Reprinted from

Eighth International Symposium

Machine Processing of

Remotely Sensed Data

with special emphasis on

Crop Inventory and Monitoring

July 7-9, 1982

Proceedings

Purdue University
The Laboratory for Applications of Remote Sensing
West Lafayette, Indiana 47907 USA

A DATABASE TO SUPPORT CROP CONDITION ASSESSMENT USING REMOTELY SENSED DATA

W.G. DRIGGERS

U.S. Department of Agriculture/
Foreign Agriculture Service
Houston, Texas

I. BACKGROUND

Few computer applications provide the potential for contributing to objectives as valuable as the timely assessment of major crops worldwide, which can help achieve better profits for farmers and assist in alleviating world food supply problems. The application of remote sensing technology to crop condition assessment provides that potential, and the Foreign Agricultural Service (FAS) of the U. S. Department of Agriculture (USDA) has for the past four years invested in the development of a minicomputer based system to apply that technology. This state-of-the-art system features a minicomputer network, an agricultural and meteorological database accessed via color graphics displays, and digital image analysis. This paper describes the application and focuses on the design, implementation, and administration of the interactive geographic database which supports crop condition assessment.

The FAS crop condition assessment system has been developed following USDA participation in the Large Area Crop Inventory Experiment (LACIE), a joint project of the National Aeronautics and Space Administration (NASA), the National Oceanic and Atmospheric Agency (NOAA), and USDA during the period from 1974-1978. The experiment demonstrated the feasibility of using remote sensing data, from LANDSAT satellites, and climatic data, from the met station network operated by the World Meteorological Organization (WMO), to perform an inventory of an important crop, chosen to be wheat. In 1978, FAS established the Foreign Crop Condition Assessment Division (FCCAD) to apply remote sensing techniques to foreign crop analysis following completion of LACIE.

Techniques developed during LACIE for collecting, maintaining and processing digital imagery data and climatic data have been important to the development and evaluation of crop condition assessment methods by FAS[1]. Experience with LACIE data management problems indicated that improved efficiency in data management was needed; for example, data handling interfaces among LACIE components were at times awkward, time consuming, and difficult to control. Analysis of the needs for improved data management indicated that requirements could be met best through implementation of a central database under control of a generalized database management system[2].

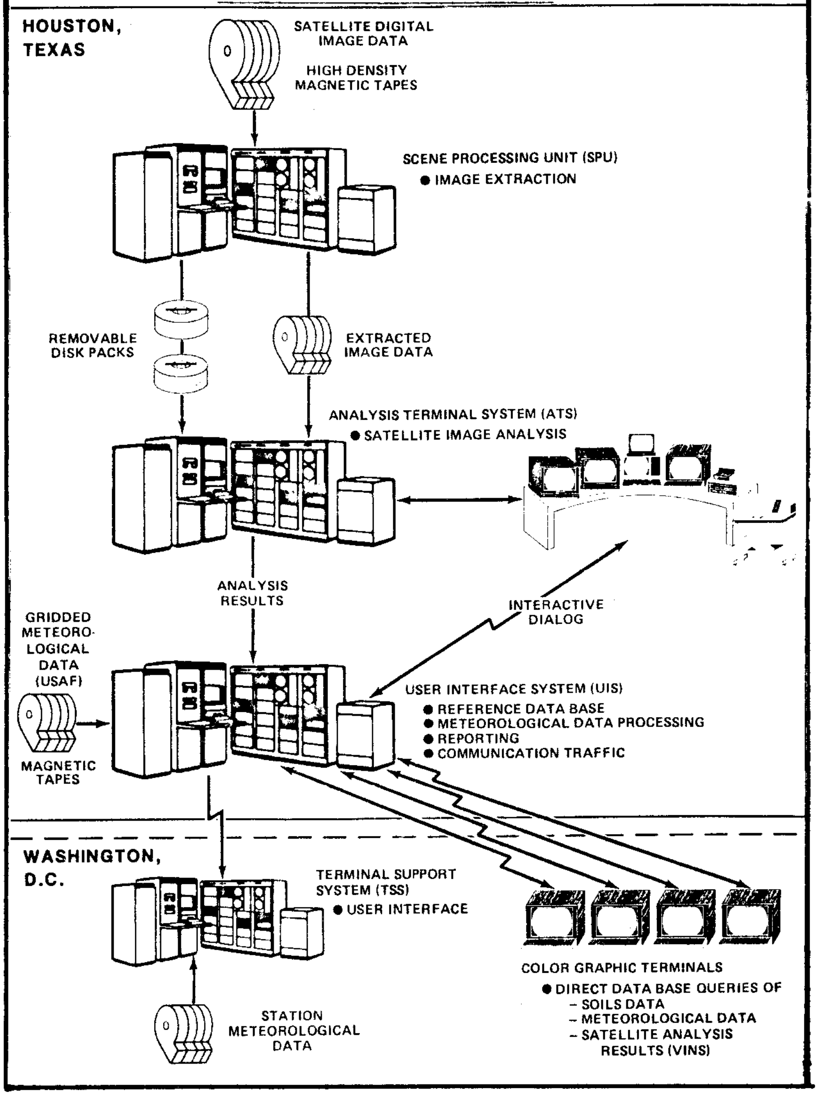
II. THE APPLICATION

A. SYSTEM FUNCTIONS

A network of three Digital Equipment Corporation (DEC) PDP 11-70 computers, all located in a site adjacent to NASA's Johnson Space Center in Houston, Texas, is used in crop condition assessment. A fourth, a PDP 11-34, is located at USDA facilities in Washington, D.C. Data in any of the systems can be accessed from the others using DECNET software and hardware. Each of the four performs a specific function, as shown in Figure 1.

SPU. The scene processing unit (SPU) is used primarily for extracting digital images of useful dimensions from LANDSAT or METSAT imagery. The SPU includes a high density tape unit for reading digital images and a Floating Point Systems AP-120B for high speed arithmetic tasks such as geometric corrections. Following image extraction, sampling and correction, LANDSAT images are recorded on computer compatible tape (CCT) or removable disk packs in

FIGURE 1. SYSTEM FUNCTIONS



universal format for access by the analyst terminal system; METSAT images are recorded on removable disk packs. No digital imagery is stored in the CADRE database.

ATS. The analyst terminal system (ATS) is the remote sensing analyst's interface for reviewing and analyzing imagery. Three analyst stations are part of the configuration; each station includes two CONRAC display consoles for image display, a video monitor and recorder for transferring images to video diskettes and an alphanumeric terminal for controlling the imagery analysis process. The ATS configuration includes an AP-120B floating point processor for high speed arithmetic for tasks such as image clustering and classification. ATS software provides capabilities for displaying and manipulating images, calculating vegetative indexes, spectral analysis functions, and techniques developed both in LACIE and by FCCAD.

UIS. Operational database residence is on the user interface system (UIS). Remote sensing analysts access ancillary data in the database either by initiating a report generation task from an alphanumeric terminal or through use of a versatile color graphics display unit. Each analyst station includes an ISC color graphics unit.

FIGURE 2. SOIL MOISTURE BY MET STATION
(Surface/Subsurface/Percent of Capacity)

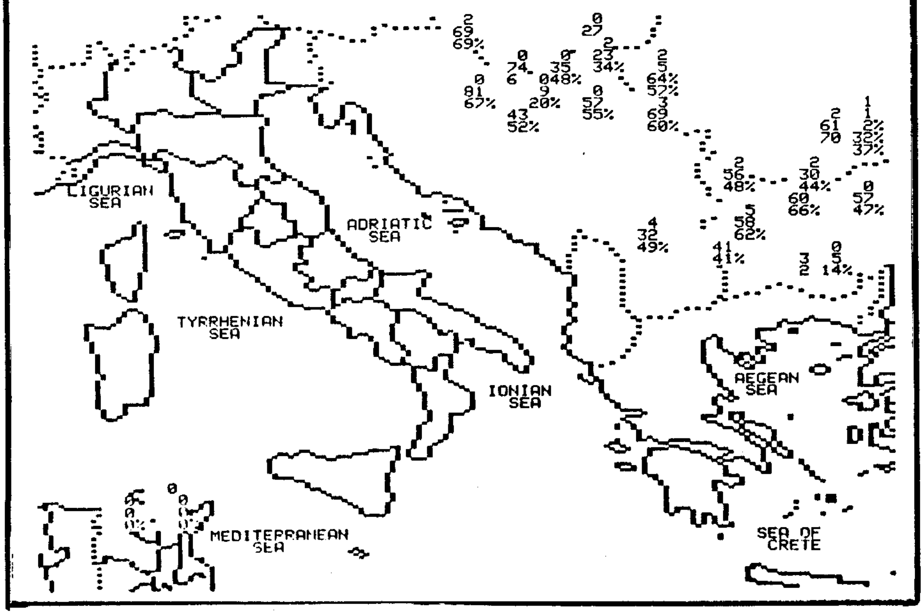
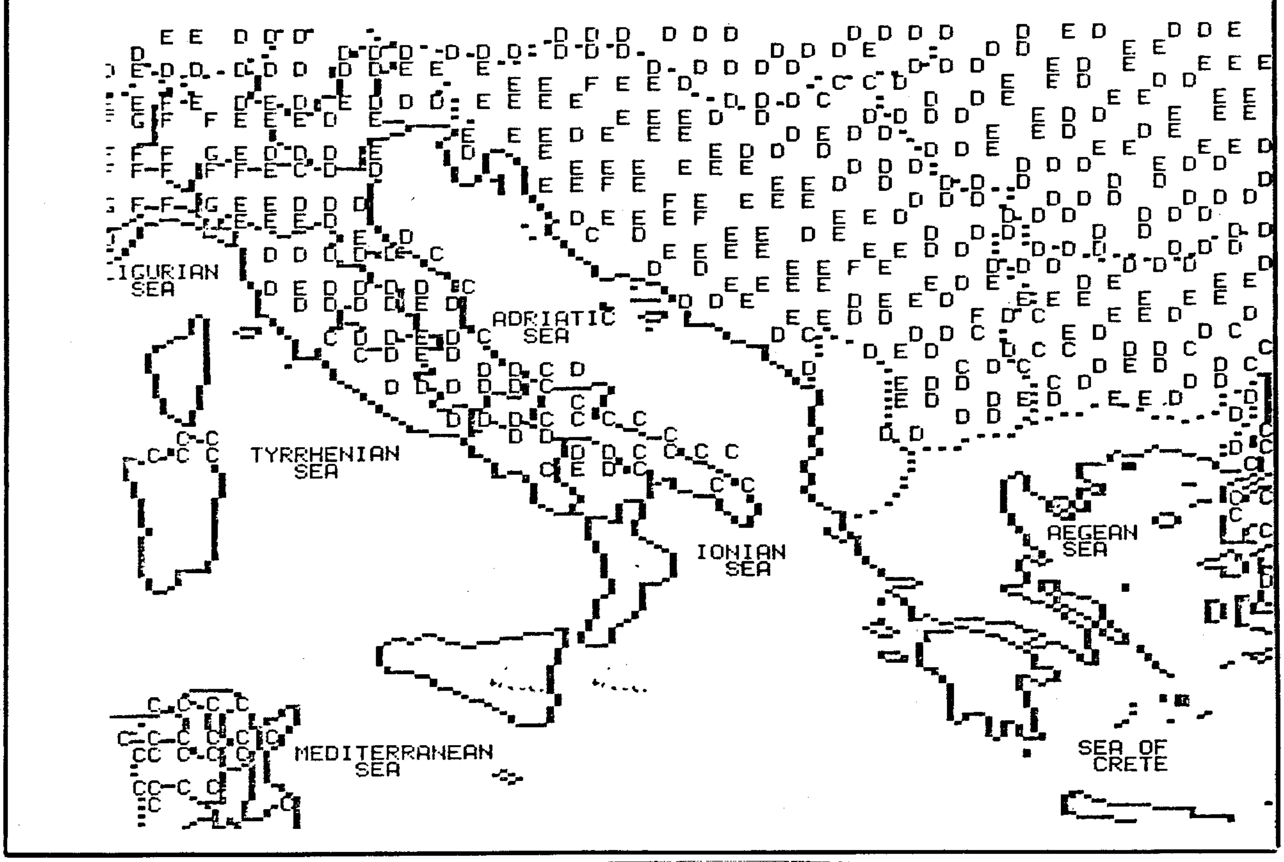


FIGURE 3. GRIDDED VEGETATIVE INDEX PLOT



TSS. The terminal support system (TSS) is a DEC PDP 11-34 located at the USDA South Building in Washington, D. C. The TSS is used to transmit daily meteorological station data to Houston via telecommunications. Reports generated from the database are also transmitted to the TSS.

B. ANALYST FUNCTIONS

FCCAD remote sensing analysts produce monthly and ad hoc crop condition assessment reports for key crops in countries of interest. Each analyst is responsible for crop assessment in one or more countries. Analysts base their reports on the analysis of both remotely sensed images stored on the ATS and ancillary data stored in the database. System access to support the analysis is performed primarily at the analyst station, as shown in Figure 1.

The two imagery analysis screens can be used to display in false color single images up to 512x512 pixels (or picture element dots) or four smaller images as extracted on the SPU. A full scene image can be displayed as a series of subscenes, or as a single image subsampled through the use of skip factors. Masks can be defined for specific fields on an image and stored for future analysis. Meteorological station locations and grid cell outlines can also be displayed as image overlays and identified for reference to ancillary data.

Ancillary data (such as climatic, soils, or crop statistics data) can be accessed either through display on the color graphics terminal or through report generation programs. Because of the geographic structure of the database, any of several data types can be displayed rapidly on a map projection on the color graphics screen. Map area and data type are determined through menu controlled interaction with the system. Figures 2 and 3 show typical data displays using the Agricultural Information Display System (AIDS), which has been developed to interface with the database.

C. DATABASE REQUIREMENTS

The CADRE (for Crop Assessment Data Retrieval and Evaluation) database design was based on the requirements of FCCAD remote sensing analysts and management, together with those imposed by the ancillary data characteristics. These requirements are extensive and only the most general are described below.

Typical analyst requirements include the capability to retrieve data as a function of geography and time. For example, the analyst may need to see total precipitation for a ten-day period plotted over some portion of the country he is analyzing. The need might also arise to analyze data in a crop growing area, as opposed to an area defined by political boundaries. The analyst needs flexible, well-defined interactions for defining queries and reasonable response for each type of query.

General requirements for the database include data characteristics, such as volume and relationships, and data management capabilities. Large volumes of meteorological parameters and products derived from those parameters must be stored in the CADRE database; significant volumes of vegetative indexes are also required. Other general requirements include data consistency and minimum redundancy.

III. DATA CATEGORIES

As stated earlier, digital imagery data are not stored in the CADRE database, although vegetative indexes derived from those images are. Only ancillary data - geographic definitions, station and gridded met data and products, vegetative indexes, agronomic data, and historical crop statistics - are stored in the CADRE database, which is the subject of this paper.

A. GEOGRAPHIC DEFINITIONS

Geographic entities defined in the CADRE database include the geographic hierarchy, agro-physical units (APU's), meteorological stations, grid cells, and Operational Navigation Charts (ONC's). Definition of the relationships among these entities is a key element in the database structure. All other data types are stored and accessed with relation to one or more of these geographic entities.

Over a period of time, geographic hierarchy definitions have varied, and the database definition must provide flexibility to change accordingly. Initially, hierarchies were defined to be those used in countries analyzed during LACIE. Names for those hierarchy levels - country, region, zone and stratum - are still used for all countries. FCCAD analysts have, however, applied different approaches to defining hierarchy levels in new countries. For the most part, hierarchy levels are now defined to correspond to those used for crop reporting hierarchy levels by each specific country; a typical result of this approach is that the number of levels varies with respect to country size.

The grid cell entity adds considerable flexibility to the geographic definitions. Defined as a rectangular mesh on a polar stereographic plane and projected onto the earth's surface from the opposite pole, the length of a side of a grid cell is approximately 25 nautical miles at middle latitudes (size and shape vary significantly with latitude and longitude). Each full grid cell is identified by an (I, J) pair, representing the matrix location with respect to the two axes on the stereographic plane. Each cell is divided further into quadrants (identified as A, B, C and D). Other geographic areas can be defined in terms of grid cell quadrants; for example, the lowest level of each country's geographic hierarchy is defined as the collection of grid cell quadrants contained in that area. Because simple algorithms can be used to convert from latitude and longitude to (I, J, K) and vice versa[3], the grid cell quadrant is an important "building block" for the CADRE database.

Meteorological stations are, for the most part, those World Meteorological Organization (WMO) stations in the crop areas of interest. Standard WMO codes are used to identify the stations. Latitude and longitude for the stations are used to relate them to other geographic definitions.

The Agro-Physical Unit (APU) is defined as an area with similar soils, climate, topography, and other agronomic factors such as land use intensity. The APU is also defined in the CADRE database as the collection of grid cell quadrants which it contains. APU definitions are useful in collecting data and analyzing crop conditions in crop growing areas, as distinguished from political or other geographic areas.

Operational Navigation Charts (ONC's) are Lambert conformal conic projection maps at a scale of 1:1,000,000 used both in data collection and for data display purposes. ONC's are used with transparent overlays, for example, in the digitizing process by which grid cell quadrants are identified with geographic entities. ONC definitions are also used to help provide good query response time for some of the AIDS displays.

The geographic orientation of the CADRE database is important and has changed significantly as data requirements have become better defined. The CODASYL-type database management system has proved to be both versatile and efficient in data retrieval, display, and report generation for this type of data.

B. METEOROLOGICAL DATA

Two types of meteorological data (WMO station and gridded) are maintained in the CADRE database, both to assist in crop condition assessment and as input to mathematical models discussed in Section 3.C below. Daily meteorological parameters available for about 3,500 selected worldwide stations include maximum temperature, minimum temperature, 24-hour precipitation, and snow depth. Normal temperature and precipitation are available by month for the stations. Daily parameters available for about 7,000 grid cells include maximum and minimum temperatures, 24-hour precipitation, evapotranspiration, solar radiation, infared radiation, and snow cover. Tri-monthly summaries (8-11 days) of both types of met parameters are stored in the database prior to purging daily data.

C. MET DATA PRODUCTS

Mathematical models have been developed for FCCAD use in estimating crop-growing variables such as soil moisture, winterkill, plant stress, and crop development stage. These models are based primarily on climatic conditions, as defined by daily met data parameters described in Section 3.2 above; other important parameters (all stored in the database) are heat coefficients derived from normal temperatures, water holding capacity for the specific soil, and approximate planting dates. Model results are estimated and stored in the database daily for soil moisture, and daily but only during the crop growing season for the other variables. Tri-monthly summaries (8-11 days) of model results, except for winterkill, are stored in the database prior to purging daily data.

D. VEGETATIVE INDEXES

Vegetative indexes (VIN's) are calculated from satellite imagery one pixel at a time and summarized to provide an estimate of plant growth and vigor for each cell. Several such indexes have been derived, some of which are based on LANDSAT imagery and others on METSAT imagery. Up to four summary indexes can be stored in the database for each satellite image acquisition.

E. AGRONOMIC DATA

For FCCAD analysts, the most important soils data elements stored in the database are available water holding capacity and land use suitability. Soils characteristics, compiled for U. S. soils are stored for soil types of interest. Soil types are then related (for comparable foreign soils) to geographic areas using soils maps and transparent overlays at the grid cell quadrant level[4].

Historical crop statistics are stored for different hierarchy levels, where reliable information is available from various sources; typically, ten years of statistics are stored for important crops. These statistics are useful for comparing current crop areas and production with historical trends.

IV. DATABASE ENVIRONMENT

A. SYSTEM HARDWARE

The DEC PDP 11-70 which serves as the database component for the system has been described briefly in Section 2.1. The configuration includes DEC floating point hardware, 768 K bytes of memory, one magnetic tape unit, a line printer, and seven disk drives, each having formatted capacity of 176 million bytes. The CADRE database resides on five of these seven disks.

B. SYSTEM SOFTWARE

Standard DEC system software is used, primarily the IAS operating system, FORTRAN, and COBOL. IAS supports interactive, time-sharing, local batch, and remotely-initiated batch processing.

C. DBMS-11

Based on the 1971 CODASYL Task Group Report, DBMS-11 supports both hierarchical and network types of data structures[5]. DBMS-11 provides separate language facilities for data definition (schema DDL and subschema DDL) and for data manipulation (DML); both FORTRAN and COBOL DMLs are available. As with any CODASYL system, all data elements, record types, set relationships, and files are defined in the schema. The user can access the database only through a subschema, a subset of the schema defined to include all (and only those) data elements and relationships needed for a specific application.

Several utility programs are available with DBMS-11[6]. FCCAD uses most of these, including programs for storage allocation, dump and restore functions, and the on-line query for applications programmers.

D. APPLICATIONS

Application programs are required for all database access functions - loading or maintaining data, querying database contents and generating reports. Data loading programs and maintenance programs (for example, for offloading or purging data) can be particularly complex, especially when multiple record types and network relationships are involved.

V. DATABASE IMPLEMENTATION

A. DESIGN APPROACH

Database design and development have evolved much as the application goals have evolved. LACIE was concerned primarily with the statistical separation of classes in order to identify crops in single fields, and the early stages of database design supported that type of analysis. Foreign crop condition assessment over broader and more diverse areas relies more on vegetative index summaries and general climate conditions, and the present design approach supports this type of analysis.

B. STRUCTURE

CADRE database structure reflects natural relationships among the various data types. Six areas have been defined, as follows:

geographic definitions
station met data and products
gridded met data and products
vegetative indexes
soils data
crop statistics

Record types within these areas are also based on the data types one would expect; for example, the geographic definition area contains records for entities such as country, region, zone, grid cell quadrant, APU, and ONC.

CODASYL set definitions provide the user with a powerful tool for optimizing response time. There are at least four important aspects to using set definitions for good performance. First, the database key for the next set member record, in the access order specified for the set, provides direct access to that record, minimizing disk head movement. Next, set definitions, combined with the location mode, can be used to cluster member records physically as close together as possible to the owner, further eliminating disk accesses. Next, the capability for defining more than one set relationship between two record types can be used to reduce still further the number of records accessed for query response. Fourth, the amount of data required to be stored in a member record is usually reduced, thereby decreasing record size and increasing how many are accessed at one time. Other advantages probably exist also; these four have been used in the CADRE database definition to provide fast response for queries.

Set definitions are not unmixed blessings, however, and must be used with care. As a practical matter, set definitions which cross area boundaries create operating problems and design problems. CADRE database integrity is provided mostly through the use of dumps, and DBMS-11 requires that these dumps be performed by area, some of which are very large. If set definitions involve record types in two areas, both must be dumped at the same time to ensure consistency. Because dumps require a lot of time, particularly with large areas, schedule problems result. Also, when a set spans two areas, record clustering benefits for the member record type are not available. To reduce this problem, redundant record types have been used (very effectively) in some parts of the CADRE database design; redundant data always create problems of data consistency.

C. ALLOCATION

Allocated disk storage, by database area, is shown in Table 1. Only the met data areas (station and gridded) have become periodically saturated to date, requiring the creation of archives and the purging of older data. The geography area is so nearly saturated that data loading has ceased. Other data types have longer retention cycles.

TABLE 1: DISK STORAGE ALLOCATION

CADRE DATABASE AREA	ALLOCATED STORAGE (MEGABYTES)	APPROXIMATE OCCUPANCY %
Geography	29	88
Gridded Met	335	84
Station Met	116	86
Soils	10	25
Crop	25	25
Veg Index	150	20

D. OPERATING FUNCTIONS

As indicated previously, CADRE database dumps are performed periodically (every two weeks at present) to ensure data integrity. Backup files of data which have been loaded are retained for a month so that they can be reloaded if a disk pack failure results in corrupted data files. Journalling capabilities are available with DBMS-11 and were used routinely at one time. Execution time for database access was so much greater with active journal files that it was decided simply to recover and reload from the most recent dump in case of failure; only two such events have occurred over the past two years, so that this has worked well. If interactive transactions or any realtime data collection resulting in database updates were performed, journal files would of course be required to ensure recovery.

Subschema definitions (user interfaces) are the only effective tools for providing database security with DBMS-11. As the number of users and applications programs grow, so do the number of subschema modules. Although the DBMS-11 data dictionary contains complete subschema definitions, this information is not accessible to the database administrator. For this reason, FCCAD has developed local subschema control software, a very valuable tool. This software is implemented as a small set of simple programs using DBMS-11. While use of the subschema control software and monitoring of activity do not ensure security, they are effective in our small installation.

E. DATA DICTIONARY

DBMS-11 data dictionary facilities are minimal for database administration purposes. No provision is made, for example, for narrative descriptions of data elements, record types, set types, areas, or files in the reports that can be generated. No provision is available for determining what record types or application programs must be modified if a data element format or representation is changed. Because the need for information of this type increases with database usage, the subschema control software developed by FCCAD has been extended to provide these, and other, capabilities.

VI. EXPERIENCE AND OBSERVATIONS

FCCAD now has three to four years of experience in implementing the geographically oriented CADRE database working with several constraints, some of which were the following:

- Available resources required that the database must be implemented on a minicomputer with commercially available software.
- Available resources required that the effort use minimum ADP staff (a total of six to eight man years of effort have been spent through FY 81 for database requirements definition, design, development, and implementation).
- Response time had to be fast enough to support the required analysis functions despite the large data volumes involved.

The DBMS-11 system has performed well. It should be noted that, at the time of initial hardware and software procurement (1976-1977), not many fully supported DBMSs were commercially available for the PDP 11-70 so that the selection process was not painful. The system does have faults, and even "bugs", however. Some of the faults have been mentioned; another example of inadequate database administration tools is the lack of a database restructure capability. DEC has compensated to some degree for shortcomings by being willing to discuss problems and "workarounds" informally to help us get past tough situations. Furthermore, use of the DML by a user with only slight FORTRAN experience has proved easy to be learned; an analyst can learn to produce useful reports in a short time. Overall, DBMS-11 must be considered a good choice for the crop condition assessment application.

Other problem solutions have also worked out well. The color graphics display capability was developed almost totally by FCCAD staff with no prior experience in that field; the capability for the analyst to view database contents on a map background greatly enhances its value. Interfacing the H. Dell Foster digitizer to the UIS was also a minimum budget, "homegrown" effort and has repaid the cost many times: data is available sooner for a larger area of the world as a result.

Alternative approaches have been, and will continue to be, considered wherever it appears advantageous. Relational DBMSs might provide better ad hoc query capabilities, for example. However, unless performance, especially with respect to query response time, can match that of DBMS-11, the capability of responding to ad hoc queries is not that important. It would be nice, however, to be able to apply relational data structure techniques to the design and definition of a CODASYL-type schema (with additional language facilities) and, as a result, to benefit from some of the operations available for envisioned for relational databases[7].

VII. SUMMARY

In summary, the CADRE database is an essential part of the FCCAD crop condition assessment system. The system is a dedicated minicomputer application which integrates digital imagery extraction, processing and display; mathematical models and statistical analysis based on worldwide meteorological data and other data types; and a central database under control of a DBMS, accessed through an impressive color graphics display.

CADRE database structure has proved to be sufficiently general to adapt to changing requirements. This has led to the conclusion that the structure should be useful even to a broader range of geographic applications. Map projections and scales can vary, the basic data types to be stored and retrieved can vary, and the scale of the application can be changed within the overall framework of the system. The approach could be applied to the analysis of other resources such as water, forests, energy, or land use. Experience with the CADRE database has provided interesting insight to a significant set of computer applications.

REFERENCES

1. Driggers, W. G., Downs, J. M., Hickman, J. R., and Packard, R. L., "Data Base Design for a Worldwide Multicrop Information System," from Proceedings of the LACIE Symposium, NASA Johnson Space Center, October 1978, pp. 1085-1096.

2. Martin, James: Computer Data Base Organization, second edition, Prentice-Hall, New York, 1977.

3. Earthsat Spring Wheat Yield System Test 1975, Final Report. Earth Satellite Corporation, Washington, D. C., April 1976, pp. 3-1 to 3-4.

4. Westin, F. C.: Landsat Interpretation - Intensive Use as Cropland. South Dakota State University, Brookings, South Dakota, 1977.

5. CODASYL Data Base Task Group Report. Association for Computing Machinery, April 1971.

6. Data Base Administrator's Guide. Digital Equipment Corporation, Maynard, MA, July 1977.

7. Date, C. J.: Database Systems, second edition, Addison-Wesley, 1977.

W. G. Driggers. Mr. Driggers has served as Database Administrator for the USDA/Foreign Agricultural Service's Foreign Crop Condition Assessment Division in Houston, Texas, since 1976. Prior to that time, his remote sensing applications experience began with work on the Large Area Crop Inventory Experiment (LACIE) in 1975. His background in computer technology includes scientific applications and system software assignments at General Electric, Lockheed and NASA, in both management and technical contributor positions. He has a Bachelor's degree in mathematics and physics from the University of Alabama.