WILDLAND FIRE AND WEATHER INFORMATION DATA WAREHOUSE

Delvin R. Bunton[1]

ABSTRACT.—The USDA Forest Service stores fire occurrence data in a relational database for planning, analysis, and other purposes. Weather observations are stored in the same database for all five federal land management agencies and some state wildland agencies. Ready access to fire occurrence and weather data enables managers, planners, and operations personnel to make informed decisions based upon the past and projections into the future. This presentation discusses the design philosophy behind the database, issues dealing with the care and feeding of the database and data stored there, and how data helps field users.

DATA WAREHOUSE CONCEPTS

Corey and Abbey (1997) define a data warehouse as "a collection of corporate information derived directly from operational systems and some external sources." This definition broadly fits the National Interagency Fire Management Integrated Database (NIFMID) discussed in this paper. Data warehouses support business decisions, not business operations. When business managers, whether in a corporation or a governmental agency, ask "What if?" questions, the data to answer the business questions will come from a data warehouse of one kind or another.

Operational databases are crucial to many businesses and government agencies. Operational databases:

1. Typically handle one record at a time (such as an invoice, weather observation, or fire).
2. Often have predictable busy and slack times during a day, week, or month.
3. Are organized to meet specific business needs.
4. Typically change slowly (if at all) over time, although some changes are abrupt and dramatic.

Data warehouses, in contrast, exhibit different characteristics. They:

1. Typically handle many records at once (all fires in 1994, weather stations in California, etc.)
2. Usage is either high or very low.
3. Are designed to cross business lines.
4. Provide ways to link data that may not be directly related.
5. Work with longer time periods to find trends.
6. Change structure over time to meet new and changing business needs.

Before computers, data warehouses existed as paper files in boxes and cabinets. Many businesses and government agencies still use this data storage method, and it works quite well for many applications. In their analysis of fire occurrence in California national forests, Show and Kotok (1923) extracted data by hand from 10,449 individual fire reports. They summarized data many different ways that required numerous passes through the paper reports. Their findings affected fire planning for several decades, and

[1] Delvin R. Bunton, USDA Forest Service, Washington Office, Fire and Aviation Management Staff, 3833 S. Development Ave., Boise, ID, USA 83705

some findings are still applicable today. One interesting sidelight is the information Show and Kotok used is very similar to what Forest Service fire managers today record about fires (Donoghue, 1982).

Barrows (1951) used computer cards for each fire (approximately 36,000) when he analyzed fire occurrence in the northern Rocky Mountains (primarily Montana and northern Idaho). Barrow's ability to quickly search for common denominators among the reports allowed him to gain insights that were hidden from earlier researchers. He presented his findings in many detailed tables and graphs that clearly show characteristics of the fires studied. Many of his findings are still applicable today.

Goldberg and Vijayan (1996) reported that managers at Victoria's Secret (a women's intimate apparel retail chain) found the company was spending too much time and money collecting information and too little time analyzing it. When they did analyze their data, they found some surprises. Stores were regularly stocked based on a mathematical average. When stocking levels at each store were based on local customer buying patterns, profits rose. The company put items on sale at many stores simultaneously even when the customers at some stores were willing to purchase the items at full price. Company management changed discounting practices, which again increased profits.

The application of modern data storage and analysis tools enables managers to find insights into relationships masked by data volume, or other trends. Data base tools allow users to quickly summarize data in ways that were possible in past times, but only with great effort. The challenge is to organize the data so those new relationships become visible, and then demonstrate to management that the findings have relevance to their decision-making.

REAL WORLD DATA STORAGE

Early managers submitted fire reports on paper forms through the organizational hierarchy. Copies were usually kept at the Forest and Regional offices. Beginning in 1940, fire reports were keypunched onto computer cards at either the Regional or National offices and stored for analysis. Some Forest Service paper fire reports from before 1970 still exist, but most were regrettably destroyed.

Weather observations were typically stored in paper form at local offices. Some data from before 1972 when AFFIRMS (Administrative Forest Fire Information Retrieval and Management System) began operations still exists. Local offices entered weather observations into AFFIRMS with a computer terminal. Weather forecasters retrieved the data and issued forecasts, which were then sent back to local fire management offices via AFFIRMS.

From 1974 to 1992, the Forest Service stored both fire occurrence and daily weather observations on computer tape at a departmental computer center (Furman, 1975; Yancik and Roussopoulos, 1982). Users ran a program to find out which tape had the information they wanted, and then ran another program to extract the data. Only one weather data format was defined, and two fire formats (neither of which gave the full data record) for the general user. Users who had special needs generally did not have their data needs met unless they either wrote the extract programs themselves, or found someone who could write the program for them.

Since fire data was added to the tapes only once per year, usually in March or April, additions and corrections were virtually impossible. Many weather observations and some fire reports were never added to the tape files, and so were missing from any analysis. Recent analysis shows most of the missing fires were very large fires that were still active (not declared out) until well after the annual report for the year was completed. Missing weather observations were most often on the first day of a fire since the weather observer was also a firefighter.

In 1986, the fire report system was converted to use a database for storage. This improved data access to fires that occurred in 1986 and later years. Field offices were finally able to correct recent fire records, and generate custom reports as needed. Units could not correct fires from prior to 1986. Fires were copied annually to the tape files for archival and retrieval purposes.

BUILDING THE FIRE AND WEATHER DATA WAREHOUSE

In 1992, the Department of Agriculture began consolidation of computer services; and eventually closed the computer center that housed the tape files. The Forest Service decided to build a database to contain both the fire and weather data using a relational data base to provide easier access to the data. The fire and weather portions of the database were built by separate development teams to speed the transition using the philosophy described below. The overall structure was named the National Interagency Fire Management Integrated Database (NIFMID). The NIFMID was originally designed around four key principles:

1. Operational weather data would be managed by the Weather Information Management System (WIMS) that was under development at the time. WIMS is an interactive system currently used by all five federal land management agencies (Forest Service, Bureau of Land Management, National Park Service, Bureau of Indian Affairs, and Fish and Wildlife Service), many states, and some local wildland fire protection agencies.
2. Forest Service fires would continue to be managed by FireStat (USDA Forest Service, 1996) to minimize field disruption. Local Forest Service offices enter fire data on their local computers and then transmit the fire data to NIFMID where the data is loaded and immediately available for analysis purposes. Department of Interior fires are entered and stored on a different computer system.
3. Historical weather and fire data would be readily available to those who need the data.
4. Users could download data for local analysis, or obtain summary reports from both historical and operational databases.

The NIFMID architecture includes over 50 tables that logically fall into six groups. Where it makes sense, codes (such as aspect, state, county, etc.) are shared by both fire and weather.

1. Fire occurrence data (information on each fire, including basic information, events, acres burned by ownership, resources used on the fire).
2. Fire code reference tables (fuel model codes, organizations, vegetation codes).
3. Current weather and weather-related tables (up to 24 observations per station per day, forecasts, station location data).
4. Historical weather tables (one observation per station per day).
5. Weather reference tables (fuel models, aspect, wind direction).
6. Common reference tables (states, counties, principle meridians).

Since we chose to blend two operational systems (weather and fire reporting) within a data warehouse, there were two problems to face. The first was to migrate historical data into the data warehouse and provide easy user access to the data. The second was to create software to capture and validate new data such as fire reports, weather observations, and forecasts.

LOADING THE WAREHOUSE WITH HISTORICAL DATA

An empty warehouse does not attract many customers (unless you are selling warehouses). An important value from the warehouse concept was the ability to easily compare past years against current years to estimate workload and possible trends.

Before NIFMID was created, the Forest Service had automated fire data records from 1970-1992 stored on tape. There were also eight years of data (1986-1992) in a database (the data records on tape for 1986 and later were derived from that database). The data was extracted from tape and loaded into flat files on the target computer. A custom loader program read the data from the flat files and populated the several tables that contain fire data. Some of the initial loads failed when non-standard data was discovered in some fields. After those problems were repaired the data loaded properly. Approximately 280,000 fires were loaded using this procedure.

The weather loading process was similar to the fire process. Data was extracted from tapes and loaded using Oracle[2] SQL*Loader. Approximately 15 million weather records were loaded by this method.

Once the historical fire and weather data was loaded, the data was indexed to speed retrieval queries. Fires were indexed by report unit, year and fire number. Weather observation were indexed by station number and observation date. Data properly indexed can be retrieved very quickly, but queries that scan for unindexed values must read the entire table to find the target records.

Numerous data problems were found as users worked with the data. Errors will undoubtedly surface for several more years. Data shifts in the old data (everything one column off) are the hardest to find and fix. When users find an error, they should have the ability to fix the error. This feature is not entirely automated, although errors do get corrected. Some records that appear to be errors (for example, a fire over-wintering under snow) may be correct.

Some of the more interesting errors and anomalies, almost all from pre-database days, included the following:

1. Some large fires were never input to the fire occurrence database, so summaries did not add to the correct number of fires nor acres burned. Missing fires were added to the database as we discovered their absence.
2. Some fires were entered two or more times and not removed from the database. This was a known but difficult to fix problem until the fires were in a database. Most of these duplicates were deleted.
3. Fire report rules at the time required each Forest that had part of a large fire to enter their portion of the acres burned (fairly rare event). In one year when numerous fires burned on several Forests, the error in double counted acres exceeded the average annual burned acres. This was quickly corrected.
4. About 1% of the fires located by latitude and longitude plotted outside the state of origin. Most of these problems were quickly found and fixed.
5. The most likely day a weather observation was missing from the observation file is the first day of a large fire on the management unit.
6. Weather observations were stored with dry bulb temperature and relative humidity. However, some observations prior to 1973 were entered with either wet bulb or dew point in the relative humidity field. The detection is more complex than the repairs, which are underway now.
7. Trace precipitation under AFFIRMS use “T” to note that there was less than 0.01” of precipitation. All “T” values were changed to 0.005” precipitation so mathematical totals became possible.
8. The weather loader program stored zero percent relative humidity for days with no observed relative humidity. When users analyzed the data, the analysis programs showed many very dry days. Since zero humidity is very unlikely, all those values were changed to null--meaning the value is missing.

FEEDING THE WAREHOUSE (OPERATIONS)

Properly feeding a data base is critical for success. Clean data is essential; there must be methods in place to enforce data integrity. Direct input without database procedures almost guarantees errors will creep into the data.

The FireStat application operates at each Forest and some District fire offices. Fire managers enter data on up to six screens about each fire. Some fields are range checked at input time. When the manager is ready to transmit the fire to NIFMID, each fire is rigorously checked again using national and local data ranges. Values that are acceptable nationally may not be within valid ranges for a specific state. For example, fires above 11,245 feet elevation would be accepted in Colorado but not in Oregon. Fires with

[2] The use of trade or firm names in this article is for reader information and does not imply endorsement by the U.S. Department of Agriculture of any product or service.

one or more serious errors are rejected, and must be corrected before FireStat will transmit the fire to NIFMID.

When the fire reaches NIFMID, it is checked again and a unique fire number is assigned. As soon as a fire is successfully loaded in NIFMID, any user can access the fire in reports, summaries, or download portions of the fire for analysis.

Catching errors before a fire is sent to NIFMID is much less expensive (and more accurate) than correcting fires loaded into NIFMID with errors. The effort to find and correct errors is significant, especially for fires prior to the current year.

Several data loading processes are implemented for WIMS. Field users can input weather observations directly into WIMS and immediately use the data for analysis and summaries. This is the typical path for manual stations or stations that have no automated link to WIMS. Users can also send the system administrator a data file to load. Over half of active weather stations (out of about 1680 active stations) are now Remote Automated Weather Stations (RAWS). Observations are taken hourly and transmitted periodically to a collection and processing facility. The processing facility reformats the data for transmission into WIMS, and WIMS loads the data into the appropriate tables.

WIMS grows daily with ingestion of weather observations and forecasts for the active stations, many of which record data year around, 24 times per day. Approximately one year of data is kept on line.

USING DATA FROM THE WAREHOUSE

Many users access and retrieve data from NIFMID, The most common usage supports fire planning work using the National Fire Management Analysis System (NFMAS) to model and plan for wildland fire suppression needs. The process relies heavily on historic fire occurrence and weather observation data to model local fire behavior and growth parameters.

KCFast (USDA Forest Service, 1995) was developed to take user input and build a batch query that is sent to NIFMID for processing. Users can request fire or weather data or summary reports without learning a complex computer control language. Output from KCFast gives users a window into NIFMID. Output generally falls into one of three categories.

1. Data files for input into analysis programs such as NFMAS, FirDat and Season (weather analysis tools), Fires (a program that allows comparison between weather and fire occurrences), RERAP (Rare Event Risk Assessment Process), and custom analysis tools such as local data bases and spreadsheets. Fire planners find many of the errors since they must review each fire during the preliminary planning process.
2. Standard reports such as weather forecasts, fire inventories, weather observation inventories, weather station lists, etc.
3. Predefined reports with user input to list or summarize fire data for the geographical area desired. Currently there are over 150 reports for fires, and a few weather summary reports.

The third output type allows managers to request summaries of various types. One recent example was, “How much would it cost the Forest Service if power transmission line owners were not held responsible for fires caused by downed powerlines?” Another requester asked for the 20 most expensive fires within the last 10 years. This type of output is very flexible and allows local managers great flexibility to answer questions and analyze their data.

FUTURE

There are ongoing discussions about other kinds of data that might be stored with the fire and weather data. One likely addition is data on management-ignited prescribed fire on Forest Service lands. At present the Forest Service does not keep such records on a national basis. Another possibility is aircraft usage, although that is less certain.

A new tool that is only a concept at this time will allow local fire managers to download fire and weather data for their management area. Once the data is loaded on their microcomputer, a wide range of tools are envisioned to enable the manager to review, correct, and analyze weather and fires. Most wildland weather analysis software only allows users to work with a single weather station. The new software will allow analysis of station aggregates with weighting factors and compare that weather data against their fire load.

Another initiative that will affect NIFMID is the National Interagency Fire Statistics Information Project (NIFSIP). This project is reviewing what data the five federal land management agencies collect with the charter to standardize federal fire occurrence data collection.

Work is on-going to determine the feasibility of using an Internet web page as the user interface for data requests (such as KCFast now does), and development of data drill-down tools that allow users to do sophisticated data searches.

LESSONS LEARNED

1. Build procedures at the database level to ensure all records added to the database meet at least minimal standards. For example, fires at elevations above 14,494 feet (elevation of Mt. Whitney in California) in the lower 48 states would be in error.
2. Scrub your data! Search for and correct obvious errors such as latitudes greater than 90 degrees, or fires that plot in the Atlantic or Pacific Oceans.
3. Check the data as thoroughly as possible at time of capture to identify data problems early. Once the input person leaves the record, corrections are much more expensive.
4. Work closely with interested parties. Sometimes the most off-beat suggestion is popular with the field users when implemented properly.
5. Be flexible. All the uses for the data haven't been invented yet.
6. If you don't use your data, the lowest acceptable accuracy becomes the norm.
7. When data isn't used, user confidence in the data is fairly low. The data may be correct but few believe it to be correct.
8. Sir Josiah Stamp[3] cautioned his peers with a statement that is still true today. He stated, "The government is very keen on amassing statistics. They collect them, add them, raise them to the Nth power, take the cube root and prepare wonderful diagrams. But you must never forget that every one of these figures comes in the first instance from the village watchman, who just puts down what he damn pleases". The dispatcher who fills out the fire report probably didn't get to smell the smoke; they do the best they can with the information available to get the report right. The dispatcher or field person is like the village watchman when it comes to fire reports.
9. One last caution: before you rely heavily on data for analysis, try to understand why the data was collected, and the data collection rules. For example, if fire cost per acre is calculated from NIFMID, the analyst should understand that only Forest Service costs are included in the total cost. The true cost may be much higher when costs other agencies accrue are included.

SUMMARY

The data warehouse provides a way to summarize and analyze fire and weather information for a variety of purposes. Building a data warehouse isn't easy, but when done well, is worth more than the cost. A well designed data warehouse should allow diverse and seemingly unrelated data to be linked so that

[3] Sir Josiah Stamp worked for the Inland Revenue Department in England from 1896-1919.

interactions and hidden relationships become visible. Many questions can be answered with accessible data.

LITERATURE CITED

BARROWS, J.S. 1951. *Forest fires in the Northern Rocky Mountains.* U.S. Department of Agriculture, Forest Service, Northern Rocky Mountain Forest and Range Experiment Station, 251 pp., illus. Missoula, Montana.

COREY, MICHAEL J., AND ABBEY, MICHAEL. 1997. *Oracle Data Warehousing.* Osborne McGraw-Hill, 356 pp., illus. Berkeley, California.

DONOGHUE, LINDA R. 1982. *The history, and reliability of the USDA Forest Service wildfire reports*. U.S. Department of Agriculture, Forest Service, North Central Forest Experiment Station, 15 pp., illus. St. Paul, MN.

FURMAN, R. WILLIAM, AND BRINK, GLEN E. 1975. *The national fire weather data library: What it is and how to use it.* Gen. Tech. Rep. RM-19. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest Experiment Station, 8 pp., illus. Ft. Collins, Colorado.

GOLDBERG, MICHAEL, AND VIJAYAN, JAIKUMAR. 1996. Data 'wearhouse' gains. *ComputerWorld* 30(19):1+ (6 May 1996).

SHOW, S.B., AND KOTOK, E.I. 1923. *Forest Fires in California 1911-20: An analytical study.* U.S. Department of Agriculture Circ. 243, 80 pp., illus. Washington, DC.

U.S. DEPARTMENT OF AGRICULTURE, FOREST SERVICE. 1995a. *Kansas City Fire Access Software (KCFast) user's guide.* U.S. Department of Agriculture, Forest Service. 204 pp., illus. Washington, DC

U.S. DEPARTMENT OF AGRICULTURE, FOREST SERVICE. 1995b. *Weather Information Management System (WIMS) user's guide.* U.S. Department of Agriculture, Forest Service. 364 pp., illus. Washington, DC

U.S. DEPARTMENT OF AGRICULTURE, FOREST SERVICE. 1996. *FireStat user's guide: how to enter fire reports using the FireStat system.* U.S. Department of Agriculture, Forest Service. 200 pp., illus. Washington, DC.

YANCIK, RICHARD F., AND ROUSSOPOULOS, PETER J. 1982. *User's guide to the national fire occurrence data library.* U.S. Department of Agriculture, Rocky Mountain Forest Experiment Station, 25 pp., illus. Ft. Collins, CO.