there is no matching observation for either the TD3280 or TD9956 record, then the TD9956 record is written with a source code of 4 and the TD3280 record is written out with a source code of 5. (Note: all ADD elements are sorted before each record is written.)
        - Special merging conditions.
            - The TD9956 data is used to fill in any missing data in the TD3280 record.
            - Retain TD9956 and TD3280 element MWn information.
            - Retain TD9956 and TD3280 EQD information.
        - Reset state variables before reading additional records. Read from appropriate file units as dictated by state variables.
- Cleanup
    - Reset linked-list and other variables to null after each day processed.
    - When TD9956 EOF mark is reached, merge and flush linked-list buffers.
    - Write out sorted element records and reset state variables.
    - Complete TD3280 file if necessary.

Merge of TD3280/9956 (ISH Format) with TD3240

- Initialization
  - Reads and process command line input, trap any associated errors.
  - Open input and output files, trap errors.
  - Call the merge routine for processing passing the I/O units
  - Load the include files.
  - Build a linked-list from the decompressed TD3240 data file for start and stop ranges.
  - Remove observations which have flags indicating erroneous data.
  - Set processing state variables.
- Processing
  - Process TD9956/3280 ISH formatted data
    - Check the condition of the file unit and state variables.
    - Read the1st record into element data structure (same as w/ the TD9956/3280 merge program.
    - Date and time variables are set.
    - Checks for specials and hourly type records (00 min or 50-59 min) are performed.
    - Write specials out with source code 4.
    - File unit conditions are checked and state variables set for the TD3240 data in ISH format.
  - Process TD3240 data
    - Pop 1st node from TD3240 linked-list.
    - Check state variables, date and time in the event the new record is a time duplicate of the prior record (source code 1 or 2). Use source code 2.
  - Merging
    - Comparison of station identifiers
      - Test station identifier equality. Throw an error if not true.
      - Function in place to fetch coop identifier from ISHPOR1
      - Checks station ordering conditions for synchronization
    - Continued processing for equality of date and time.
      - Add TD3240 information to (00 and 50-59 minute) ISH records.
      - Increment precipitation elements.
      - Insure all elements are sorted before written.
      - Write out data in the 50-59 linked-list not merged with TD3240 data.
    - Trap from data gaps and duplicate time records (1 or 2 source code)
- Cleanup.
  - Reset variables to null after each day processed
  - When 9956/3280(ISH) EOF is reached write all buffers.

Relational Database Structure

The ISH relational database contains more than just the tables needed to make the dataset available on-line. It also stores some support tables that provide station history, format properties, and inventory data. In fact, there are four groups of tables:

- Format Properties Tables - These are Field, Format, and Format_Field. The purpose of these tables is to provide formatting properties such as data type, length, default missing value, etc., to programs needing to parse or assemble data in the ISH format.
- ISH Data Tables - These three tables, ISH_StnDate, ISH_ElemFlds, and ISH_RemEqdQnn, store the actual ISH data that is to be available to users of the CDO system.
- Inventory Tables - There are a number of tables that store various parts of the ISH inventory. They are: I_3280, I_9956, I_Dasv, I_Diff, I_ISH, and I_IshElem.
- Station History POR Tables - These tables, Adjust, Bypass, IshPOR1, IshPOR2, StnHist, and TimeConv provide the necessary station history information needed to supply location, time-frame, and matching information to various components of the system.

The design of the relational (Oracle) database changed and grew over the course of the development effort. To that end, the three core data tables, ISH_StnDate, ISH_ElemFlds, and ISH_RemEqdQnn, are structured in an attempt to balance the rules of normalization with some denormalization in order to save space and to facilitate partitioning.

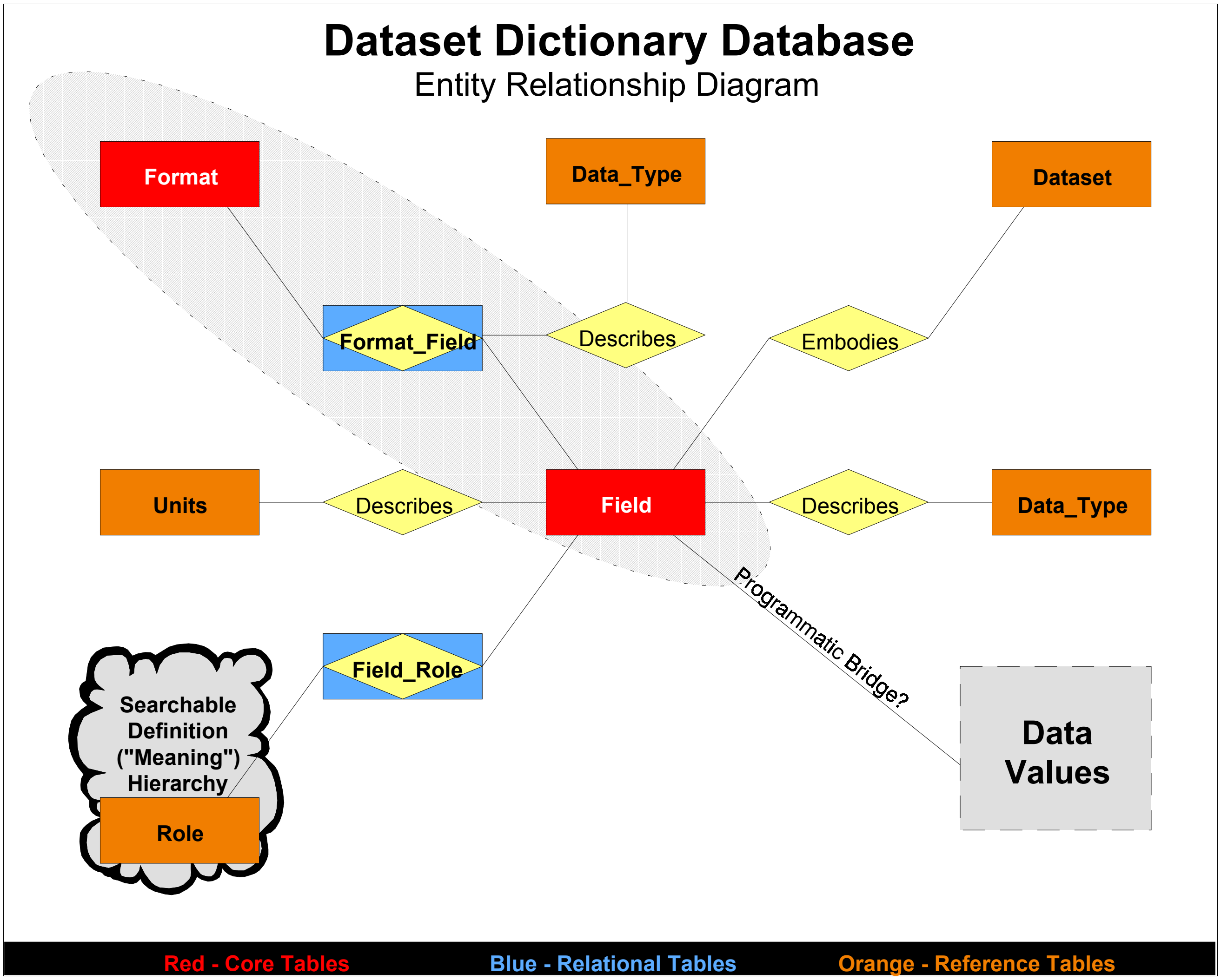


Figure 7 - Dataset Dictionary Entity Relationship Diagram

<u>Dataset Dictionary: Format Properties</u>

The intent of the Format Properties Tables also requires some explanation. These tables were developed from an older table which provided a self-referential description of the element properties. The design calls for a Dataset Dictionary Database that is intended to achieve four goals:

- <u>Parsing (Input)</u> - This feature provides the parsing intelligence needed so that a program can know how many positions a field takes up in the data file. It also contains data type information so the value can be stored in the program correctly.
- <u>Assembly (Output)</u> - This feature allows an application to reverse the parsing process to assemble data in the documented format.
- <u>Searching</u> - This feature builds in some indexing so that each of the fields of a format can be assigned a role. These roles can be added to increasingly more and more general roles in a self-referential hierarchy that allows a user to begin with a broad brush and "drill down" to those fields that interest them.

- Documenting - This feature attaches long names and text descriptions to the fields so that a format can be documented right out of the database.

If the database were to be normalized completely, it would call for separate tables for each of the elements with keys to the ISH_StnDate table. However, this makes any SELECT statements potentially quite long and juggling that many tables can be confusing. So all the elements are stored in the ISH_ElemFlds table with a sequence number primary key to hold it all together. See Appendix B for details on the elements contained within the database, and the ASCII file structure prior to loading into Oracle.

The ISH_RemEqdQnn table holds the remarks, EQD, and QNN sections, respectively. These sections are not split out into individual fields, but are kept intact for further parsing later, if desired. There is a one to one relationship between an entry in ISH_StnHist and in ISH_RemEqdQnn. The reason for this is to save space since not all ISH records will have these sections.

The Year column is repeated in all three tables so that they can be partitioned. This improves query performance and allows part of each of the tables to be taken "off-line" by year for loading or repair while the rest of the tables are available for use.

The Station History Tables

The core station history tables IshPOR1 and IshPOR2 are the ones used to add missing station data such as lat/lon/elev, call signs, etc. They are also used to match station ID's so that the correct AWS station ID can be paired up with the correct WBAN and COOP stations for the same time period. The Adjust, Bypass, StnHist, and TimeConv tables contribute to the building of these two tables.

The Inventory Tables

Each of these tables holds monthly observation counts (ie, number of data records for each station) for the respective data source. For example, I_9956 holds the counts for the TD9956 data for each station and year processed. The I_3280 holds the TD3280 counts, etc. The I_Diff table is the difference of the TD3280/9956 (whichever is greatest) and the I_ISH count for the same station and date. This table is used to identify any possible loss of data during the processing.

System Design and Development

Phases of the Development Process

There were four main phases used in the development of this system: Design, alpha, beta, and production. The term, "development" is used to mean the initial plans, programming, testing, and setup necessary to move the project from concept to use. This is important because there is another phase, maintenance, that is important, but is beyond the scope of development.

The Four Phases of Development

1) Design *(Planning)*
2) Alpha *(Construction)*
3) Beta *(Testing)*
4) Production *(Deployment)*

The dividing line between the phases is not always distinct; in fact, there are often overlap periods where two or even three of the phases are occurring simultaneously. Often a program will have parts of it that demote back to previous phases before "getting it right." But eventually all the phases are completed for all the components of the system at which point the development is finished.

Design Phase

Any information system must begin here. If the requirements are not understood clearly, then the development, testing, and production phases will be frustrating, time-consuming, and less stable. Ideally, the development team should meet together frequently to hash out the design ideas on paper. They should also involve the end-users as much as is reasonable. Once the ideas are expressed on paper, the actual programming and testing can begin. The rule of thumb is, "A problem well-defined is half-solved."

Alpha Phase

The alpha phase is the time involved in the initial programming work. The only testing done during this phase is by the programmer who does some rudimentary testing to insure that the program is at least performing the basic functionality in a way that is expected. At this stage, the developer is free to code and organize his or her files in whatever way he or she chooses.

Beta Phase

By the time the code is promoted to beta phase, more control must be placed on the development to insure uniformity and maintainability. The user circle of this phase grows from the primary developer to the other developers and decision-makers in the project. Most of the testing is done during this time with revisions being made by the developer as the testing reveals bugs or design changes.

Production Phase

In this final phase, the scope of users expands further to include those outside the development group who will interact with the system on a regular basis. The early stages of this phase will also include some testing, mainly to verify that the system will actually run in the production environment. Some small coding may also take place, primarily for process control scripts.

Configuration Management (CM) and Revision Control System (RCS)

Revision control helps manage the various versions of source files that are created as the system evolves. It allows the developers to roll back to prior versions if necessary and to inject comments with each update that describes the changes made.

The tool used in the development of ISH was RCS. This tool was chosen because it is simple to use, freely available, and seemed to match the level of sophistication needed by this system. It was applied only to code in the beta phase.

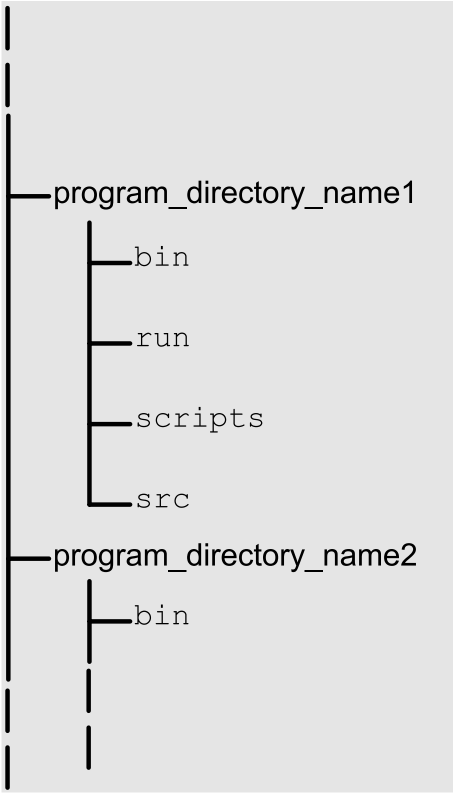


Figure 8 - Beta and Production Directory Tree Structure

Directory Structure and File Locations

There were also special directories used in the beta phase to further organize the programs. The basic structure is shown in Figure 8. Each program has its own directory under which are four subdirectories: bin, run, scripts, and src. The ./bin directory holds the executable. ./run holds any test input, output, log, etc., files that are run against the program. The ./scripts directory holds any scripts used to launch the program or control the compilation, and the ./src directory holds the source files.

Shared Object Code Libraries

Shared object code libraries are repositories of pre-compiled code that can be linked into a program to add functionality. Here are the strongest advantages to using these libraries:

- Faster development with shared code. Once the code has been developed and tested, it can be reused by may different programs with very little additional effort. The code is readily available, providing immediate functionality for the developer.
- Uniformity in style and behavior. Because code can be so easily shared, it provides a uniform way of dealing with common issues that all these programs face.
- Easier maintenance. Because the libraries are actually referred to at runtime, they can be updated without having to recompile the programs that use them.
- Hides complex coding. When there are tasks to be done that require a specific knowledge and/or system paths to develop, these libraries can really make the job easier. For example, accessing Oracle from a C or Fortran program can be a headache because of the lack of clear, platform-specific documentation for the precompiler. Writing the embedded SQL code is fairly easy. Compiling it correctly, though, has proven to be a separate, non-trivial matter. If this functionality can be made available in a shared object library, however, then the task of writing C and Fortran programs that work with Oracle becomes far less daunting.

There are a total of three of these libraries built in conjunction with the ISH system with two of them intended for use as outlined above. These two are discussed in the next two parts below. The third one, the wrapper.c library, is intended for exclusive use with the TD3280 to ISH conversion program.

The Linked List Library

The purpose of this library is to create some code to handle the use of linked lists in a way that can be reused, improved, and expanded as needed to handle the management of linked lists. These lists are designed to be generic enough to handle any type of data. They are doubly-linked, capped off by a NULL pointer on both ends, and consist of a series of "links" chained together.

The Oralib Library

This library was created to make the process of incorporating calls to Oracle from Fortran and C easier. After some difficulties with the Pro*C precompiler, this quickly idea quickly took on merit. Generally, Oralib allows a parent program to submit a SELECT statement and receive the result set back as a link list in the style described above for the linked list library. It also contains functions to submit other SQL commands such as DELETE, INSERT, and UPDATE.

Oralib requires two things to be done in order to run smoothly with a parent program. The first is to include the oralib.h file and the second is to use a make file to correctly link the library in with the parent code at compile time. In order to use Oralib with Fortran, it is important to understand that the library, written in C, depends heavily on the use of pointers to functions for implementing the SELECT and returned linked list result set. There are reports of Fortran 90 being able to handle this directly, but the way it is implemented in this ISH system is through the use of a C wrapper program which bridges between the Fortran code and Oralib.

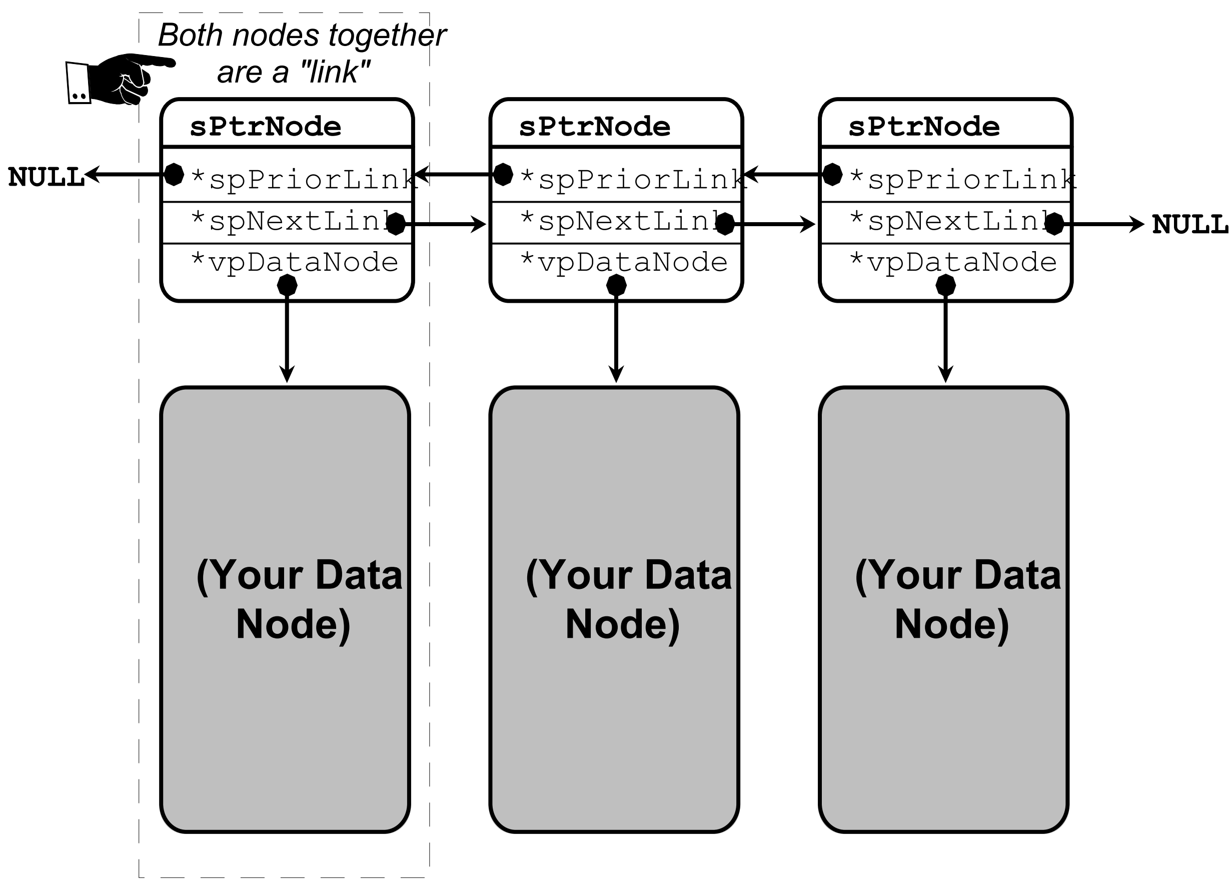


Figure 9 - Linked list structure used by the *linklist* shared object library.

Appendix B - ISH File Format

The summary below provides a quick review of the ISH ASCII data format. For further details, see the ISH format documentation.

Format specifications: A = Alpha characters only; 9 = Numeric characters only; X = Can be either alpha or numeric; + = Numeric sign position.

Position	Field	Len	Format	Description
Control Data Section				
1 - 4	Rec Length	4	9999	Length, in bytes, of the portion of the record beyond position 105.
5 - 10	AWS Id	6	999999	AWS station identifier, padded on the left with zeros.
11 - 15	WBAN Id	5	99999	WBAN station identifier, padded on the left with zeros.
16 - 19	Year	4	9999	Observation year.
20 - 21	Month	2	99	Observation month, padded with zero.
22 - 23	Day	2	99	Observation day, padded with zero.
24 - 25	Hour	2	99	Observation hour in 24-hr. notation, padded with zero.
26 - 27	Minute	2	99	Observation minute, padded with zero.
28 - 28	Data Source Flag	1	9	Data source flag. This field is filled with values when these data are merged with Datsav3.
29 - 34	Lat	6	+99999	Latitude in decimal degrees, scaled to 1000.
35 - 41	Long	7	+999999	Longitude in decimal degrees, scaled to 1000.
42 - 46	Report Type	5	XXXXX	Report Type. Indicates the type of observation reporting method.
47 - 51	Elev	5	+9999	Elevation in meters.
52 - 56	Call Letters	5	99999	Call letter identifier.
57 - 60	QC Name	4	9999	Quality control process name.
Mandatory Data Section				
61 - 63	Wind Dir	3	999	Wind direction.
64 - 64	Wind Dir QC	1	X	Wind direction QC code.
65 - 65	Wind Type	1	X	Wind observation type code.
66 - 69	Wind Speed	4	9999	Wind speed in m/s.
70 - 70	Wind Spd QC	1	X	Wind speed QC code.
71 - 75	Cloud Hgt	5	99999	Ceiling height in meters.
76 - 76	Cld Hgt QC	1	X	Ceiling height QC.
77 - 77	Clg Det Cd	1	X	Ceiling determination code.
78 - 78	Cavok Cd	1	X	CAVOK code.
79 - 84	Visibility	6	999999	Visibility in meters.
85 - 85	Visibility QC	1	X	Visibility QC code.
86 - 86	Vis Var Cd	1	9	Visibility variability code.
87 - 87	Vis Var QC	1	9	Visibility variability QC code.
88 - 92	Temp	5	+9999	Air temperature in degrees Celsius.
93 - 93	Temp QC	1	X	Air temperature QC code.

94 - 98	Dew Pt Temp	5	+9999	Dew point temperature in degrees Celsius.
99 - 99	Dew Pt QC	1	X	Dew point QC code.
100-104	Sea Lev Pres	5	99999	Sea level pressure in mb.
105-105	Sea Lev QC	1	X	Sea level pressure QC code.
Additional Data Section (Variable Length)				
106-108	Add Ind	3	AAA	Additional data section indicator. Always “ADD”.
AA1-4	• Liquid Precip Id	3	AA9	Liquid precipitation identifier. This is set to “AAx” where x = 1, 2, 3, or 4 depending on how many occurrences appear in the data.
	• Period Qty	2	99	Period quantity.
	• Depth Dim	4	9999	Depth dimension, amount in mm.
	• Cond Cd	1	9	Condition code.
	• Quality Cd	1	9	Quality code.
AC1	• Prcp History Id	3	AA9	Precipitation observation history identifier.
	• Duration Cd	1	9	Duration code.
	• Characteristic Cd	1	X	Characteristic code.
	• Dur Char QC	1	9	Duration and characteristic quality code.
AG1	• Prcp Bogus Id	3	AA9	Precipitation bogus observation identifier.
	• Discrepance Cd	1	9	Discrepancy code.
	• Est Equiv Dim	3	999	Estimated water equivalency dimension..
AJ1	• Snow Dep Id	3	AA9	Snow depth identifier.
	• Snow Dep Dim	4	9999	Dimension, depth in cm.
	• Cond Cd	1	9	Condition code.
	• Snow Dep QC	1	9	Snow depth quality code.
	• Equ Wtr Dep	6	999999	Equivalent water depth dimension, in mm.
	• Equ Wtr Cond Cd	1	9	Equivalent water condition code.
	• Equ Wtr Cond QC	1	9	Equivalent water condition quality code.
HL1	• HailId	3	AA9	Hail identifier.
	• HailSize	3	999	Hail size.
	• HailSize QC	1	9	Hail size quality code.
AL1-4	• Snw Accum Id	3	AA9	Snow accumulation identifier. This is set to “ALx” where x = 1, 2, 3, or 4 depending on how many occurrences appear in the data.
	• Period Qty	2	99	Period quantity.
	• Depth Dim	3	999	Depth dimension, in cm.
	• Cond Cd	1	9	Condition code.
	• SnwAccumQC	1	1	Snow accumulation quality code.
MW1-9	• Pr Wth Id	3	AA9	Present weather identifier. This is set to “MWx” where x = 1, 2, 3, 4, etc depending on how many occurrences appear in the data.
	• Pr Wth	2	99	Manual atmospheric condition code.
	• Pr Wth QC	1	X	Manual atmospheric condition quality code.

MV1-7	• Pr Wth Vc Id	3	AA9	Present weather in the vicinity identifier. This is set to "MVx" where x = 1, 2, 3, 4, 5, 6, or 7 depending on how many occurrences appear in the data.
	• Pr Wth Vc Cd	2	99	Atmospheric condition code.
	• Pr Wth Vc QC	1	X	Atmospheric condition quality code.
AW1	• Auto Pr Wth Id	3	AA9	Automated present weather in the vicinity identifier.
	• Auto Cond Cd	2	99	Automated atmospheric condition code.
	• Auto Cond QC	1	X	Automated atmospheric condition quality code.
AY1-2	• Man Occ Id	3	AA9	Past weather observation - Manual occurrence identifier.
	• Man Atmos Cd	1	9	Manual atmospheric condition code.
	• Man Atmos QC	1	9	Manual atmospheric condition quality code.
	• Man Period Qty	2	99	Period quantity.
	• Man Period QC	1	9	Period quality code.
AZ1-2	• Auto Occ Id	3	AA9	Past weather observation - Automated occurrence identifier.
	• Auto Atmos Cd	1	9	Automated atmospheric condition code.
	• Auto Atmos QC	1	9	Automated atmospheric condition quality code.
	• Auto Period Qty	2	99	Period quantity.
	• Auto Period QC	1	9	Period quality code.
ED1	• Runway Vis Id	3	AA9	Runway visual range observation identifier.
	• Runway Dir Angle	2	99	Direction angle.
	• Runway Des Cd	1	A	Runway designator code.
	• Run Vis Dim	4	9999	Visibility dimension in meters.
	• RunwayObsOC	1	9	Runway visual range observation quality code.
GA1-6	• Sky Cov Layer Id	3	AA9	Sky cover layer identifier.
	• Sky Cov Cd	2	99	Coverage code.
	• Sky Cov QC	1	9	Coverage quality code.
	• Base Hgt Dim	6	+99999	Base height dimension in meters.
	• Base Hgt Dim QC	1	9	Base height dimension quality code.
	• Cld Type Cd	2	99	Cloud type code.
	• Cld Type QC	1	9	Cloud type quality code.
GF1	• Sky Cond Id	3	AA9	Sky condition observation identifier.
	• Tot Cov Cd	2	99	Total coverage code.
	• Tot Op Cd	2	99	Total opaque code.
	• Sky Cond QC	1	X	Sky condition QC code.
	• Tot Low Cd	2	99	Total lowest cloud cover code.
	• Tot Low Qc	1	9	Total lowest cloud cover quality code.
	• Low Genus Cd	2	99	Low cloud genus code.
	• Low Genus Qc	1	9	Low cloud genus quality code.
	• Low Cld Base	5	99999	Lowest cloud base height dimension in meters.

	• Low Cld Base Qc	1	9	Lowest cloud base height dimension quality code.
	• Mid Cld	2	99	Mid cloud genus code.
	• Mid Cld Qc	1	9	Mid cloud genus quality code.
	• Hi Cld	2	99	High cloud genus code.
	• Hi Cld Qc	1	9	High cloud genus quality code.
GD1-6	• Sky Sum Id	3	AA9	Sky cover summation identifier.
	• Cov Code 1	1	9	Coverage code 1.
	• Cov Code 2	2	99	Coverage code 2.
	• Cov QC	1	X	Coverage QC code.
	• Hgt Dim	6	+99999	Height dimension in meters.
	• Hgt Dim QC	1	X	Height dimension QC code.
	• Char Code	1	9	Characteristic code.
GG1-6	• Bel Stat Cld Id	3	AA9	Below station cloud layer identifier.
	• Bel Stat Cov Cd	2	99	Coverage code.
	• Bel Stat Cov QC	1	9	Coverage quality code.
	• Top Hgt Dim	5	99999	Top height dimension in meters.
	• Top Hgt Dim QC	1	9	Top height dimension quality code.
	• Type Cd	2	99	Type code.
	• Type QC	1	9	Type quality code.
	• Top Code	2	99	Top code.
	• Top QC	1	9	Top quality code.
GJ1	• Sun Obs Id	3	AA9	Sunshine observation identifier.
	• Sun Dur Qty	4	9999	Sunshine duration quantity.
	• Sun Dur QC	1	9	Sunshine duration quality code.
IA1	• Grd Sfc Obs Id	3	AA9	Ground surface observation identifier.
	• Grd Sfc Obs Cd	2	99	Ground surface observation code.
	• Grd Sfc Obs Qc	1	9	Ground surface observation quality code.
IA2	• GSO Min Tmp Id	3	AA9	Ground surface observation minimum temperature identifier.
	• GSO Min Tmp Qty	3	999	Minimum temperature period quantity.
	• GSO Min Tmp	5	+9999	Minimum temperature in Celsius.
	• GSO Min Tmp QC	1	9	Minimum temperature quality code.
KA1-2	• Xair Temp Id	3	AA9	Extreme air temperature identifier.
	• Period Qty	3	999	Period quantity.
	• Xair Temp Cd	1	X	Extreme air temperature code for minimum or maximum.
	• Xair Temp	5	+9999	Extreme air temperature in Celsius.
	• Xair Temp QC	1	9	Extreme air temperature quality code.
MA1	• Press Id	3	AA9	Atmospheric pressure identifier.
	• Altimeter	5	99999	Altimeter setting in mb.

	• Alt QC	1	9	Altimeter quality code.
	• Stn Press	5	99999	Station pressure in mb.
	• Press QC	1	9	Pressure quality code.
MD1	• Pres Change Id	3	AA9	Atmospheric pressure change identifier.
	• Tendency Cd	1	9	Atmospheric pressure tendency code.
	• Tendency QC	1	9	Atmospheric pressure tendency quality code.
	• Three Hr Qty	3	999	Three hour quantity in mb.
	• Three Hr QC	1	9	Three hour quality code.
	• Twty Four Hr Qty	4	+999	Twenty four hour quantity in mb.
	• Twty Four Hr QC	1	9	Twenty four hour quality code.
ME1	• Geo Hgt Id	3	AA9	Geopotential height isobaric level identifier.
	• Geo Hgt Id Cd	1	9	Geopotential height isobaric level code.
	• Geo Hgt Dim	4	9999	Height dimension in meters.
	• Geo Hgt Dim QC	1	9	Height dimension quality code.
OA1-2	• Sup Wnd Id	3	AA9	Supplementary wind observation identifier.
	• Sup Wnd Typ Cd	1	9	Supplementary wind type code.
	• Sup Wnd Per Qty	2	99	Supplementary wind period quantity.
	• Sup Wnd Spd	4	9999	Speed in m/s.
	• Sup Wnd Spd Qc	1	9	Supplementary wind speed rate quality code.
OC1	• Wnd Gust Id	3	AA9	Wind gust observation identifier.
	• Wnd Gust Spd	4	9999	Speed in m/s.
	• Wnd Gust QC	1	9	Wind gust observation quantity code.
SA1	• Sea Sfc Tmp Id	3	AA9	Sea surface temperature observation identifier.
	• Sea Sfc Tmp	4	9999	Sea surface temperature in Celsius.
	• Sea Sfc Tmp QC	1	9	Sea surface temperature quantity code.
UA1	• Wav Meas Id	3	AA9	Wave measurement identifier.
	• Wav Meas Meth	1	X	Method code.
	• Wav Per Qty	2	99	Wave period quantity.
	• Wav Hgt Dim	3	999	Wave height dimension.
	• Wav Meas QC	1	9	Wave measurement quality code.
	• Wav Sea State Cd	2	99	Sea state code.
	• Wav Sea State QC	1	9	Sea state quality code.
UG1	• Wav Prim Id	3	AA9	Wave measurement primary swell identifier.
	• Wav Prim Qty	2	99	Period quantity.
	• Wav Prim Hgt Dim	3	999	Height dimension.

	• Wav Prim Dir Angl	3	999	Direction angle.
	• Wav Prim QC	1	9	Primary swell quality code.
UG2	• Wav Sec Id	3	AA9	Wave measurement secondary swell identifier.
	• Wav Sec Qty	2	99	Period quantity.
	• Wav Sed Hgt Dim	3	999	Height dimension.
	• Wav Sec Dir Angl	3	999	Direction angle.
	• Wav Sec QC	1	9	Secondary swell quality code.
WA1	• Plat Ice AccId	3	AA9	Platform ice accretion identifier.
	• Plat Ice Src Cd	1	9	Source code.
	• Plat Ice Thk Dim	3	999	Thickness dimension.
	• Plat Ice Tend Cd	1	9	Tendency code.
	• Plat Ice QC	1	9	Platform ice accretion quality code.
WG1	• Wtr Ice Hist Id	3	AA9	Water surface ice historical observation identifier.
	• Ice Edge Brng Cd	2	99	Ocean ice observation edge bearing code.
	• Edge Dist Dim	2	99	Edge distance dimension.
	• Edge Orient Cd	2	99	Edge orientation code.
	• Form Type Cd	2	99	Formation type code.
	• Nav Effect Cd	2	99	Navigation effect code.
	• Wtr Icd Hist QC	1	9	Water surface ice historical observation quality code.
WD1	• Wtr Ice Id	3	AA9	Water surface ice observation identifier.
	• Ice Edge Brng Cd	2	99	Ocean ice observation edge bearing code.
	• Uni Con Rate	3	999	Uniform concentration rate.
	• Non Uni Con Rate	2	99	Non-uniform concentration rate.
	• Ship Rel Pos Cd	1	9	Ship relative position code.
	• Ship Penet Cd	1	9	Ship penetrability code.
	• Ice Trend Cd	1	9	Ice trend code.
	• Dev Cd	2	99	Development code.
	• Gr Ber Bit Pres Cd	1	9	Growler-bergy-bit presence code.
	• Gr Ber Bit Qty	3	999	Growler-bergy-bit quantity.
	• Iceberg Qty	3	999	Iceberg quantity.
	• Wtr Icd QC	1	9	Water surface ice observation quality code.
Remarks Section (Variable Length)				
REM	• Rem Sect Id	3	AAA	Remarks section identifier.
SYN	• Syn Rem Id	3	AAA	Synoptic remark identifier.
	• Rem Len Qty	3	999	Remark length quantity.
	• Rem Text	250	X (250)	Remark text.

AWY	• Awy Rem Id	3	AAA	Airways remark identifier.
	• Rem Len Qty	3	999	Remark length quantity.
	• Rem Text	250	X (250)	Remark text.
MET	• Met Rem Id	3	AAA	METAR remark identifier.
	• Rem Len Qty	3	999	Remark length quantity.
	• Rem Text	250	X (250)	Remark text.
EQD (Qxx/Nxx) Section (Variable Length)				
EQD	Eqd Sect Id	3	AAA	Element quality data section indicator. This marks the beginning of this section which saves the data and flag values of data portions that were “erroneous.” Each of the erroneous values are stored in their own portion as described below. NOTE: The converted TD3280 data will use “N” for the counter whereas TD9956 and ISH use “Q”. For example, “N01” indicates an EQD section that comes from TD3280 data and should be parsed differently from a TD9956 EQD section.
Qxx	• Qxx Id	3	A99	EQD portion counter for Datsav3 data where ‘xx’ ranges from 01 to 99.
	• Orig Value	6	999999	Original data value.
	• Reason Cd	1	9	Reason code.
	• Parameter Cd	6	XXXXXX	Parameter code.
Nxx	• Nxx Id	3	A99	EQD portion counter for TD3280 data where ‘xx’ ranges from 01 to 99.
	• Orig Value	6	999999	Original data value.
	• Units Cd	1	X	Units code.
	• Elem Cd	4	XXXX	TD3280 element code.
	• Meas Flag	1	X	TD3280 measurement flag.
	• Quality Flag	1	X	TD3280 data quality flag.
QNN Section (Variable Length)				
QNN	Qnn Id	3	AAA	QNN indicator. This marks the beginning of this section which preserves the TD3280 source codes and data flags for each data portion found in the TD3280 file. NOTE: There are a maximum of 35 occurrences for these QNN entries.
	• Qnn Elem Cd	1	X	TD3280 element code.
	• Qnn Src Cd1	1	X	TD3280 source code 1.
	• Qnn Src Cd2	1	X	TD3280 source code 1.
	• Qnn Meas Flag	1	X	TD3280 measurement flag.
	• Qnn Quality Flag	1	X	TD3280 quality flag.
	• Qnn Data Val	6	+99999	Original TD3280 data value.

Appendix C - ISH Oracle Table Structures

The summary below provides a quick review of the ISH relational table formats. It is not intended as a detailed review of the table structures.
(I = indexed field, P = primary key)

Climate Data Online (CDO) ISH Tables (used in building and providing web access to the data):

		Column Name	Data Type
ISH_Country_Alias Table			
I	1)	CDO_Abbrev	Varchar2(2)
I	2)	AWS_Country_Id	Varchar2(2)
	3)	Source	Varchar2(10)

ISH_Hist Table		
1)	AWS_Id	Varchar2(6)
2)	WBAN_Id	Varchar2(5)
3)	CoOp_Id	Varchar2
4)	Begin_Date	Varchar2
5)	End_Date	Varchar2
6)	Stn_Name	Varchar2
7)	SStn_Name	Varchar2
8)	State_Prov_Id	Varchar2
9)	LatDegDec	Varchar2
10)	LongDegDec	Varchar2
11)	ElevMetersTenth	Varchar2
12)	Call_Sign	Varchar2
13)	ICAO_Id	Varchar2
14)	AWS_Country_Id	Varchar2
15)	WMO_No	Varchar2
16)	Country	Varchar2
17)	County	Varchar2
18)	Climate_Dv	Varchar2
19)	DataCnt	Number
20)	FIPS_Country_Id	Varchar2

Format Properties Tables for CDO:

Dataset_Field Table			
P	1)	DatasetId	Number
P	2)	Field_Id	Number
	3)	Source_Table	Varchar2
	4)	Source_Col	Varchar2
	5)	Where_Cond	Varchar2

```
DsElem Table
   1) DatasetId         Number
 P 2) ElemId            Number
   3) ParElemId         Number
   4) Abbrev            Varchar2
   5) Name              Varchar2
   6) Units             Varchar2
   7) Scale             Number
   8) Len               Number
   9) ElemPosition      Number
   10) Selectable       Number
   11) ParAbbv          Varchar2

Field Table
 P 1) Field_Id          Number
   2) Data_Type_Id      Number
   3) Header            Varchar2
   4) Name              Varchar2
   5) AccessType        Varchar2
   6) Description       Varchar2
   7) Table_of_Values   CLOB

Field_Data_Type Table
 P 1) Data_Type_Id      Number
   2) Data_Type_Desc    Varchar2

Field_Role Table
 P 1) Role_Id           Number
 P 2) Field_Id          Number

Format Table
 P 1) Format_Id         Number
   2) Name              Varchar2

Format_Field Table
P 1) Format_Id          Number
P 2) Field_Id           Number
  3) Field_Pos          Number
  4) Header             Varchar2
  5) Name               Varchar2
  6) Length             Number
  7) Scale              Number
  8) Physical           Number
  9) Fixed              Number
  10) Default_Val       Varchar2
  11) Sign_Type         Number
  12) Padding_Char      Character
  13) Sort_Order        Number
  14) Sort_Direction    Character
  15) AccessType        Varchar2
  16) Parent_Field_Id   Number
  17) Field_Unit_Id     Number
```

```
Format_Field_Unit Table
 P 1) Field_Unit_Id     Number
   2) Field_Unit_Desc   Varchar2

Role Table
 P 1) Role_Id           Number
   2) Header            Varchar2
   3) Name              Varchar2
   4) Parent_Role_Id    Number
```

ISH Data Tables for CDO:

```
ISH_ElemFlds Table
 P 1) ObsKeyId          Number
 P 2) Elem_Id           Varchar2
   3) Year              Number
   4) ElemFld1          Varchar2
   5) ElemFld2          Varchar2
   6) ElemFld3          Varchar2
   7) ElemFld4          Varchar2
   8) ElemFld5          Varchar2
   9) ElemFld6          Varchar2
   10) ElemFld7         Varchar2
   11) ElemFld8         Varchar2
   12) ElemFld9         Varchar2
   13) ElemFld10        Varchar2
   14) ElemFld11        Varchar2
   15) ElemFld12        Varchar2
   16) ElemFld13        Varchar2

ISH_RemEqdQnn Table
 P 1) ObsKeyId          Number
   2) Year              Number
   3) RemString         Varchar2
   4) EqdString         Varchar2
   5) QnnString         Varchar2

ISH_StnDate Table
 P 1) ObsKeyId          Number
 I 2) AwsBan_Id         Varchar2
 I 3) AWS_Id            Varchar2
 I 4) WBAN_Id           Varchar2
 I 5) Year              Number
 I 6) Month             Number
 I 7) Day               Number
 I 8) Hour              Number
 I 9) Minute            Number
 I 10) DataSrcFlag      Character
 I 11) RptType          Varchar2
   12) LatDegDec        Varchar2
   13) LongDegDec       Varchar2
   14) ElevMeters       Varchar2
   15) Call_Id          Varchar2
   16) QcName           Varchar2
```

```
Temp_ElemFlds Table
 P 1) ObsKeyId           Number
 P 2) Elem_Id            Varchar2
   3) Year               Number
   4) ElemFld1           Varchar2
   5) ElemFld2           Varchar2
   6) ElemFld3           Varchar2
   7) ElemFld4           Varchar2
   8) ElemFld5           Varchar2
   9) ElemFld6           Varchar2
   10) ElemFld7          Varchar2
   11) ElemFld8          Varchar2
   12) ElemFld9          Varchar2
   13) ElemFld10         Varchar2
   14) ElemFld11         Varchar2
   15) ElemFld12         Varchar2
   16) ElemFld13         Varchar2

Temp_RemEqdQnn Table
 P 1) ObsKeyId           Number
   2) Year               Number
   3) RemString          Varchar2
   4) EqdString          Varchar2
   5) QnnString          Varchar2

Temp_StnDate Table
 P 1) ObsKeyId           Number
   2) AwsBan_Id          Varchar2
 I 3) AWS_Id             Varchar2
 I 4) WBAN_Id            Varchar2
 I 5) Year               Number
 I 6) Month              Number
 I 7) Day                Number
 I 8) Hour               Number
 I 9) Minute             Number
 I 10) DataSrcFlag       Character
 I 11) RptType           Varchar2
   12) LatDegDec         Varchar2
   13) LongDegDec        Varchar2
   14) ElevMeters        Varchar2
   15) Call_Id           Varchar2
   16) QcName            Varchar2
   17) StatsDone         Number
```

ISH Database Generation Tables (used in building and verifying the ISH database):

Inventory Tables:

I_3280

P	1) WBAN_Id	Varchar2
P	2) Year	Number
	3) M01	Number
	4) M02	Number
	5) M03	Number
	6) M04	Number
	7) M05	Number
	8) M06	Number
	9) M07	Number
	10) M08	Number
	11) M09	Number
	12) M10	Number
	13) M11	Number
	14) M12	Number

I_9956 Table

P	1) AWS_Id	Varchar2
P	2) Year	Number
	3) M01	Number
	4) M02	Number
	5) M03	Number
	6) M04	Number
	7) M05	Number
	8) M06	Number
	9) M07	Number
	10) M08	Number
	11) M09	Number
	12) M10	Number
	13) M11	Number
	14) M12	Number

I_Dasv Table

P	1) AWS_Id	Varchar2
P	2) Year	Number
	3) M01	Number
	4) M02	Number
	5) M03	Number
	6) M04	Number
	7) M05	Number
	8) M06	Number
	9) M07	Number
	10) M08	Number
	11) M09	Number
	12) M10	Number
	13) M11	Number
	14) M12	Number

```
I_Diff Table
 P 1) AWS_Id             Varchar2
 P 2) WBAN_Id            Varchar2
 P 3) Year               Number
   4) M01                Number
   5) M02                Number
   6) M03                Number
   7) M04                Number
   8) M05                Number
   9) M06                Number
   10) M07               Number
   11) M08               Number
   12) M09               Number
   13) M10               Number
   14) M11               Number
   15) M12               Number

I_ISH Table
 P 1) AWS_Id             Varchar2
 P 2) WBAN_Id            Varchar2
 P 3) Year               Number
   4) M01                Number
   5) M02                Number
   6) M03                Number
   7) M04                Number
   8) M05                Number
   9) M06                Number
   10) M07               Number
   11) M08               Number
   12) M09               Number
   13) M10               Number
   14) M11               Number
   15) M12               Number

I_IshElem Table
 P 1) AWS_Id             Varchar2
 P 2) WBAN_Id            Varchar2
 P 3) Year               Number
 P 4) Month_Id           Number
 P 5) Elem_Id            Varchar2
   6) Elem_Count         Number
```

Reference Tables:

```
Data_Source Table
   1) DataSrcCd          Character
   2) DataSrcDesc        Varchar2
```

```
ISH_Load_Stats Table
 I 1) AWS_Id            Varchar2
 I 2) WBAN_Id           Varchar2
 I 3) Year              Number
 I 4) Month             Number
 I 5) Day               Number
 I 6) DataSrcFlag       Character
 I 7) RptType           Varchar2
   8) Pre_S_Cnt         Number
   9) Pre_E_Cnt         Number
   10) Pre_R_Cnt        Number
   11) Post_S_Cnt       Number
   12) Post_E_Cnt       Number
   13) Post_R_Cnt       Number

ISH_QC Table
 P 1) AWS_Id            Varchar2
 P 2) WBAN_Id           Varchar2
 P 3) Year              Number
 P 4) Algorithm_Num     Number
   5) Errors_Corrected  Number
   6) Suspects          Number
```

Station History Tables:

```
Adjust Table
   1) AWS_Id            Varchar2
   2) WBAN_Id           Varchar2
   3) CoOp_Id           Varchar2
   4) Begin_Date        Varchar2
   5) End_Date          Varchar2
   6) Override_Flag     Number

Bypass Table
   1) WBAN_Id           Varchar2

IshPOR1 Table
   1) AWS_Id            Varchar2
   2) WBAN_Id           Varchar2
   3) CoOp_Id           Varchar2
   4) Begin_Date        Varchar2
   5) End_Date          Varchar2
   6) LatDegDec         Varchar2
   7) LongDegDec        Varchar2
   8) ElevMeters        Varchar2
   9) Call_Sign         Varchar2
   10) Updated          Number

IshPOR2 Table
   1) AWS_Id            Varchar2
   2) WBAN_Id           Varchar2
   3) Begin_Date        Varchar2
   4) End_Date          Varchar2
   5) End_Date2         Number
```

MSC_Stns Table

1) AWS_Id	Varchar2
2) Stn_Name	Varchar2
3) S_StatPr	Varchar2
4) FIPS_Cd	Varchar2
5) LatDegDec	Varchar2
6) LongDegDec	Varchar2
7) ElevMeters	Varchar2
8) ICAOId	Varchar2
9) Call2	Varchar2
10) Updated	Number

StnHist Table

1) Stn_Id_Num	Number
2) Stn_Name	Varchar2
3) SStn_Name	Varchar2
4) City	Varchar2
5) S_StatPr	Varchar2
6) Begin_Date	Varchar2
7) End_Date	Varchar2
8) Wban_No	Varchar2
9) CoOp_No	Varchar2
10) Stn_Net	Varchar2
11) WMO_No	Varchar2
12) Call_Sign	Varchar2
13) General	Varchar2
14) Latitude	Number
15) Longitude	Number
16) Elevation	Number
17) Alias_Flag	Varchar2
18) Country	Varchar2
19) Type	Varchar2
20) Name2	Varchar2
21) ASOS	Varchar2
22) NEXRAD	Varchar2
23) County	Varchar2
24) ICS	Varchar2
25) AFID	Varchar2
26) Climate_Dv	Varchar2
27) NWS_Loc_Id	Varchar2
28) ICAO_Id	Varchar2
29) Time_Zone	Varchar2
30) Lat_Dir	Varchar2
31) Lat_Deg	Number
33) Lat_Sec	Number
34) Lon_Dir	Varchar2
32) Lat_Min	Number
35) Lon_Deg	Number
36) Lon_Min	Number
37) Lon_Sec	Number
38) LatLon_PRC	Varchar2
39) Elevation1	Varchar2
40) Elevation2	Varchar2
41) Elev_Type	Varchar2
42) AWSId	Varchar2

43) ToStnHist Varchar2

TimeConv Table

1)	WBAN_Id	Varchar2
2)	CoOp_Id	Varchar2
3)	Station_Name	Varchar2
4)	Country_Name	Varchar2
5)	State_Code	Varchar2
6)	Conv_Factor	Varchar2
7)	Begin_Date	Varchar2
8)	End_Date	Varchar2
9)	Conv_Factor_Test	Varchar2