JPL PUBLICATION 79-70

Data Base Management Systems Panel Workshop: Executive Summary

(NASA-CR-162105) DATA BASE MANAGEMENT SYSTEMS PANEL WORKSHOP: EXECUTIVE SUMMARY (Jet Propulsion Lab.) 18 p HC A02/MF A01 CSCL 05A

N79-31066

Unclas

G3/81 31791

August 1, 1979

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

JPL PUBLICATION 79-70

Data Base Management Systems Panel Workshop: Executive Summary

August 1, 1979

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ABSTRACT

An executive summary of the discussions of the first workshop for a panel on Data Base Management Systems (DBMS) for space-acquired and associated data is presented. The panel was concerned with investigating the full range of DBMS needs, which include acquiring, managing, storing, archiving, accessing and dissemination of data for an application. Recommendations were made to NASA regarding the future development and support of DBMS technology. The discussions further brought up existing bottlenecks in NASA DBMS operations, expected developments in the field of remote sensing, communications, and computer science, and presented an overview of existing conditions and expected problems. The requirements for a proposed spatial information system and characteristics of a comprehensive browse facility for NASA earth observations applications were discussed.

FOREWORD

The Data Base Management Systems Panel Workshop was held at the California Institute of Technology, Pasadena, California, March 5-7, 1979. The workshop was sponsored by the Applications Division of the Office of Space and Terrestrial Applications, National Aeronautics and Space Administration, and was coordinated by the Jet Propulsion Laboratory.

TABLE OF CONTENTS

INTRODUCTION

The Applications Systems Division of the Office of Space and Terrestrial Applications (OSTA) held its first workshop for a panel on Data Base Management Systems for space-acquired and associated data at the California Institute of Technology on March 5-7, 1979. The panel has been requested to consider future applications needs for cataloguing, data integration/preparation services and geocoding of imaging and non-imaging data derived from earth observation systems. A summary of the panel's findings is presented here, with a final report to be issued in the near future.

The Panel was concerned with investigating the full range of future data base management systems (DBMS) needs. Thus, the concerns of DBMS operations included cataloguing, acquiring, managing, storing, archiving, accessing and disseminating all data for an application. Furthermore, it was recognized that a DBMS must be able to provide various levels of support to research activities, and therefore provide both manipulating functions (e.g. scale, rotation and projection changes) as well as analytic functions (i.e. convert raw data to processed data and processed information). From the panel discussions of each of the above operations, and their status and future needs for earth observations data management, the following recommendations were made.

RECOMMENDATIONS

1. Central Browse Facility. NASA needs to develop a central, but not centralized, catalogue/browse facility which stores the location of all the kinds of data needed for applications. The facility should provide a human engineered (i.e. query system) functional index aid which would key earth observations data and associated ancillary files needed for an application, and provide information on methods needed to acquire the data. An economic feasibility study should be undertaken to determine potential overhead cost of such a facility.

2. Archiving. As an agency that is chartered to conduct and support research, the problem of permanent data archives of NASA and related data is important and should be considered as a basic element in a data management program. Within NASA, NASA should address the question of who archives what and how much data should be archived for how long. It is suggested that an investigation be undertaken to determine a hierarchy of archiving requirements and their associated levels of spatial and temporal resolution of data types. Furthermore, NASA should investigate the feasibility of maintaining state-of-the-art storage mediums and pursue the most cost-effective alternatives.

3. Spatial Information System. NASA, perhaps with other agencies, should establish a multi-center activity to produce a comprehensive/unified plan for the development of a new generation of spatial information systems designed to integrate NASA and Non-NASA geolocated data. The activity should identify and prioritize all technical issues and problems related to developing a comprehensive spatial information system including (but not limited to) the following: a) assigning responsibility for solving the identified problems to specific centers/agencies having the proper expertise/needs, b) developing and fostering standards, c) planning for wide dissemination of the technology developed; and d) planning for data system demonstration projects.

4. Standards. NASA, in cooperation with other agencies and the ANSI, should foster and develop standards in a variety of areas which presently impede the formation of general purpose DBMS operations. These include data base and data format standards for data sharing, software engineering standards to ensure systems transferability, standard protocols, interface standards, quality control and validation standards for data, and standardized labeling procedures which would summarize the history of processing functions applied by a facility.

5. Follow Communications Technology. NASA should, in cooperation with other agencies, investigate ways of improving data sharing capabilities, of linking users and data suppliers, and of providing improved data networking capabilities as the state-of-the-art in communications hardware develops.

6. Data Base Sharing. NASA should investigate the possibility of having large data base producers within the agency (e.g. LACIE, Landsat-D Project, Climate Program, etc.) share their data bases on a cost recovery basis and, more importantly, subsidize the archiving and initial data preparation costs prior to storage. The accessibility of data should not depend upon the funding of the project/agency that collects/generates the data but should be assured by the availability of specially identified funds.

7. Non-NASA Data Sources. NASA/OSTA needs to clearly identify needs for non-NASA data and make the necessary arrangements to tap these non-NASA data sources and bring to the attention of non-NASA facilities the NASA/OSTA data bases that may support their activities.

8. Data Management Program. OSTA should identify the goals and objectives of a comprehensive data management program and should develop a detailed plan which specifies the activities and milestone schedules needed to accomplish the established goals and objectives. OSTA should form a Steering Committee of knowledgeable NASA and non-NASA personnel to help formulate an OSTA data management program.

9. Mission Reviews. NASA should evaluate proposed data management practices for new programs to assure their use of existing technology and data standards prior to funding approval.

10. Review Workshop Findings. OSTA management should review the report produced as a result of this workshop and should provide guidelines needed to scope and focus future workshop planning activities.

PURPOSE OF THE WORKSHOP

As NASA moves into the 1980's it intends to provide a key role in the development of integrated analyses of global resources. Data base management systems (DBMS), given their function of keeping track of and integrating data for user applications, promise to assume a growing importance in assuring successful efforts in global resource monitoring and analysis. As NASA/OSTA has attempted to more fully provide the applications user community with more information (as opposed to data), the needs for engineering technology/scientific research and applications have become inextricably interrelated. In such complex areas as water runoff assessment, synoptic climatology, and modelling potential changes in land resources, the data base generated from an operational mission may very well contribute to a purely research project. Thus, while parallel efforts in R&D and operational programs are unwise according to all logic, they must be done in the light of practical requirements to satisfy users and justify missions. A common NASA-wide approach to solving data accessibility and data integration problems may well provide the only cost effective alternative to the present policy of internalized data gathering, storage and retrieval, and integration for analysis.

The workshop brought together a panel of NASA and non-NASA experts who have designed and constructed DBMS that support earth observation systems. The panel was formed to assess the state-of-the-art in DBMS hardware and software design technology and determine the potential for improvement in those areas where earth observation systems interface with DBMS applications. The need to effectively address the future requirement and functional capabilities needed for applications DBMS in becoming readily apparent as NASA research moves toward addressing more complex issues through the use of earth observations satellites.

DISCUSSION

The workshop discussions began with four presentations on existing data base management systems currently being applied to earth observation systems data. The presentations enabled the panel to focus on the issues both at hand and potential, and set the tone for the group discussions that ensued. While some areas of concern could be dismissed as specific problems which NASA/OSTA was aware of and currently addressing, the majority of the issues brought up were of a general nature which promise to continue to be with us until remedied. In an effort to highlight the discussion results which can be addressed most effectively by NASA technology initiatives, this section is broken out into three parts: bottlenecks, expected developments, and existing conditions and expected problems.

Bottlenecks

1. The most serious problem today is the difficulty in finding and securing all the data needed for a given application. In addition, data sometimes become unusable since some vital data input or information about the data is unavailable. Two examples illustrating the point were brought forth during the discussion: 1) LACIE non-image data were very sparse and electronic linkages with meteorological information were not always evident. 2) The lack of digital data archiving of weather satellite data, particularly SMS/GOES, will severely cripple NASA's contribution towards weather and climate programs for years to come.

2. Even after locating the required data, moving it from one organization/location to another is a slow, often error prone and time consuming process.

3. Improvements are needed in NASA's addressing and making use of mass storage technology to reduce the cost, increase the shelf life usefulness, and increase the capacity of data storage. A quantum reduction in cost of data archiving would, for instance, eliminate the current SMS/GOES digital data archive problem. In addition, technology assessment needs to address the problem of increasing by a large amount the number of access paths to mass storage, thereby assuring ease in accommodating the large number of users expected with operational systems.

4. No comprehensive/easily used capability exists for integrating data spatially and temporally or to analyse or use such data for modelling studies. Existing systems have many limitations, they have not been fully developed nor are they widely available. Solving this problem requires the development of a comprehensive geocoded information system which addresses both the NASA and non-NASA components, particularly:

a. A need for further research in more efficient/accurate algorithms involved in spatial data manipulation.
b. A need to assist in development of more efficient systems associated with digitizing technology, as it is a principal form of non-image data base contribution of spatial information.
c. The need to complete the ground control point file for the Landsat Master Data Processor as soon as possible, as precise geolocation is the first step required to achieve a general purpose integration of imaging and non-imaging data for spatial DBMS operations.

Expected Developments

1. For economic reasons, satellite remote sensing may well replace aircraft and other means of acquiring earth observations data within ten years.
2. Hardware developments in the next decade are expected to provide for increases by factors of two or more in storage technology, in microprocessor capabilities and in I/O channel capabilities, resulting in faster/cheaper access to more data.
3. Software developments will concentrate on comprehensive spatial information system(s). The system will steadily become easier to use with increased user familiarity and increased attention to user/system interfaces.
4. The costs of performing duplicated processing operations at multiple locations will force the need for shared processing, a functional division of responsibility, shared data bases, shared systems technology, and will result in de facto standards.

Existing Conditions and Expected Problems

1. Range. The range of users of earth observations data who must be serviced by a DBMS facility for an application is great with respect to needs and also with respect to technological sophistication. At one extreme are state and local governments who need someone to develop a DBMS technology and then with a large investment in training support receive a system transfer. At the other extreme are large institutional/programmatic users (e.g. Landsat-D Project, Climate Program, LACIE) who have strong DBMS technology needs but are capable of not only using existing technology but also of developing needed technology. Thus, NASA/OSTA needs to determine user requirements for remotely sensed data both for OSTA programs, and potential outside users as well, to properly plan for future data management requirements. While it is recognized that the establishment of requirements is frequently an iterative educational process, several basic points are evident today. First, users for many applications need to be able to access all data relevant to their application. Operationally they can do this either through a series of integrated data bases or through access to distributed data bases. Second, the ease with which DBMS technology can serve the range of applications is dependent upon the universality of spatial data storage standards and information extraction procedures. Thirdly, timeliness requirements for data delivery vary widely among users of the same sensor data, and the range in timeliness requirement and level of near-operational application should pace the DBMS systems technology development for a mission.
2. Transfer. NASA must effectively disseminate and transfer DBMS technology to a wide community of potential users if it expects to accelerate the use of its earth observation data and thereby encourage the development of operational missions. A wide spectrum of technology transfer activities are envisioned, including COSMIC,

the Regional Applications Program Centers, and documenting and widely disseminating generic case studies which include not only the software modules but also technical reports on the scientific rationale and implementation procedures used. The most effective presently available DBMS technology is associated with cataloguing systems which are presently available from industry. It was pointed out that microfiche technology can be interfaced directly with computer systems, representing a cost-effective means for information dissemination. Similarly, it was noted that network and hierarchical models for cataloguing DBMS frequently compare poorly with relational DBMS, because of their restructuring/rerunning requirements when new information fields are needed. While the establishment and maintenance of a catalogue DBMS will always incur expense for a new mission/facility, its implementation is now straightforward and should be subcontracted to knowledgeable industry suppliers.

3. Data Integration. User applications have come to increasingly require the integration of NASA earth observation data with non-NASA data and a variety of types of satellite imagery if the application is to be integrated with the existing information gathering functions of an agency. To date, NASA DBMS have been able to achieve spatial data integration for only a few specific case studies and often at great cost. The inability for a particular DBMS technician to perform the straightforward task of preparing a "packet" of information with ease is caused by many conditions beyond his control. The external conditions that impede DBMS technicians from performing their data integration functions were the topic of much discussion. Several key points were brought up:

a. There is a need for facilities and communications capabilities to allow remote data access and data base/data management systems sharing.

b. The data producer/data base owner should be responsible for data validation and data quality assurance, especially in a distributed system.

c. The problem of archival storage of spatial and related data by NASA is of great importance to many user applications. A rationale for archiving which considers periodic reviews of spatial and temporal resolution requirements for data would help avoid unnecessary loss of historical data and yet permit a logical reduction in storage volume as time progresses.

d. Strong requirements exist for the creation of a wide variety of standards to promote accessibility to existing data bases and to data bases still in the planning stage. For data bases there is a need for data compatability through the development of interface standards, and for data content translators among data bases. With this in mind, NASA should make its imaging data conform to standard map projections, and data file contents conform to ANSI standards. For data recording there is a need to improve the state of the art in high density tape hardware systems, to assure the highest available quality control of data in a data base.

e. We are presently not adequately assessing and estimating data volumes of existing and proposed spatial data handling projects. Procedures need to be developed which translate estimates into resource requirements and a responsible archiving policy.

f. Any application can have a full range of timeliness requirements and as near-realtime functions come to be used the DBMS hardware and software become more special purpose and focussed. As user applications are requiring near-realtime data with greater frequency today than in the past, the need for (d) standards and (e) procedures to assess data volumes becomes more acute.

g. NASA must develop a comprehensive spatial information system which is able to manage all data types (e.g. image/raster, grid cell, polygon, tabular). Earth observations systems data represents only a portion of the information needed by other agencies, and they therefore cannot to be expected to develop systems that are comprehensive.

CHARACTERISTICS OF A SPATIAL INFORMATION SYSTEM

One of the recommendations of the panel is that a multi-center activity be initiated to produce a comprehensive/unified plan for the development of a new generation of spatial information systems designed to integrate NASA and non-NASA data. Presently available spatial information systems are not generally applicable nor widely disseminated, with the result that NASA earth observations data input to a user application is frequently less than adequate. Characteristics that should be addressed are included below:

A spatial information system should accept and handle/process all data types, including: point, line, contour, polygon/boundary (including arc-segment, double digitized, and pen-up/pen-down digitizing convention formats), areas, raster/image data, tabular data.

A spatial information system should provide system level geographic analysis capabilities, such as: polygon overlay, point to surface and line to surface interface and data integration, rubber-sheeting/overlay of data sets, and spatial integration procedures that incorporate a variety of interpolation methods and sampling strategies used by various disciplines in mapping derived information.

A spatial information system should provide extensive reporting and automatic map generation as output product generation capabilities.

A spatial information system should provide standard interface capabilities to: applications software systems, image processing systems, standard statistical analysis packages, tabular data management systems, report generation systems, and polygon-based geographic information systems.

A spatial information system should be designed so that it can be used as a menu/command/procedure based system which supports interactive, batch and real-time operations. Furthermore, parts of the system which are hardware dependent should be isolated to low-level modules to ease transferability, and the system should be implemented to permit maximum ease of software transferability.

While it is recognized that such a system will in reality develop more rapidly in some discipline areas than others, and be implemented at different degrees of rapidity at different facilities, NASA/OSTA should make a commitment to support spatial information system architectures which conform to prescribed standards.

CHARACTERISTICS OF A COMPREHENSIVE BROWSE FACILITY

DBMS technology associated with cataloguing systems represents the most readily available data management software, and is well supported by private industry. It should be noted, however, that while the implementation of cataloguing function systems is now straightforward, the establishment and maintenance of a cataloguing DBMS will always incur expense for a new mission or facility. There is a further need to link the presently available and anticipated NASA and associated non-NASA earth observational systems data bases into a comprehensive browse facility to permit applications users to obtain complete and rapid inventories of available data. Several important characteristics for a comprehensive browse facility need to be adopted if the full potential for earth observations systems data use is to be realized:

1. For the higher order decisions associated with a comprehensive browse facility, network and hierarchical models for cataloguing DBMS may compare poorly with relational DBMS, because of the formers' restructuring/rerunning requirements when new information fields are added.

2. Microfiche technology can be interfaced directly with computer systems, and as it represents a considerable cost reduction in information dissemination should be incorporated as a primary mode of data base description.

3. It would be a mistake for NASA/OSTA to attempt to centralize the catalogue/browse facility with storage of the spatial data, but rather should support a distributed network of data catalogues and data bases.

4. The facility should provide a functional index aid to the earth observations data which in a human engineered query system would guide the applications user to <u>all</u> relevant data types that may support his information needs.

5. The facility should tap non-NASA data sources that have been identified as supportive of earth observations user applications.

APPENDIX: WORKSHOP RELATED MATERIAL

A. DBMS SURVEY QUESTIONNAIRE

B. AGENDA

C. WORKSHOP ATTENDEES

D. REVIEWERS UNABLE TO ATTEND

PRECEDING PAGE BLANK NOT FILMED

A. DBMS SURVEY QUESTIONNAIRE

1. Survey available cataloguing DBMS at your Center/Facility. Using a separate sheet of paper for each system:

   a) Note the mission(s), if any, that the system has been developed for.
   b) Note the system hardware configuration it operates under.
   c) Note the system software language(s) used in developing the DBMS and whether or not it is in COSMIC.
   d) List key bibliographic references (reports and articles) that describe the system or cognizant individuals at your Center/Facility.
   e) Briefly describe the DBMS functions and the system's current status.

2. Survey available satellite derived spatial data integration DBMS at your Center/Facility using the framework a) - e) in question 1.

3. Survey available satellite and non-satellite derived spatial data integration DBMS at your Center/Facility using the framework a) - e) in question 1. Responses here should include geographic information systems which may only peripherally include satellite data or which at this time have yet to directly interface with satellite data.

4. To the extent you are aware, list the user surveys, missions plans, and project study reports, which you expect to impact the data base management systems functions at your facility between now and 1985. List cognizant individuals at your facility where possible.

5. Note the missions and/or programmatic areas that are to impact the frequency and turn-around time for the following DBMS functions:

   a) Archival storage and retrieval
   b) Catalog browse file procedures and protocols
   c) Diverse data set integration requirements

6. Note the missions and/or programmatic areas that are to impact the following DBMS functions of:

   a) Analysis of diverse spatial data sets (e.g., image processing, graphic display, statistical/mathematical analysis)
   b) Accessing archived data that may be several months/years old
   c) Providing near-real time analysis of incoming data

7. For missions and/or programmatic areas that require DBMS development, chart the current/anticipated data flow through the system noting in list form:

   a) The kinds of systems hardware improvements that would increase the speed, analysis power, and storage capacity of your facility, (e.g., where would mass storage devices and micro-processors be appropriate?)
   b) The kind of algorithms/software that would increase the speed, analysis power, ease of data storage/retrieval, and ease of variable data set integration.

B. AGENDA

Monday Morning, March 5, 1979 (Location: Athenaeum)

8:30AM -	Welcome and opening remarks, N.A. Bryant and H.W. Shaffer
9:00AM - 11:00AM	Review Sampling of Existing Data Base Management Systems 1. Burt Horsted, The EROS Data Center Browse File and Digital Image Preparation Facilities. 2. John Lyon, the LACIE Data Base and Follow-on Requirements. 3. Pete Bracken, The Weather and Climate Analysis Facility Using AOIPS and Associated Systems. 4. Al Zobrist, The VICAR/IBIS Image and Non-Imaging Data Integration Software System.
11:00AM - 11:30AM	Discussion of Panel Session Grouping Arrangement and Topics to be addressed.
11:30AM - 12:30PM	Lunch

Monday Afternoon, March 5, 1979 (Locations: Athenaeum, Crellin 151, Thomas 08)

1:00PM - 5:00PM	Separation of the group into three panels, each with a chairman, to begin to address the following points: 1. Inventory of existing DBMS available within Space Community and their association with existing flight missions. 2. Survey of generic types of DBMS, i.e., a) cataloguing, b) satellite derived spatial data integration, c) satellite and non-satellite spatial data integration. 3. Review of generic components of DBMS (e.g. archiving, access protocol, hardware/software integration as a function of progressively near-real time analysis). 4. Listing of expected DBMS technological developments, by generic type, for 1985, 1990, 2000. 5. Priority of DBMS technology bottlenecks to be addressed by NASA community in facilitating implementations of DBMS requirements/standard expected by user community.

Tuesday Morning, March 6, 1979 (Location: Athenaeum)

9:00AM - 11:30AM	Review by panel chairmen of each panel's findings. Discuss progress, difficulties, further writing/discussion assignments for panels.
11:30AM - 12:30PM	Lunch

Tuesday Afternoon, March 6, 1979 (Locations: Athenaeum, Crellin 151, Stelle 14)

1:00PM - 5:00PM	Panels reconvene, incorporate additional information in open panel session and undertake individual writing as appropriate.

Wednesday, Morning, March 7, 1979 (Location: Athenaeum)

9:00AM - 12:00PM	Panel chairmen present results. Executive summary format and outline of topics addressed will be prepared for JPL personnel write-up during following week.
12:00PM	Workshop adjourn.

Follow-up work:

March 14, 1979	Draft Executive Summary sent to workshop attendees for comment, to be returned by March 21, 1979.
April 4, 1979	Transmit Executive Summary to Applications Data Service, GSFC, Review meeting of User Requirements.
March 21, 1979	Extended comments and bibliographic backup due at JPL for drafting of final report.
April 23, 1979	Transmittal of draft final report to attendees for comment, to be returned by May 1, 1979.
May 7, 1979	Review of report by selected non-federal agency personnel for comment.
June 4, 1979	Transmission of final report to NASA, with outside reviewer's comments appended.

C. WORKSHOP ATTENDEES

Pete Bracken
Information Systems Development Branch
Information Extraction Division
Goddard Space Flight Center

Nevin Bryant
Science Data Analysis Section
Earth Observations Division
Jet Propulsion Laboratory

Richard desJardins
Study Manager
Applications Data Service
Goddard Space Flight Center

Richard Durstenfeld
Mission Control and Sequence
Systems Design Section
Jet Propulsion Laboratory

John Hall
Water Pollution Data Systems
Environmental Quality Program Office
Langley Reserach Center

Burt Horsted
Computer Operations
EROS Data Center
U.S. Geological Survey

Jim Kibler
Earth Radiation Budget Data Systems
Langley Research Center

Don Lozier
Regional Applications Program
Ames Research Center

John Lyon
Systems and Facilities Branch
Earth Observations Division
Johnson Space Center

William Mitchell
Geographic Information Systems
Geography Branch
U.S. Geological Survey

Roland Nagle
Environmental Prediction Research Facility
Fleet Numerical Weather Center
U.S. Navy

Ron Pearson
Data Management
Earth Resources Laboratory

Tom Renfrow
Systems Analysis Section
Jet Propulsion Laboratory

Eve Schwartz
Environmental Prediction Research Facility
Fleet Numerical Weather Center
U.S. Navy

H. William Shaffer
Information Systems
Applications Division
Office of Space and Terrestrial Observations
NASA Headquarters

Ed Sherry
Mission Design Section
Jet Propulsion Laboratory

Walt Truszkowski
Control Center Operations Section
Mission Operations Division
Goddard Space Flight Center

William Whitney
Information Systems TEchnology
Jet Propulsion Laboratory

Al Zobrist
Science Data Analysis Section
Earth Observations Division
Jet Propulsion Laboratory

D. REVIEWERS UNABLE TO ATTEND

Jim Bagwell
Systems Concept Branch
Space Systems and Technology
Directorate
Lewis Research Center

Jack Estes
Land Resource Inventory
Renewable Resource Observations
Office of Space and Terrestrial
Applications
NASA Headquarters

Patricia Fulton
Topography Division
U.S. Geological Survey

Eric Greunler
Advanced Systems
Satellite and Missile
Organization
U.S. Air Force

Dale Lumb
Regional Applications Program
Ames Research Center

David Small
NOSS Data System
National Oceanic and Atmospheric
Administration

John Theon
Weather Program
Environmental Observations
Division
Office of Space and Terrestrial
Observations
NASA Headquarters

Doug Thomas
Data Systems Laboratory
Marshall Space Flight Center

NASA—JPL—Coml., L.A., Calif.