

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION  
FUNDAMENTAL RESEARCH PROGRAM  
INFORMATION UTILIZATION AND EVALUATION

Principal Investigator: John E. Estes

Co-Principal Investigator: Ludwig Eisgruber

WORKING GROUP

Glen Bacon

Ida Hoos

Nevin Bryant

Robert MacDonald

Christopher Clayton

Bruce Scheer

Ludwig Eisgruber

Ronald Shelton

John Estes

Charles Vars

Forrest Hall

NASA Contract Number NAS 9-16077

FINAL REPORT  
APPENDICES

JUNE 1, 1981



(NASA-CR-167591) NATIONAL AERONAUTICS AND  
SPACE ADMINISTRATION FUNDAMENTAL RESEARCH  
PROGRAM. INFORMATION UTILIZATION AND  
EVALUATION, APPENDICES Final Report  
(California Univ.) 132 p HC A17/MF A01

N82-24134

Unclass

G3/82 09815



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION  
FUNDAMENTAL RESEARCH PROGRAM  
INFORMATION UTILIZATION AND EVALUATION

Principal Investigator: John E. Estes  
Co-Principal Investigator: Ludwig Eisgruber

WORKING GROUP

Glen Bacon	Ida Hoos
Nevin Bryant	Robert MacDonald
Christopher Clayton	Bruce Scheer
Ludwig Eisgruber	Ronald Shelton
John Estes	Charles Vars
Forrest Hall	

NASA Contract Number NAS 9-16077

FINAL REPORT  
Appendices

June 1, 1981

## Table of Contents

Executive Summary . . . . .	1
Introduction . . . . .	8
Background . . . . .	10
Research Recommendations . . . . .	13

### Appendices in Separate Volume

Appendix A . . . . .	A1 - A18
Appendix B . . . . .	B1 - B17
Appendix C . . . . .	C1 - C49
Appendix D . . . . .	D1 - D18
Appendix E . . . . .	E1 - E5
Appendix F . . . . .	F1 - F41

## Working Group Summary and Conclusions

The material in this section summarizes the Working Groups discussions held at the workshops detailed in Appendix C of this report. It is important to note that the discussion held by the working group surrounding these presentations were far reaching and included some topics not specifically presented by individual speakers. As such, while the summary and conclusions here are presented with reference to specific workshops some conclusions presented summarize essential end-point thinking of working group members (i.e., these are conclusions on research issues as of the writing of this report).

### Asilomar Workshop on Information and Decision Processes

The fact that today's government policymaking process places increased emphasis on rigorous program evaluation. For example, Presidential Decision 54 (November 1973) requires that private users of future systems pay product prices sufficient to ensure maximum recovery of system costs consistent with the public good. Benefits to users must be evaluated against system costs. As such future satellite remote sensor systems will develop only if they can be economically justified (i.e., justified in terms of their ability to provide data which has potential value as information to users which can be effective input into the user decision system (information system)).

Asilomar Workshop participants agreed that no general theory or methodological framework exists for estimating the value of user decision-oriented information systems. A number of factors contribute to this methodological problem. First, there is no market price for the output of

most public information systems, and therefore value cannot be estimated in conventional ways. Second, the quality of an information system is based on its multiple characteristics, including accuracy, timeliness, reliability, continuity, and so on. Moreover, some types of information possess the characteristics of a public good (the use of the good by one individual does not reduce the amount available to others), and its private value may differ substantially from its social value. Third, many sources of information are available to decision-makers, few are regarded as certain, and choices among types of information to obtain and use are made within the uncertain environments that confront decision-makers.

Thus, the Working Group on Information Utilization and Evaluation has concluded that current knowledge and research provide little guidance for the evaluation of modifications in the public sector components of complex user application decision-oriented renewable resource information systems. As previously stated, the future capabilities of alternative remote sensing systems are all limitless; but the design of future satellite sensor systems in a meaningful fashion; however, is quite imperfect. Additionally, no generally accepted methodology exists for evaluating changes in information systems, structures, sources, and products. Thus, the Working Group recommends that an integrated research program be undertaken to:

- (1) Review and document existing renewable resource information systems and examine the potential of implementing advanced expert system approaches; and,



- (2) Develop the underlying theoretical and methodological basis for estimating the value of renewable resource data and information systems and their modification.

The review and documentation of theoretical and methodological research should proceed together and involve some of the same scientists; they are complementary endeavors. Thus, research should focus on systems to monitor cereal grain production, land cover/use, and soil moisture where remotely sensed data are or may be combined with data collected by other means.

The review and documentation process should not just include systems which obviously lend themselves to remote sensing applications. The option of combining remotely sensed data with data collected by other means should be studied. In fact, this latter possibility indicates why the review process should place special emphasis on why, how, and with what effects are multiple-sources of data used within existing information systems.

Uses of remotely sensed data typically involve data and information with a wide variety of attributes such as function, scale, timeliness, precision, and others. These attributes should be examined in the context of systems where tradeoffs among specific attributes and institutional constraints can be directly addressed. Systematic reviews and documentation are needed to understand how the attributes of data and information derived from it interact and affect analytical and decision processes.

Most studies on the value of information assume, for example, that a farmer, an elevator operator, and a buyer all use the same information in the same way, but evidence indicates that this is unrealistic. Some information is available as a public good, but it is not equally used by all who receive it. For these reasons, the identification of major

market and non-market users of information, their sources of information, and the adequacy of these sources become important if new systems are to be satisfactorily evaluated. Particular attention should be given to the documentation of how the behavior of public and private users and producers of information are changed when information is viewed as a public rather than a private good.

As future information systems are developed careful attention should be given to techniques emerging from the rapidly developing field termed "artificial intelligence." Research in this field is addressing new approaches to knowledge representation, language understanding, heuristic search, and other symbolic reasoning problems directly pertinent to many of the problems encountered by paradigms employed in existing information systems (Shortliffe, et al., 1979).

Analyses of many human decision-making processes suggest that as decisions move from simple to complex, the reasoning style becomes less algorithmic and more heuristic, while qualitative judgmental knowledge and the conditions for invoking it appear to increase. The artificial intelligence techniques used in "expert systems" more closely mimic this approach than previously used paradigms. In proposing research areas for information and decision processes, Bacon in a letter to Estes dated 2 June 1981, noted:\*

"I believe the most likely approach and therefore the first priority is that associated with expert systems. In order for the general public or laymen "government agencies" to obtain value from satellite data, a great deal of highly technical assistance must be given. I believe the expert system approach would be ideal and would make a very useful demonstration project. At the same time, this area is considered to be one of the more advanced areas of computer science discipline."

---

\*The entire letter from Bacon to Estes is included in Appendix D.

Research exploring the linkages and possible use of expert systems techniques are needed in at least three important areas:

(1) Discipline Concepts -

Traditional paradigms have no true "understanding" of the discipline involved. Although explicit decision trees can give decision theory programs a greater sense of pertinent associations, true discipline knowledge and the heuristics for problem solving are not explicitly represented nor used. So-called "commonsense" is often clearly lacking when existing systems fail, and this is often what most alienates potential users. In contrast, expert systems make explicit use of production rules that relate observations to associated inferences that may be drawn.

(2) Conversational Capabilities -

Both for capturing knowledge from collaborating experts and for communicating with users. The need is strong for the development of computer-based linguistic capabilities.

(3) Explanation -

Alternative systems seldom emphasize an ability to explain the basis for their decisions in terms understandable to the user. This can leave the user with no basis for deciding whether or not to accept the system's results and can lead to resentment at what could be perceived as an attempt to dictate the decision-making process. Once again in contrast, the heuristics of an expert system can often form a coherent explanation of system reasoning.

Working Group members feel strongly that work is required in this new and important approach and that there are potentially great benefits to be gained far beyond their application in the information and decision-making process area.

Toward developing the underlying theoretical and methodological basis for estimating the value of renewable resource data and information systems top priority is necessary. The long-run goal would be the formulation of quantitative models capable of contributing to the evaluation of trade-offs between existing and potential new information systems. Such models should be able to describe and explain existing systems and to provide the means to identify and estimate the magnitude of future consequences of new systems with alternative features. Specific research tasks include:

- \* Development of a general theory of the economics of information systems and their modification.
- \* Development of relevant measures for attributes of data and information (e.g., minimally sufficient statistics) and empirical methods for investigating information system performance.
- \* Development of models, at various levels of aggregation, of information systems with public, private, and international components.
- \* Investigation of analytical systems and processing strategies that pertain to the extraction of information from data based on disparate concepts.
- \* Analyses of data and information dissemination systems, including studies of accessibility and decentralization, economies of scale, confidentiality and property rights, public good



feels that this type of analysis could best be accomplished from a fundamental perspective if examined within the context of potential implementation of expert system approaches to the generation of information from remotely sensed data.

#### San Jose Workshop on Data Base Management and Use

During the San Jose Workshop on Data Base Management and Use it was discovered that the computer science community beyond NASA's direct concern are working on improving our ability to handle larger files and perform special purpose (e.g., array) processing to increase the speed of image.

There is currently a major commercial motivation for highly integrated computer assisted design and manufacturing systems that require interaction with graphical information in a manner similar to the useage of remotely sensed image information. General purpose data manipulation and accessing facilities for these systems should provide the base for NASA facilities in the future.

The potential exists for major breakthroughs changing the basic structure of data base management systems as they exist today. A number of Working Group and workshop participants felt that it is the responsibility of NASA and the funders of the Workshop (with respect to the Renewable Resources Program in particular) to make their needs understood so that the computer science community can focus on improvement. Today, data base management systems are primarily designed to serve the banking/accounting community. Special purpose hardware systems are also being developed for military application which are near- and real-time oriented. In addition, federal agencies are cooperating with the commercial community and the American National Standards Institute (ANSI) on developing data handling standards and interchange formats. By focusing research on data base

management through systems level demonstration projects NASA fundamental research program participants should provide a solid framework for the Renewable Resources Program to articulate both its immediate and long range needs to the research community and the commercial sector.

As discussed here system level demonstration projects involve the validation and testing of data base generation and use scenarios, and can therefore be considered orthogonal to technology and technique development programs being pursued by NASA/OAST, the Data Systems Branch within NASA/OSSA, and the general communications field. Systems level demonstration projects sponsored by the Fundamental Research Program would reveal current and potential problems associated with data handling gaps and bottlenecks for applications and system tests applied to resource inventory and modeling and serve to illuminate the potential for the integration and implementation of artificial intelligence/expert system approaches to resource management problems. In addition, renewable resources should keep abreast of work in the Data Systems Branch on a Transportable Applications Executive (TAE) as this concept has important implications for furthering the combined research potential of the NASA center. Such fundamental work as TAE can improve NASA's overall potential for advancing the application of advanced image processing to a wide variety of studies.

Working Group members feel that in the area of Data Base Use and Management that the case study (and scenario) approach is a practical means for determining essential components for data base management and use as they are applied to earth observations applications. We feel three general types of case studies, emphasizing present and future sensor platforms, could be employed to analyze a range of potential applications. Suggested topics are:

- a) crop yield forecast in a foreign country;
- b) watershed runoff forecasting;

and c) prediction of changes in land use within a region. Each of these studies can be thought of as exercising different aspects of data base management technology being developed by the communications and computer industries. They will also test the basic assumption that conceptual models designed to analyze a process and derive a prediction can in fact incorporate satellite sensor data efficiently and effectively to accurately address a problem of general concern to users. Each of the three suggested case study areas are being actively researched by government agencies and the private sector today. Thus, good bench marks for timelines, accuracy, and level of detail required for the end-product exists. The massive amounts, and the specialized requirements for calibration, interpretation, and formatting of satellite sensor data have in past NASA application demonstrations become the overriding technical concern. As a result, for each case study proposed the following elements need to be investigated:

- 1) Impacts of data base structure on sensor and ancillary data input and archiving formats;
- 2) Impacts of and architectures for query capabilities and processing rates on system development and use;
- 3) Impact of data availability, archiving, and opportunity costs of data storage; and,
- 4) Impact of decentralized data bases/information systems on cost/benefit ratios for systems use.

In addition, technology research parallel to that being undertaken in other branches of OSSA and OSTA needs to be undertaken to assume the availability of viable processing strategies in future systems capabilities to integrate disparate data types. Key issues include:

1. Understanding generic functions in spatially oriented data manipulation.

These include data capture, data registration, and data analysis. Data capture involves the encoding of image, vector, and tabular data types which represent or are keyed to areas on the ground. Data registration involves both the conversion of satellite and ancillary data products to a given map projection, and assurance that the information content of a data set is properly represented for the map scale used in the analysis. Data analysis involves the combination of one or more of the following primitive functions: a) Given a point and a district, does the point lie within the district. b) Given a point and a district file, which district contains the point. c) Given a particular district in a district file, what are its neighbors. d) Given a district file and an area classification file, what are the acreages of each area classification in each district. e) Given two district files, one major and one minor, what are the proportions of each minor district in each major district. f) Given a district file and a line segment, what are the mileages of the line segment in each district. g) Given a point p and a point file, what is the point in the point file nearest to p. h) Given a point p and a point file, what is the distance from p to the nearest point in the point file. i) Given a point and a line segment file, which line segment passes closest to the point. j) Given a density map and a district file, what are the volumes in each district (spatial integral of density). k) Given a district, what is the centermost point (an inside point which is farthest from the boundary).

2. Effects of positional accuracy on estimates derived from multiple data planes. All mapping, whether satellite derived or obtained by conven-



tional means involves global and local positioning errors. When products are derived from the analysis of several mapped phenomena, the error is cumulative. Investigations need to be undertaken to determine:

- a) The sensitivity of models to positioning errors in input data sets,
- b) The potential for the high resolution satellite data to improve spatial integration/mapping functions such as trend surface mapping and development of multi-stage sampling designs.

3. Investigate the performance and capacity requirements for the large record size and special purpose processing required for imaging and geographic applications. Is there a need to change the basic structure of Data Base Management Systems? Or, can we expect fundamentally different approaches to satellite data use in the future, such as sub-scene sampling for trend analysis rather than large area inventories? What would be the impact on data administration and use for expected scenarios of satellite data use determined by selected case studies.
4. Examination of possible and probable future environments which will impact renewable resource data base management and use. Factors such as timeliness and repetitiveness of satellite overflights will continue to be essential drivers of satellite systems. All stages in ground data acquisition, calibration, geocoding, archiving, transferring, and analysis need to meet the satellite overflight requirements if an end product is to be provided. Given certain assumptions in future analysis requirements, what are the key stages in data preparation which need to have improved through-put? How would each stage in data preparation and analysis best be served (i.e., centralized or decentralized function)? Are basic structural changes implied if

a move from research demonstration to operational capability is to be achieved?

#### Houston Workshop on Information Performance

Summarizing the discussions held after the presentations during the Houston Workshop on Information Performance several important points were stressed. Professor Jensen made the point that if we want agencies such as the USDA's Statistical Reporting Service to change over and employ new procedures based in some part on the analysis of remotely sensed data then we need to come to understand the sources and magnitudes of errors associated with all parts of the sampling systems to be used by such agencies. This point was echoed in subsequent conversations held with George Rosenfield of the U.S. Geological Survey in Reston, Virginia (see correspondence Appendix D). In addition, Gaylord Warden stressed that while the office of Federal Statistical Standards and Policy (OFSSP) criteria are appropriate to natural resource statistical systems, neither the institutions which collect such information nor the concepts employed in the sampling systems are sufficiently developed to permit detailed tradeoff and cost benefit analysis which could be considered in any way comparable to those undertaken for economic and other Federal statistical systems.

A number of participants discussed the need for an improved method for getting at the question of user data requirements. Comments such as Clough's: A basic examination of appropriate methods for determining what agencies and institutions are trying to achieve must be accomplished before any bona fide analysis can take place; i.e., optimization cannot occur until we understand the nature of the market. O'dell talked of the potential for the use of case studies and/or the value of a series of well documented

data sets. These could be employed to test and improve statistical sampling strategies or to develop new strategies which are more robust and sensitive to environmental variations. Such new or improved strategies are necessary because for many renewable resource survey problems classical statistical techniques are "nearly applicable." O'dell employed the phrase "nearly applicable" because most of the assumptions upon which classical statistical procedures are based are violated in their application to specific resource inventory problems. This is important as while sampling is still a major concern a number of participants felt that methods for achieving improved verification of local identification and classification in a non-sampling approach is of far more importance.

There was also a considerable discussion of the nature of the accuracy question in remote sensing applications. Cobberly pointed out that there are many levels of accuracy assessment needed. We need to study the accuracy vs. information questions, as well as an analysis of accuracy vs. information, content vs. decision-making. The whole question of original data products and how this fits into a total systems accuracy concept in remote sensing must also be addressed. In addition, it was felt that processing accuracy must be addressed in terms of levels required within given time constraints imposed by a decision system. This led to a comment by Houston that four major factors seemed to be recurring in the discussion which had considerable relevance to performance and required more study. These factors are accuracy, timeliness, applicability and affordability.

Finally, a number of study areas requiring fundamental research were put forward based on these discussions. These would include studies on the:

- content of future uses of remote sensing systems, including the needs for truly global resources information systems;

- need for detailed censuses of resources vs. sampling of given resources should be assessed;
- development of concepts to facilitate tradeoffs and cost benefit analysis of economic and other federal statistical systems. These studies should address questions of privacy, costs, and accuracy verification;
- use of remote sensing to improve the accuracy of local statistics;
- new approaches to sampling more responsive to variations from natural (normal) conditions;
- resolutions necessary for the production of key indications of important resource conditions.

Here again the context of future uses of remote sensing including the needs for global information systems can be viewed from a scenario or case study approach. Research here should address the range of socio-economic and geopolitical issues within which future global resources information systems will operate. Remote sensor systems of the future can serve a variety of user groups. The combination of artificial intelligence for on-board and ground processing and analysis of the data/information produced by these systems for input to potential global resource planning and management issues offers an exciting and potentially rewarding area for future study.

To date most large scale resource applications of remotely sensed data have involved sampling in some form. Yet, resource managers typically require maps as inputs in their decision process. This question of techniques and methodologies for providing a detailed census in cartographic form for a given resource should be examined. Issues which must be addressed include scales, geometric corrections registration, projections, transformations, classification procedures, and classification and locational accuracy verification



procedures. In this area what must be looked at carefully is the potential for achieving mature mensuration status for future satellite remote sensor systems.

The need for studies directed towards the development of concepts to facilitate tradeoff and cost benefit studies of economic and other federal statistical systems could potentially be accomplished as part of the studies aimed to develop a general framework for assessing the value of an information system which was discussed earlier. Included in these studies should be questions of privacy, costs, public good, accuracy verification, utility factors, and the development of tradeoff options for obtaining a "minimally sufficient statistic."

In crop estimation and forecasting of acreages and yields often remote sensing data are regarded as merely another source of information that can be used to improve the precision of current methods, if it can be shown that it would be useful for the purpose. Statistically, ordinary local sample data can be made more effective for estimation if auxiliary data can be found that is functionally related to the primary data and is available over the whole universe of interest. In the simplest case, if the directly obtained ("ground truth") data and the auxiliary (remote-sensed) data are correlated (either positively or negatively) then the auxiliary data are regarded as statistically useful for this purpose. If the remote-sensed data are available at no cost then it would be irrational statistically, to reject the use of those data in crop estimation, for those crops processing non-zero correlations between the two kinds of observations. Hence apart from technical details and costs, the determination of these correlation coefficients, the conditions of their measurement, their stability over time, etc., are required before remote-sensed data can be expected to be accepted for crop

estimation. If the correlations exist, then the intercept and slope coefficients of regression,  $\alpha$  and  $\beta$ , are of interest and how stable they are over species, districts, time, etc. If we presume that the values of these coefficients are peculiar to each of these situations and therefore, for suitable accuracy, must be determined for each estimating situation. If they were constant then, once determined, the primary data ("ground truth") can be dispensed with. Since  $\rho$ ,  $\alpha$  and  $\beta$  are also dependent on the interpretation ("classification") of remote-sensed data and the size and type of ground truth unit used, such interdependencies suggest that a set of coefficients, and the nature of their dependencies on factors affecting them, would be very useful for evaluating the effectiveness of remote-sensed data for crop estimation. Perhaps such information exists now in some form. However, if uses are to be made in any part, or in every part, of the world a systematic collection of these "facts" would be very helpful indeed.

Many sampling approaches employed in the analysis of remotely sensed data of renewable resources are based on concepts which involve the calculation of means from records acquired over a given time frame. Yet in many instances it is the short-term episodic events, whose true impact may be masked in normative data, which can produce disastrous consequences for resources. Research into the best ways to account for these occurrences must be conducted if we are to design sensor systems capable of providing early indications of impending problems so that effective remedial management action can be taken.

Somewhat related to this is the question of the best methodology for determining the optimum resolutions necessary to produce key indications of important resource conditions. Resolution is an important concept in remote sensing; current methods of testing the information which can be

extracted from remotely sensed data of varying resolutions leave much to be desired. Methods must be developed wherein optimal tradeoffs between information extraction potential and varying degrees of resolution are clearly laid out. We must know that when we want to survey the forest that we may not need to see the limbs on the trees. Yet, we must constantly be aware of the question as to whether multiple uses demand the determination of a minimally sufficient resolution and account for the tradeoffs involved in producing this data.

After briefly summarizing from all working groups the more specific research areas which should be addressed in the areas of data and information performance along with data and information attributes, the panel feels the following research topics could profitably be pursued:

#### DATA AND INFORMATION SYSTEMS PERFORMANCE

- Examine Fundamental Aspects of the Accuracy of Products of Remote Sensing
- Develop Relevant Measures which Characterize Data/Information (e.g., Minimally Sufficient Statistic)
- Develop Procedures Necessary for Renewable Resources Satellite Sensor Systems to Achieve "Measurement Systems" Status for Given Applications
- Continued Examination of Fundamental Aspects of Sampling Theory as they Apply to Spatially Distributed, Temporally Varying, Renewable Resources Parameters; at Local, Regional, National, and International Scales. Key Issues Include:
  - greater sensitivity to deviations from "normal conditions"
  - identification of key parameters and determination of correlation coefficients between remote sensor scene derived data and ground conditions for use when no independent source of verification exists
  - explore the potential of nonparametric test of data

#### DATA AND INFORMATION ATTRIBUTES

- Develop Improved Techniques for Measuring Tradeoffs Between Use of Existing and New Information Systems for Renewable Resource Decision-Making
- Investigate Impacts of Timeliness, Reliability, Accuracy, and Assured Product Delivery on Cost Benefit Potential
- Examine Alternative Means of Assessing the Public Good and Multiple Use Aspects of Data/Information
- Explore Use of Existing Collateral Data to Reduce Requirements for Sensor and Ground Truth Data Acquisition

In conclusion, an important point which has reoccurred in our discussion again was raised by non-working group workshop participants: there is a lack of a theoretical framework for evaluating the products of remote sensing. This question must be addressed. As Clough pointed out, an analysis of this research area must address a variety of issues including credibility, public good, policies and standards, real time and archival temporal requirements, uniformity and standardization, data integration and more.

It is a big task. It is a task, however, which participants felt must be strongly addressed if we are to truly begin to realize the information potential inherent in remotely sensed data.

The material contained here is an outgrowth of a series of studies to develop a program of fundamental research which is being sponsored by the Renewable Resources Branch of the Earth and Planetary Exploration Division of NASA's Office of Space Science and Applications (OSSA). The purpose of these studies is to define a three-to-five year research program to broaden and strengthen the scientific base for future applied research and development programs directed at better use of aerospace and remote sensing in monitoring the Earth's renewable resources.

The future potential of aerospace remote sensor systems is truly extraordinary. The opportunities they provide, however, will require public decisions concerning satellite sensors, ground facilities, and analysis systems of great complexity and cost. The potential benefits of future systems are large, but similarly large are the data management and storage problems they generate. Various systems could provide information to address important national and global resource planning and management concerns, but they would also create legal, social, and economic problems of national and international significance.

OSSA currently has a number of applied research and development projects (e.g., AgRISTARS) which employ essentially existing knowledge to design, engineer and demonstrate the capabilities of aerospace remote sensing as a valuable resource management information source. The basic research program is intended to compliment these projects by focusing on concepts and issues at the frontiers of the relevant sciences. A successful program will provide the essential building blocks for future applied projects to support the

design of sensors, ground facilities, and institutions to facilitate national and possibly increased international use of remote sensing systems in the late 1980's and 1990's. The research will be conducted by scientists who are knowledgeable concerning current and potential applications of aerospace remote sensing technology to renewable resources. Research results are expected to be communicated worldwide by the publication of papers in recognized scientific journals and other types of appropriate publications.

The Program Definition Study of Basic Research Requirements in remote sensing is divided into four general areas of study:

- (1) Scene Radiation and Atmospheric Effects Characterization;
- (2) Pattern Recognition and Image Analysis;
- (3) Electromagnetic Measurements and Data Handling; and,
- (4) Information Utilization and Evaluation.

This document represents the final report of the working group on information utilization and evaluation. Each of the four research area working groups were composed of scientists and practitioners from universities, research institutions, industry and government. These groups were formed to plan and conduct a series of workshops to identify critical research topics and to disseminate knowledge of current and projected applied research and development needs. Each working group is developing a prioritized list of research topics and a preferred method for implementing and coordinating its recommended program of basic research. Each group leader or leaders is or has been supported by selected members of the working group who are responsible for documenting the group's findings and recommendations in a report to NASA.

The study undertaken by NASA is based on the premise that the development of a successful basic research program requires careful planning and communications. Planning and the ultimate success of the research effort can be enhanced by extensive use of scientists and practitioners with different interests who are familiar with potential applications and problems with information based on remotely sensed data. Communication between NASA personnel and the research and user communities is more likely to identify worthwhile research topics and stimulate the scientific community to engage in such research than a purely internal effort by the agency. Therefore, NASA asked working groups to conduct a series of workshops where qualified experts from universities, research institutions, private industry, and government hear and discuss one another's views concerning the current state of knowledge, on-going work at the frontiers, and critical areas where research could produce significant results. The workshops have and are providing opportunities for working groups to draw broadly on the knowledge of the scientific community and deepen their understanding of relevant issues before preparing recommendations to NASA.

The working group on Information Utilization and Evaluation organized and then met in organizational and planning sessions in April, 1980 in Santa Barbara, California. Members of the Working Group Steering Committee are seen in Table I.

TABLE I

INFORMATION UTILIZATION AND EVALUATION WORKING GROUP  
STEERING COMMITTEE MEMBERS

GLEN BACON, IBM Corporation

NEVIN BRYANT, NASA Jet Propulsion Laboratory



CHRISTOPHER CLAYTON, University of California

LUDWIG EISGRUBER, Oregon State University

JOHN ESTES, University of California

FORREST HALL, NASA Johnson Space Center

IDA HOOS, Space Science Laboratory, University of California

ROBERT MACDONALD, NASA Johnson Space Center

BRUCE SCHEER, The Planning Economic Group, Boston

RONALD SHELTON, Michigan State University

CHARLES VARS, Oregon State University

A subsequent meeting was then held in July 1980 at Johnson Space Center to: (1) identify those areas where basic research is appropriate and (2) plan a series of workshops to address these research needs. A listing of participants for each information utilization and evaluation meeting and workshop can be found in tabular form in Appendix E of this report.

Following these initial meetings of the Information Utilization and Evaluation Working Group a series of three major workshops and a special session (held in conjunction with a National Meeting on Remote Sensing for Resource Management) were held. The first session was conducted in September 1980 on Information and Decision Processes at Asilomar, California, and a special session in Kansas City in October 1980; the second workshop on Data Base Use and Management in San Jose, California in January 1981; and, the third workshop on Data and Information Performance in Houston, Texas in February 1981.

In February 1981, just prior to the final workshop, a third planning session was held in Houston to discuss progress that had occurred to date and to plan for the final workshop along with how this final report would

be assembled. As previously stated, participants at these meetings are listed in Appendix E of this report. Table II, however, presents a breakdown of the individuals attending each of the meetings at the Information Utilization and Evaluation working group by employment.

TABLE II									
BREAKDOWN OF TOTAL PARTICIPANTS BY EMPLOYMENT									
	PL1	PL2	AW	KC	SJW	PL3	HW	OTHER	TOTALS
PRIVATE INDUSTRY		2	2	3	7			2	16
FEDERAL GOVERNMENT	2	3	3	2	7	2	5	1	25
STATE AND LOCAL GOVERNMENT			3	1					4
UNIVERSITY	8	5	8	4	3	4	7	4	43
TOTALS	10	10	16	10	17	6	12	7	88

Since certain individuals may have attended more than one meeting a breakdown of individual participants by employment is also seen in Table III.

TABLE III	
BREAKDOWN OF INDIVIDUAL PARTICIPANTS BY EMPLOYMENT	
PRIVATE INDUSTRY	14
FEDERAL GOVERNMENT	16
STATE AND LOCAL GOVERNMENT	4
UNIVERSITY	22
TOTAL INDIVIDUALS PARTICIPATING	56

These tables illustrate that while overall state and local governmental representation in the working group was low, participation by University, private industry, and federal government representatives was relatively evenly split.

The initial planning meetings were basically a time to exchange ideas and formulate a perception of the task ahead. Members of the Working Groups Steering Committee found in their discussions that although there is no single theory of information, they could come to some agreement on a general framework encompassing the major features of information systems. The major elements of the framework, as well as a paradigm within which to evaluate information systems problems, are sketched here to provide a understanding and guide to the concerns which underlie the recommendations for research presented later in this report.

At their most basic level information systems are designed to help decision-makers make decisions toward solving problems. This indicates why theoretical or conceptual work on information should focus to some extent on problem-solving. Since problem-solving requires some understanding of the nature of the problem to be solved, information can be viewed as the product of some process of inquiry. From this perspective, any knowledge about the problem is dependent on the system of inquiry used to obtain that knowledge. However, to contribute to solving problems, an information system must represent the reality relevant to those problems and the decisions to be made. As a consequence, the goals and values of the decision-maker affect the design of an information system, and the decision-maker must be considered as a major component of the information system.

The problems of information system design and evaluation are relatively simple if the decision-maker confronts well-structured problems, but very troublesome if problems are ill-structured. Problems are well-structured when known relationships exist between actions, outcomes, and their value to the decision-maker. The mode of inquiry is straightforward if problems

are well-structured, and the evaluation of information systems addressing such problems pose few difficulties.

In contrast, ill-structured problems are those where knowledge of states of "nature," acts, outcomes, and even the values of the decision-maker are not known with confidence. The uncertainty inherent in ill-structured problems often mean there is difficulty in defining: (1) the reality and problems that information systems must address; and, (2) the scope of information systems themselves. Owing to these problems of definition and scope it is our opinion that a multidisciplinary systems approach to problem-solving can provide the surest foundation for the information systems paradigm presented here. As discussed in this report information systems are broadly viewed as integrating the results of many different modes of inquiry to solve problems. Decision-makers apply heuristic as well as algorithmic approaches to the acquisition of information and decisions.

At this point it is important to note that data is often incorrectly equated to information. The paradigm outlined below: (1) distinguishes between data and information; and, (2) relates data collection to the inquiry and analytical processes used to solve problems and to reach decisions. A data system seeks to represent reality empirically. Reality is complex, and data systems commonly report on multiple dimensions of the following fields of observation: (a) internal operating environment; (b) external operating environment; (c) external peripheral environment; and, (d) self-representation. Data can be quantitative or qualitative.

Measurement occurs only after deciding what to measure. Superior sampling procedures and measurement techniques have little value unless the data generated are based on meaningful concepts that are relevant to the decisions

being made. Therefore, data collection involves three steps: (1) conceptualization; (2) operationalization of the concept; and, (3) measurement.

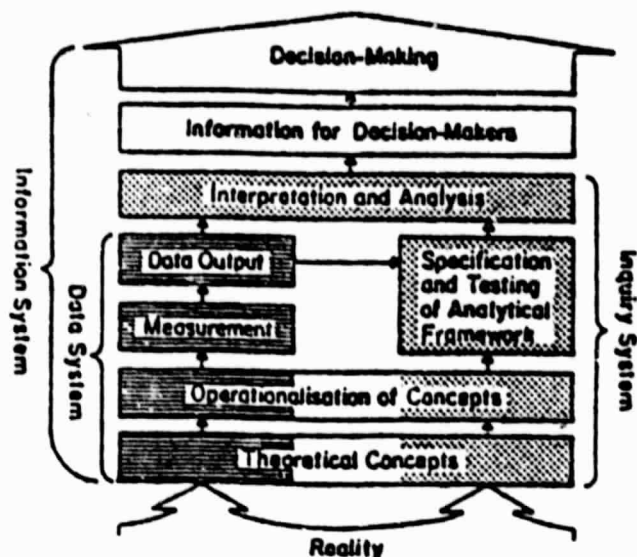
Bonnen (1977) emphasizes that the statistical reliability of a data system has three corresponding meanings. The first is conceptual reliability, (i.e., do the concepts accurately represent reality and pertain to the decisions being made). The second is operational reality (i.e., the categories of empirical variables should be highly correlated with the conceptual representations of reality). The third is measurement reliability as usually defined by statisticians.

Data systems produce data not information. To become information, data needs analysis and interpretation to acquire meaning for use in decision-making. As Riemenscheider and Bonnen (1979) observed, "an information system is a process which imposes form and gives meaning to data." An information system has three components: (1) a data system; (2) the analytical and interpretative capabilities to convert data into information; and, (3) the decision-maker. This is depicted in Figure 2.

Information systems have multiple components because decision-makers seek different types of information on the various "fields of reality" with which they must deal. Conceptual obsolescence is a problem for information systems because changes occur in reality and the agenda for decisions. The memory of an information system is invariably more appropriate in part to some prior reality and agenda than it is to the present. Modification of information systems attempts to overcome recurring mismatches of reality, decision, and available information.

Communications within and between organizations require interpersonal transmission of data and information. Problems of codes, formats, and "noise"

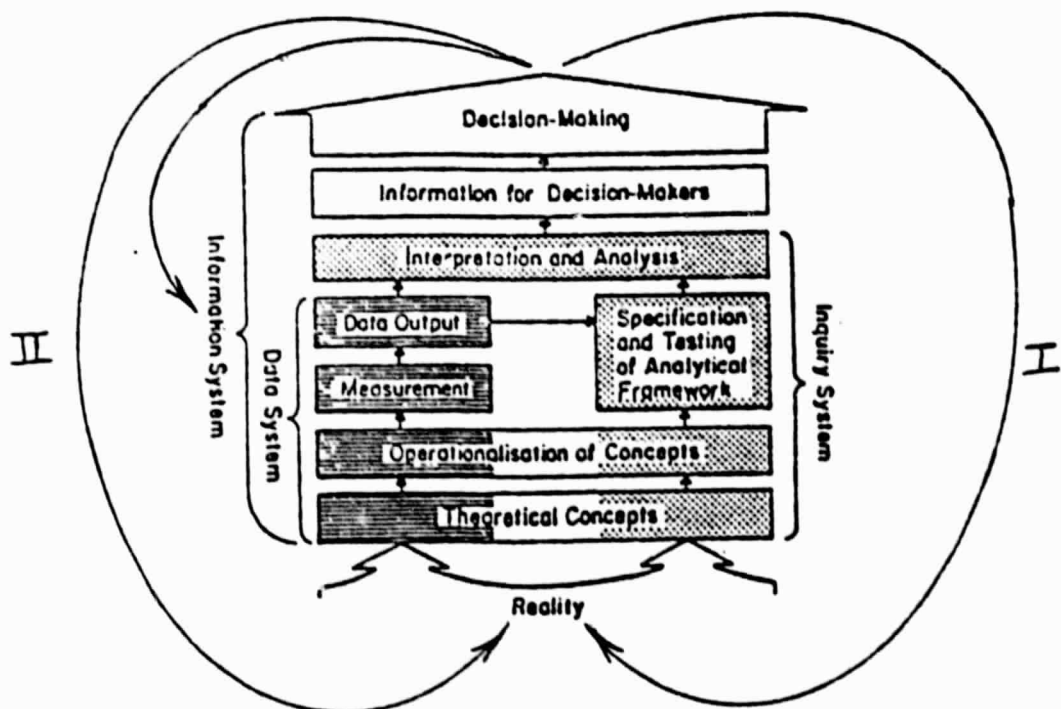
Figure 1. An information systems paradigm.



Reality: Fields of observation:

- Internal operating environment
- External operating environment
- External peripheral environment
- Self representation

Figure 2. Two modes of information use.





are widely understood, but those deriving from different conceptions of reality, problems, and needs are not so well understood; although they can significantly affect the use and value of information systems.

Information systems are used in both an ordinary and an extraordinary mode. The first mode occurs when information accords reasonably well with expectations and only routine adaptations are called for within the operating environment. In ordinary circumstances recurrent mismatches between observations and expectations do not occur, and there is no need to change the information system. Path I in Figure 3 depicts the ordinary mode of information use.

In the extraordinary mode, however, information does not accord with expectations concerning the operating or peripheral environments or itself. Non-routine actions are taken and learning, development, and innovation are then initiated to transform the information system. Components of the system are modified, deleted, or added in the extraordinary mode of use. Path II in Figure 2 illustrates the extraordinary use of information to alter the information system itself.

The distinction between the ordinary and extraordinary use of information is significant but difficult to draw precisely. For situations exist where both modes are carried on simultaneously, and learning can occur when an information system operates in the ordinary mode. Further, some systems may have limited capacity to review their own performance and to decide how to modify themselves.

Based then upon our general agreement the Working Group's Steering Committee put forth four specific areas of critical importance to the area of information utilization and evaluation as we understand it. These are

four areas where we feel fundamental research must be conducted if we are to truly advance the applications potential of future satellite remote sensor systems. These areas identified and addressed in the discussions are as follows:

- (1) Information and decision processes;
- (2) Data and information attributes;
- (3) Data base management and use; and,
- (4) Data and information systems performance.

Briefly expanding upon the core of the rationale for the choice of these areas the Working Group Steering Committee members presented at the planning meetings their feelings in respect to:

(1) Information and Decision Processes

Present resource decision systems and their actual potential use of remotely sensed data are not well understood. Fundamental research must focus not only on the physical design and image acquisition characteristics of perspective remote sensor systems but on current and probable future institutional arrangements, decision processes, and their information requirements if rational choices are to be made among potentially competing remote sensing systems.

Information components of decision systems must be reviewed to specify and understand the characteristics, credibility, and utility of currently used data from all sources. It is not feasible to analyze all systems where remotely sensed data are currently being or potentially could be used. Therefore, the Working Group feels that research should focus on a limited

number of application scenarios where remotely sensed data are or may be combined with data collected by other means for use in the decision process.

Research efforts in this area should address both private and government decisions with national and possibly international implications. Utilization of biological and social science models in the decision process should also be analyzed. Actual practice in the use of scientific models and data could be compared against best possible practice given the current state of knowledge. These comparisons may then be employed to establish the quality of data and information currently used in decision-making. Such comparisons could also identify where modified information systems and scientific models that rely on remotely sensed data could be employed to specify the potential benefits which might accrue from improved sensor system capabilities and how those advances could improve decision processes in the next 10-20 years. Potential improvements would be judged by criteria appropriate to evaluating the various consequences of monitoring the application scenarios chosen on a local, state, regional, national, and/or international basis.

(2) Data and Information Attributes -

Users of remotely sensed data typically utilize data and information with a wide variety of attributes. Attributes such as function, scale, timeliness, precision, and others should be examined in the context of systems where tradeoffs among attributes and institutional constraints can be directly addressed.

Functional requirements of remotely sensed data are important and the desired scale, spatial resolution, and other attributes of the desired information will determine the resolution requirements of future sensors. Will system users want to see a forest, a timber circle, a stand of timber, or individual trees? How will these elements be related to data needs? How will remotely sensed data interface with data from other sources? Fundamental research is needed to understand how the attributes of data and information derived from it interact and could affect decision processes in the future. Therefore, the attributes of data and information will be examined and evaluated within expected future decision environments concerned with cereal grain production, land use/cover, and soil moisture. Studies will address the wide range of data potentially produced by remote sensing technology.

(3) Data Base Management and Use -

Many resource managers and scientists believe that in the future various renewable resource management questions will require new forms of data base management (storage, manipulation, and retrieval) and uses (data outputs) to function efficiently. Research in this area will analyze the basic factors that affect:

- (a) Input of remotely sensed and other data into information systems;
- (b) Storage, retrieval, and management issues;

(c) The interaction and transmission of remotely sensed data and other data once they have been entered into a system; and,

(d) The types of products demanded of information systems.

With respect to the first and second areas, research will seek answers to basic questions concerning data types, quantities, formats, storage, accessibility, retention spans, and accuracy. In the third area, research will address how resource management data systems can be merged, manipulated, analyzed, and interpreted once remotely sensed data is incorporated into the system. What are the models employed, the types of data they require, how can these data be combined, surveyed, and sampled? The final area involves examination of the factors affecting the presentation of information to its users. Topics will include hard vs. soft copy, tabular, graphic, carto-graphic, and combination products, as well as analyst-decision-maker interactions.

(4) Data and Information Systems Performance -

This general area is concerned with investigating approaches to performance evaluation of the data and information generated by remote sensing mapping and inventory systems. The history of remote sensing reveals recurring problems of (a) defining performance parameters meaningfully to both technology users and developers, and (b) establishing approaches for estimation of those parameters in constrained situations where independent reference data may not be available.

To establish the meaningful performance parameters both technology developer and user interests must be taken into account. While as the developer may be primarily concerned with the accuracy of the estimates of the subcomponents of the estimation models, the user may be more generally concerned with questions revolving around the timeliness, utility, and reliability of the estimates, their cost, and interpretability and familiarity of the products of the technology, along with ultimately understanding the reports, etc. Additionally, the technology developer may be concerned primarily with quantitative measures of performance, while the user will be concerned with more qualitative social and economic factors.

It is difficult to estimate the impact of performance parameters once they have been established. In most cases remote sensing mapping and inventory systems generate data and information which are independently available from in-place conventional systems. In all cases, the data and information from the remote sensing system are being evaluated for that improvement in either timeliness, spatial dimension or accuracy as compared to such conventional systems. This then raises the fundamental question of how the performance of the remote sensing system is to be evaluated (i.e., to what standard are its outputs to be compared?). How are inferences as to its performance to be made in situations where independent data is either unreliable or non-existent? What should be the measures of performance?



The sections which follow present a brief synopsis of the individual presentations at the workshops which addressed these topic areas. This is followed again by a summary and conclusion sections and recommendations.

## BACKGROUND SECTION

### REFERENCES

1. Bonnen, J.T. Assessment of the Current Agricultural Data Base: An Information System Approach. In: L.R. Martin (ed), A Survey of Agricultural Economics Literature. Minneapolis: Univ. of Minnesota Press, 1977, pp. 385-407.
2. Dunn, E.S. Social Information Processing and Statistical Systems - Change and Reform. New York: Wiley-Interscience, 1974.
3. Miller, Thomas A. "Value of Information--A Project Prospectus." Mimeographed. U.S. Department of Agriculture, Economic Research Service, Commodity Economics Division, Agricultural Policy Analysis Program Area.
4. Riemenschneider, C.H. and Bonnen, James T. "National Agricultural Information Systems: Design and Assessment." In: M.J. Blackie and J.B. Dent (Editors), Information Systems for Agriculture. London: Applied Science Publishers, 1970.

The following material presented in chronological order provides an overview of the important points presented by speakers at the three information utilization and evaluation workshops. In addition also presented are the salient points acquired by steering committee members in special sessions and discussions held with participants at Kansas City in conjunction with the Soil Conservation Society of Americas meeting on Remote Sensing for Resource Management.

#### Asilomar Workshop

The first workshop conducted by the Working Group on Information Utilization and Evaluation was held at Asilomar, the University of California Conference Center at Monterey, September 16-17, 1980. Attendees are listed in Appendix A while the topics and discussants are seen in Table WR-I:

#### TABLE WR-I

##### Titles of Presentations and Speakers at

##### INFORMATION AND DECISION PROCESSES WORKSHOP ASILOMAR, CALIFORNIA SEPTEMBER 1980

THEORY OF INFORMATION SYSTEMS: PUBLIC AND PRIVATE DIMENSIONS -  
C. Bart McGuire

IMPLICATIONS OF GLOBAL REMOTE SENSING SYSTEMS: A DISCUSSION -  
Ludwig Eisgruber, leader

INFORMATION SYSTEMS USED IN PUBLIC AND PRIVATE DECISION MAKING -  
Bob Barker, Bob Potter, Leonard Sloski

ARE THERE COMMON RESEARCH NEEDS?: A DISCUSSION -  
Charles Vars, leader

TABLE WR-I (continued)

CRITICAL ANALYSIS OF EXISTING INFORMATION SYSTEMS -  
James Bonnen

CULTURAL, LEGAL AND POLITICAL IMPLICATIONS OF GLOBAL INFORMATION SYSTEMS -  
Oswald Ganley

CRITICAL ISSUES RESEARCH: A DISCUSSION -  
Ludwig Eisgruber, leader

The workshop focused on the topic, Information and Decision Processes.

The objective was to identify those fundamental aspects of information and decision systems that are incompletely understood, important to future applications of aerospace remote sensing technology, and which required fundamental research in order to achieve application ready status. Participants sought to identify research topics that are common to many information systems and are on the frontiers of particular disciplines, or cross-cut the usual boundaries between disciplines.

The Workshop opened with a presentation by Robert B. MacDonald concerning the goals and status of the NASA program definition study of basic research requirements. MacDonald was followed by John E. Estes who explained the purpose of the Workshop on Information and Decision Processes. Six major presentations were made by well-known authorities. Summaries of these presentations, identified by participants, are provided here. Critical Research areas identified by participants can be found in Section V of this report.

Theory of Information Systems: Private and Public Dimensions

C. Bart McGuire, Graduate School of Public Policy, University of California, Berkeley.

McGuire reviewed the nature of benefit-cost analysis, its successes and

failures, and the applicability of its methods to the evaluation of information systems. He argued that decision analysis provides an appropriate theoretical framework for addressing many questions about information systems; its practical impact, however, has been limited because the model of rational choice breaks down where there are two or more persons, each trying to achieve their own objectives. Although considerable recent work has focused on these problems, no consensus has yet emerged on the "correct" way to look at group decision-making problems.

McGuire also discussed the problems involved in evaluating public information systems. Problems for public systems are greater than for private systems because publicly provided information has special characteristics and impacts on private and public decisions. Evaluations of public systems must answer questions about how to disseminate and finance, the avoidance of politicization, errors, and the creation of false expectations, as well as the basic question of what to provide for whom. McGuire concluded that although some good work has addressed these questions, it does not provide an adequate basis for confident evaluations of existing or proposed public information systems.

#### Information Sources Used in Private and Public Decision-Making

1. G. Robinson Barker, Manager, Forest Resource Information Systems, St. Regis Paper Company.

Barker described how management decision-making by St. Regis Paper Company depends on forest inventory data acquired from remote sensing. He described the problems as well as the benefits of obtaining and maintaining accurate inventory estimates. St. Regis now uses high flight aerial photographs and Landsat data in combination with traditional

methods to acquire forest inventory data. Aerial photography and Landsat products are employed to design timber cruise strategies, annotate previously acquired inventory and cartographic materials, and develop documents for work in the field. Remotely sensed data are very useful, but high acquisition and storage expense, rapid obsolescence, timely availability, vendor reliability, and sample design remain important problems.

Barker emphasized that Landsat data are ancillary data for St. Regis, maps are the desired primary data. The digitizing and interactive graphics capabilities that St. Regis has developed to produce maps with Landsat data generate significant benefits to the company. Landsat products permit St. Regis to sample more efficiently, reduce design costs for aerial sampling, and improve its ability to monitor forest resources important to the company. Landsat products facilitate the dissemination of information and almost eliminate the need to draft maps by conventional means. Barker concluded by expressing his concern whether in the future Landsat data will be (1) continuously available in a timely fashion and (2) compatible with existing data.

2. Robert Potter, Assistant Chief of Planning, California Department of Resources.

Potter began by describing the activities of the California Department of Water Resources. The Department engages in planning activities that affect the California Water and Central Valley Projects as well as the use of water throughout California. The Department designed, constructed, and operates the California Water Project which transfers water from northern to southern California. Persuasion and litigation are important activities by which the Department seeks to influence the operation of the Central Valley Project as well as the use of water throughout California.



ORIGINAL PAGE IS  
OF POOR QUALITY

Potter explained the various types of data the Department uses in its activities. These include data on land use, water run-off and storage (surface and underground), snowpack, soil moisture, land cover, and weather conditions. The data are used with various models to estimate existing, and predict future, magnitudes of floods, water supplies, and water use. The Department also uses the data to identify seepage and plant stress, control encroachment, and conduct various conservation programs.

Potter indicated that the models utilized by the Department are generally inferior to the available data. (The exception would be the flood prediction models which are superior to the data now available.) The recent drought put great pressure on the Department to generate estimates of the costs and benefits involved in various potential incremental reallocations of water. The Department seeks to avoid the politicalization of its data; this is difficult, however, because pre- and post-project assessments by the Department sometimes lead to actions with significant economic and political consequences.

3. Leonard Slosky, Assistant to the Governor of Colorado for Science and Technology.

Slosky first discussed the needs of state and local government for natural resource and environmental data. He stressed that both state and Federal legislation and regulations have increased the authority and responsibility of state and local governments and regional agencies to plan and manage natural resources. To execute these responsibilities successfully within their limited budgets, state and local officials need spatially oriented data which are comprehensive, timely, and can be

ORIGINAL PAGE IS  
OF POOR QUALITY

acquired at reasonable cost on a repetitive basis. Conventional data collection techniques are often costly or incapable of meeting these needs.

Slosky reported that the Intergovernmental Science, Engineering, and Technology Advisory Panel (ISETAP) Natural Resource and Environment Task Force found Landsat to be an important technology that can make significant, often unique contributions to the information base required for state and local government's management of natural resources. Surveys conducted by the Task Force indicated that state and local governments find Landsat to be valuable because:

- Landsat can provide the only feasible means for conducting certain natural resource inventories. The cost, time and manpower involved in the use of conventional techniques is often prohibitive.
- The repetitive coverage of Landsat can permit users to monitor frequently the changes occurring over time in large areas.
- The multi-spectral and synoptic characteristics of Landsat can provide a new means of viewing the environment, providing information virtually impossible to acquire with conventional techniques.
- Landsat can provide uniform standardized data in a digital format which is easily incorporated into computerized information systems.

Slosky reviewed the actual use of Landsat by state, regional and local agencies, and then outlined the constraints on its increased utilization. These constraints include:

- Lack of Federal Commitment to Data Continuity and Compatibility.  
Until an operational Landsat Information System is established, states will hesitate to invest in the development of Landsat capabilities and will delay the incorporation of Landsat data into management programs and decision-making.
- Data Timeliness: Delays and uncertainty in obtaining data from the ground system have and will preclude the use of Landsat by state and local governments.
- Inadequate Federal Technology Transfer: State, regional and local agencies do not have sufficient technical capacity or specific expertise in remote sensing and digital analysis techniques. They need more technical assistance to develop an initial capability and continuing assistance to stay abreast of technological developments.
- Ill-defined Federal Agency Responsibilities. Lack of coordination among Federal agencies and ill-defined responsibilities for the Landsat system have led to limited services to state and local governments.
- Failure of Federal Agencies to Use and Encourage Landsat Use.  
Federally mandated programs impose great demands for new data. However, from time to time Federal agencies have not permitted states to use Federally provided funds for Landsat processing to meet program requirements.
- Lack of State Involvement in Landsat Decision-making. The Landsat system has been planned and operated for the Federal and international sectors with little input from state and local governments. Federal agencies have not recognized the operational

requirements of state and local agencies or the unique technical and institutional setting of each government.

- State Constraints to Use of Landsat. The fragmentation of natural resource responsibilities and the attendant difficulty of coordination, the lack of trained staff, bureaucratic inertia, and difficulties in obtaining funding for new programs have also constrained the use of Landsat.

#### Critical Analysis of the U.S. Federal Statistical System

James T. Bonnen, Department of Agricultural Economics, Michigan State University, and erstwhile Executive Director, President's Reorganization Project for the Federal Statistical System.

Bonnen reviewed the development, characteristics, and problems of the U.S. Federal Statistical System. Unlike most countries, the United States has a decentralized statistical system. For most of its history the system has worked well. Agencies have developed model statistical programs to serve policymakers and other users; they have also pioneered in the development of new technologies.

In recent years, however, the system has lost some of its effectiveness. Major new demands have been placed upon it and there have been significant changes in the environment in which it operates. Billions of dollars in the Federal budget are allocated to State and local governments under legislated formulas based on Federal statistical series. In addition, certain Federal expenditures impact private sector wages and are in turn affected by Federal statistics on prices. Government agencies require statistical data to support policy decisions and the evaluation of their activities.

Bonnen argued that national policy information needs now require a statistical system with strong overall coordination and planning. Without such direction, a decentralized system cannot deal effectively with the problems that afflict the current fragmented, uncoordinated collection of statistical activities. These problems include:

- Lack of policy relevance. While the current system provides a wide variety of data, these data are not employed adequately for objective analyses of complex issues which cut across present Department and policy decisions structure. The system has frequently failed to anticipate the need for data to be used in dealing with emerging policy problems. Failure to anticipate major data requirements leads to unnecessary conflict over the nature of problems and policy options; it also leads to rising political pressures on statistics and statistical agencies.
- Periodic threats to integrity. For the public to trust the data, it must be sure that the data are not subject to actual or perceived manipulation and that the objectivity of Federal statistical data is fully protected. While the record is good, there have been occasional problems.
- Inadequate quality. The statistics produced in small units, or as by-product, are often unreliable and poorly designed for their purpose. The growing volume of contracted data collection and analysis are too frequently of poor statistical quality.

- Inadequate protection of privacy of respondents.
- Excessive paperwork.

Bonnan concluded by explaining why the Statistical Reorganization Project recommended, and the President approved, establishment of an Office of Statistical Policy (OSP) in the Executive Office of the President for the coordination of the Federal Statistical System. This Office would function as a separate agency reporting to the President and accountable to the Congress.

The mission of the Office would be to:

- Ensure the policy relevance, quality, and integrity of statistical data and analyses produced by agencies of the Federal Statistical System.
- Develop plans so that the Federal Statistical System will be able to meet future information needs efficiently in the face of technological, social and economic changes;
- Assist in minimizing the burden on all persons and organizations asked to supply statistical or other data to the Federal Government;
- Maintain a proper balance between protecting individual and business rights to privacy and confidentiality and meeting information needs for public policy.

#### Political, Cultural, and Legal Problems of Global Information Systems

Oswald Ganley, Executive Director, International and Allied Areas, Harvard University Program on Information Resources Policy, and erstwhile Deputy Assistant Secretary of State for Science and Technology.

Ganley discussed the political, cultural, and legal aspects of global remote sensing systems and their use. He addressed three questions:

1. What will the political world look like during the next twenty years?

ORIGINAL PAGE IS  
OF POOR QUALITY

2. What problems are associated with large, repetitive, and long-term flows of data across national borders?
3. What legal and political problems affect the international use of remotely sensed data?

Ganley predicted that the world during the next two decades will experience recurring political struggles. Uncertainty will characterize most relations between nations, regions, and groups, and leadership will be ephemeral and limited. Neither the United States nor any other nation will dominate or consistently provide leadership in international affairs. The United States, despite its great power and leverage, must negotiate with other nations and international organizations to achieve its objectives.

Ganley took the position that the United States government has just begun to understand that the world has entered an "Information Age." Information is a basic resource like materials and energy. The need to ask "Who has a resource? Who wants it? How can you get it? and What are the terms of trade?" links any resource to political or economic power. For information resources, Ganley argued that traditional answers are now inadequate owing to the growing abundance and versatility of modern technology.

To illustrate the changing and uncertain answers to questions concerning information resources, Ganley drew on his previous analyses of the role of information resources in Canada and of their influence on U.S.-Canadian relations. The possibility of rapid, massive, and cheap flow of data across borders and the preeminence of U.S. facilities and services has



led many countries to worry about transborder data flows. Western European governments justify these concerns in terms of individual privacy. But the Canadian government and industry view them as essentially economic - a matter of lost jobs, negative effects on the balance of payments, and foregone management opportunities. (See Figure WR-1).

Research and development in high-technology industries has also been at issue. The Canadian government has expressed displeasure over the low level of research and development performed by U.S. subsidiaries in Canada and the substantial U.S. ownership of the Canadian electronics industry. These sentiments are analogous to those expressed by U.S. industries about Japan.

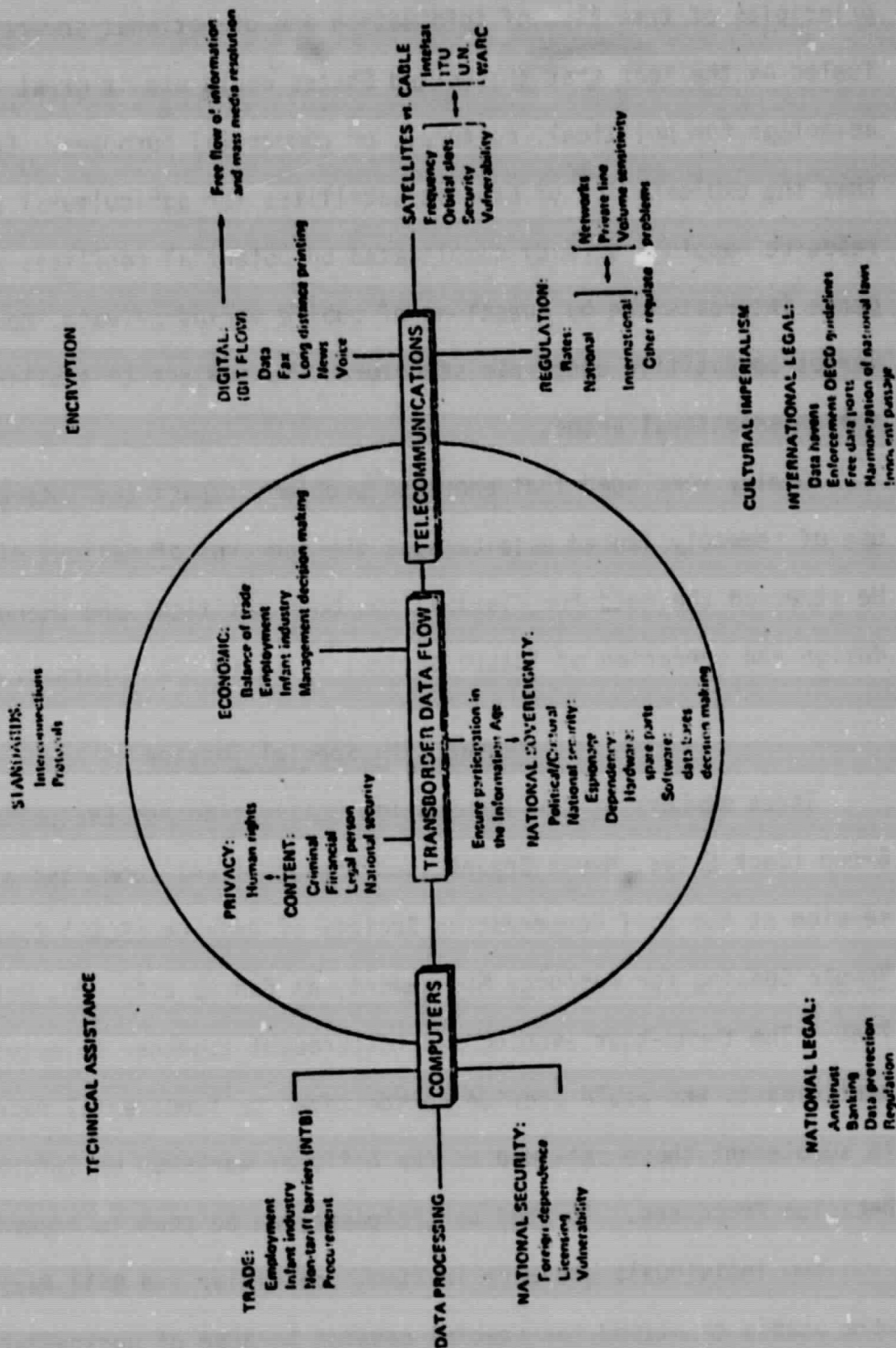
The Canadian government perceives Canada as being overwhelmed by American media content that robs Canada of its national identity. Directly, this mainly affects such U.S. industries as publishing and broadcasting. But similar perceptions in the Third World pit restrictionist attitudes against American devotion to the principle that information should flow freely. The resulting clashes have ideological as well as economic overtones.

The possibility of broadcasting directly to homes via satellites raises similar issues. Canada wants to have domestic television broadcasts from satellites rather than local transmitters. But this could mean importation of even more U.S.-made content than at present. Globally, many nations believe that direct television broadcasting by satellite from one country to another without the prior consent of the receiving state is a

ORIGINAL PAGE IS  
OF POOR QUALITY

FIGURE WR-1

*Transborder data flow  
(TDF) affects all interna-  
tional business, not just the  
data processing and tele-  
communications industries.*



violation of national sovereignty. The resulting conflict between the principles of free flow of information and of national sovereignty is fueled by the fear that the United States would use its great technological advantage for political, cultural, or commercial purposes. Ganley argued that the exploitation of Landsat satellites for agricultural and mineral resource mapping, already complicated by potential conflicts with intelligence interests and by questions of public versus private, monopolistic versus competitive ownership structures, is subject to similar fears in the international arena.

Ganley concluded that enormous problems impact the international use of remotely sensed data because the concerns of nations differ greatly. He stressed the need for flexibility, both political and economic, in the design and operation of future systems to serve the international community.

#### Kansas City Special Session

Three members of the Information Utilization and Evaluation Working Group (Jack Estes, Nevin Bryant, and Charles Vars) conducted a special session at the Soil Conservation Society of America (SCSA) Conference, Remote Sensing for Resource Management, at Kansas City, MO, October 28, 1980. The three-hour evening session brought together selected conference participants who would provide perspectives on fundamental research needs to supplement those obtained at the Asilomar workshop on Information and Decision Processes. Invited participants can be seen in Appendix A.

Two individuals who were invited, Joe Cihlar and Bill MacFarland, were unable to attend the special session because of unexpected changes in their schedules, but Estes was able to meet with them in subsequent

interviews and their views are included herein. In addition to the synopsis of our discussions we have included in a copy of the results of a data needs question circulated at this meeting (Kansas City USER NEEDS SURVEY, Appendix C). These results were provided by Cihlar to Estes in December, 1980.

The session opened with a presentation by Estes concerning the goals, status, and activities of the NASA Program Definition Study of Basic Research Requirements. Bryant and Vars then briefly reviewed the Workshop on Information and Decision Processes at Asilomar, as well as planned future activities of the Working Group on Information Utilization and Evaluation.

One major research recommendation emerged from the two-hour discussion session. This recommendation was that existing information systems, their use and users, and related processes need to be reviewed and documented before potential future systems can be evaluated. The objective should be the review and documentation of multi-purpose, government and private systems with international dimensions. The review and documentation process should not just include systems which obviously lend themselves to remote sensing applications. The option of combining remotely sensed data with data collected by other means should be studied. In fact, this latter possibility indicates why the review process should place special emphasis on why, how, and with what effects are multiple-source of data used within existing information systems.

Most studies on the value of information assume, for example, that a farmer, an elevator operator, and a buyer all use the same information in the same way, but evidence indicates this is unrealistic. Some information

is available as a public good, but it is not equally used by all who receive it. For these reasons, the identification of major market and nonmarket users of information, their sources of information, and the adequacy of these sources become important if new systems are to be satisfactorily evaluated. Particular attention should be given to the documentation of how the behavior of public and private users and producers of information are changed when information is viewed as a public rather than private good.

### San Jose Workshop

The second workshop conducted by the Working Group on Information Utilization and Evaluation was held at Santa Teresa Laboratories of IBM, near San Jose, California, January 19-20, 1981. Participants are listed in Appendix A. The workshop focused on the topic, Data Base Use and Management. A list of individuals making presentations and the title of their talks are seen in Table WR-II. The objective was to review the state of the art in computerized data base management systems and technologies associated with the use of data-particularly remote sensing data. Eight presentations were made by individuals well versed in the field.

TABLE WR-II

Titles of Presentations and Speakers at  
DATA BASE USE AND MANAGEMENT WORKSHOP  
SAN JOSE, CALIFORNIA  
JANUARY 1981

SPACE APPLICATIONS DATA SYSTEMS PROGRAM -  
Nevin Bryant for Peter Bracken

PLANETARY SCIENCE DATA ACQUISITION -  
Adrian Hooke

TABLE WR-II (continued)

REQUIREMENTS AND CHARACTERISTICS OF SCIENTIFIC DATA BASES -  
Stan Hansen

CARTOGRAPHIC DATA BASES -  
Delmar Anderson

COMMITTEE ON DATA MANAGEMENT AND COMPUTATION REPORT REVIEW -  
Ralph Bernstein

DATA BASE ISSUES IN GEOGRAPHIC APPLICATIONS -  
Ray Lorrie

PROBLEMS OF HIGH RATE COMPUTATION -  
William Sharpley

DATA BASE CONSIDERATIONS FOR NAVY WEATHER FORECASTING -  
Eve Schwartz

Summaries of these presentations are provided below. Critical Research issues from the discussion held at this workshop are found in Section V of this report. An appendix of selected viewgraphs given by each speaker is provided as background information (see Appendix B).

1) P. Bracken/N. Bryant (presenting)

NASA/OSTA Data Systems R&D Program

The Data System has as its goal the development of key enabling technologies to assist the OSTA Discipline Branches more efficiently address the ground data handling of earth observation systems sensor data. The program cuts across all of the Discipline Branches, with the result that it has been structured to address both immediate and longer-range concerns to OSTA applications. Immediate concerns in improved data handling are being addressed through the development of Pilot Demonstrations and focusing on critical mission requirements such as the Landsat D ground data handling system. Longer range concerns are being addressed through the

development of NASA expertise in information science and contributions to the data handling design for missions scheduled for FY86+ launch. Liason with OSTA research is maintained, but emphasis again is on the Pilot Systems. The Branch's philosophy is to learn by doing, and build specific blocks of expertise. The Pilot Systems may in fact be too research scientist oriented for a complete evaluation of system requirements, given the stated OSTA and user community being composed of federal, state, and private organizations with high-volume and repetitive data processing characteristics. However, there are some very challenging issues in extracting indexes and storing data associated with satellite information. These probably are not amenable to significant theoretical analysis given assumptions by high-volume and repetitive data processing characteristics but are best handled on an experimental basis.

## 2) A. Hooke

### Planetary Science Data Acquisition

The planetary science data acquisition differs from the earth observations program in two important respects: much lower data rates and a well defined and limited set of users. The process, however, suffers similar problems to that experienced by the earth observations missions. Being small, however, has enabled the program to recently redesign its information system from end-to-end, thereby highlighting elements of generic value to the Fundamental Research Program. Key elements include: (1) exploring the value of data compression to trade off information for volume, (2) developing standardized data links between components for staging/transmission, (3) developing standards to permit expansion and externalization of the data network to more end users/



scientists. A principal enabling element has been the development of layers of protocols to keep track of packets of data being transmitted from spacecraft to ground and data base to data base. Layered protocols probably are the best way to completely inform the recipient of a data sets characteristics, but it is impractical to expect retroactive upgrading of protocol information in the initial data bases.

### 3) S. Hansen

#### Requirements and Characteristics of Scientific Data Bases

The Boeing computerized graphics and data base systems applied to airplane design, manufacture scheduling and logistical support was discussed in the context of their technological requirements. Basic problems in the Boeing environment are analogous to those in satellite information processing and utilization. Some key points brought up were: (1) we are not just computerizing data and procedures that were already there, we are creating new approaches to analysing data and making comparisons through a greatly reduced cycle time for information processing. (2) Computerizing systems are vertically integrated to perform specialized tasks with function-oriented software, while in fact what is needed is a global-oriented transfer capability so that transferability of data can occur through horizontal integration. (3) Where many components characterize a system the number of horizontal interactions (e.g., facility-to-facility) should be minimal and they should not be complex. (4) The costs associated with moving through a heterogeneous environment are associated with reformatting and converting data; moreover, image and graphical data types and hardware do not have

ORIGINAL PAGE IS  
OF POOR QUALITY

complete compatibility of interfaces. (5) Compatibility can be helped through the development of a neutral format all components must conform to, although this does place an overhead on operations. (6) There needs to be decentralized data base structures, distributed data processing, and a common carrier if the variety of users/technology managers are to have a sufficiently flexible research and design environment to meet the variety of customer/end-user needs/requests. (7) Users do not have requirements, they have reactions to situations imposed upon them.

4) D. Anderson

Cartographic Data Bases for Exploiting Remote Sensing Imagery

It is important to note the difference between the cartographic data base manipulation and use for analysis versus that required for presentation. There is a need to provide simplified cartographic products to the end user which get the salient points across. However, this same data must be used for analysis as well as presentation. Therefore, you do not want to have to change the data encoding structure over time, but rather have levels of generalization to handle more spatially detailed information. Furthermore, the cartographic data bases must be organized to accommodate a wide variety of spatial attributes (points, lines, areas) and integrate them in a spatially consistent manner for analysis and simplifying them uniformly for presentation. A final problem was noted, that of anticipating questions you need to answer. This affects the cost-effectiveness of data capture and storage/archiving as well as data formatting.

ORIGINAL PAGE IS  
OF POOR QUALITY

5) P. Bernstein

Committee on Data Management and Computation (CODMAC) Report Review

The viewpoints expressed by CODMAC on the whole parallel those generated by a variety of NASA-sponsored user-requirements studies and workshops. Concerns with data quality, timeliness, and accessibility were prevalent. This report has significant importance, in that it represents the views outside NASA of NASA's data distribution role. Two points concerning the report were noted: (2) the report reflected an assessment of the space data handling situation from the perspective of research scientists, b) scientists are sometimes resistant to technological change compared with their existing research medium largely because they feel unfamiliar with the new way of doing things and/or processing of data is beyond their control or preferred specifications.

6) R. Lorrie

Data Base Issues in Geographic Applications

The use of the relational data base approach to handling geographically encoded data files and graphics data was discussed. The System R approach was described. A significant point was that the system operated on records not sets or arrays of data. Up to date this has created severe overhead problems for graphical data representations, and image arrays have not been addressed. System R, as most DBMS, have concentrated on the accounting/inventory data management problem. The relational data base concept would be useful for many applications if it could efficiently handle spatial oriented data other than nominally encoded files. The value of the relational data base technology for

ORIGINAL PAGE IS  
OF POOR QUALITY

spatial data will depend strongly upon the ability to extract quantitative features of that data. These are necessary as a base for the inquiry and manipulation facilities which are key to relational systems

7) W. Sharpley

Problems of High Rate Computations

The technology of high data rate computing, even for the specialized case of image arrays, is developing independent of NASA to serve other user needs. There is probably a compatability with the development of high rate computing where NASA may not need to support hardware development for sensor data interpretation, but rather the formatting of data bases. A basic bottleneck in image file computing is I/O. Potential solutions include back end machines and floppy disks. It is important to note that geographical/image data bases use more streamlined DBMS functions than generally provided by commercial systems. This is because many of their concerns are limited in scope and do not involve requirements such as frequent updating of all records in a file.

8) E. Schwartz

Data Base Coordinate System for the Storage and Retrieval of Satellite Data

The Navy's procedure to place satellite cloud cover information into an easily addressable working data base for meterological models to operate on was described. The system is a good example of the kinds of software and hardware architectures which can be developed to achieve a very specific product derived from satellite imagery. The data is not archived (i.e., each new overflight replaces the previous one) and the map projection process is not reversible to the intial scanner data.



Flexibility in satellite data analysis capabilities has been sacrificed for a comprehensive input capability to complex weather forecast models which require systematically formatted/mapped data to operate efficiently.

#### Houston Workshop

The third workshop conducted by the Working Group on Information Utilization and Evaluation was held at the Lunar and Planetary Science Institute near Johnson Spacecraft Center at Clear Lake City, Texas, February 23 and 24, 1981. The objective of this workshop was an examination of some of the important attributes of data which give it value as information and methods for evaluating the ability of remote sensor systems to provide accurate data/information for resource management purposes. Attendees at this workshop are listed in Appendix A. Seven presentations were made at this workshop. TABLE WR-III is a copy of paper presentation schedule followed during this meeting. Summaries of these presentations along with important identified research issues in Section V of this report.

#### TABLE WR-III

##### Titles of Presentations and Speakers at DATA AND INFORMATION PERFORMANCE WORKSHOP HOUSTON, TEXAS FEBRUARY 1981

PERFORMANCE PARAMETERS AND CRITERIA FOR JUDGING FEDERAL INFORMATION SYSTEMS -  
Gaylord Worden

DATA AND INFORMATION FROM THE SRS SYSTEM -  
Raymond Jessen

ASSESSMENT OF THE STATISTICAL ACCURACY OF ESTIMATES MADE FROM REMOTELY  
SENSED DATA -  
Patrick Odell

MAINTAINING THE SPATIAL COMPARENT IN STATISTICAL AGGREGATION -  
Alan Strahler for Reginald Golledge and Larry Hubert

TABLE WR-III (continued)

FEDERAL-PROVINCIAL AND INTERNATIONAL DIMENSIONS OF CANADIAN REMOTE  
SENSING PROBLEMS -  
Donald Clough

MODELLING THE EFFECTS OF MISREGISTRATION ON MULTI-SPECTRAL CLASSIFICATION-  
Fred Billingsley

THE DATA REPRESENTATION ACCURACY QUESTION IN REMOTE SENSING-  
Bill Cobberly

Summary of Worden's Presentation, JSC,  
Performance Parameters and Criteria for Judging Federal  
Information Systems  
February 23, 1981

Worden began by describing the Federal Statistical System, its problems, and the missions of the Office of Federal Statistical Standards and Policy (OFSS&P). The United States has a decentralized Federal statistical system: many agencies produce statistics, but each is responsible for the design, collection, quality, and dissemination of its statistical outputs. Some agencies contract with private firms to collect statistics, whereas others have established large and professionally sophisticated organizations to produce statistics.

The quality and policy relevance of Federal statistics vary widely. Statistics produced by small units, or as a by-product of administrative and regulatory activities, are sometimes unreliable, poorly designed, and inadequate for important policy problems. Failure to anticipate major data needs has led to political pressures on statistical agencies. The objectivity of Federal statistical data has been subject to periodic threats, but the record to date is good. Protection of the privacy of respondents who provide records, as well as minimization of the burden of response are problems of increasing significance today.

Worden indicated that the OFSS&P is the only central Federal agency concerned with Federal statistics. The specific missions of the office are to:

- Ensure the policy relevance, quality, and integrity of statistical data produced by Federal agencies;
- Develop plans so that the Federal Statistical System will be able to meet future information needs efficiently;
- Assist in minimizing the burden on all persons and organizations asked to supply data to Federal agencies;
- Maintain a proper balance between protecting individual and business rights to privacy and confidentiality and meeting the information needs for public policy.

The OFSS&P seeks to ensure policy relevance by addressing broad issues that cut across Department and policy decision structures. The office represents these issues throughout the decentralized Federal statistical system, and it encourages communications among agencies, reporting populations, and users so that alternative needs are recognized in agency planning. The objective is to promote the development of an integrated statistical system with sufficient uniformity in concepts and measurements that the system can serve current and emerging national policy needs satisfactorily.

The OFSS&P routinely reviews the contracts let by small units for data collection and other statistics-related activities. Subjective benefit-cost analysis, as well as studies of objectivity, access, and use, are conducted by the office to maintain the credibility of Federal statistics and the confidentiality of data on which they are based. For Federal statistics to be trusted, the public must be sure the (1) data are not manipulated to serve narrow political interests and (2) the objectivity of the data is fully protected. Effective protection must be given to the confidentiality of data supplied by persons and businesses. Regulatory agencies rarely pledge confidentiality because they are concerned with the activities of specific entities. Some actions of regulatory agencies concern the major



agencies producing economic statistics because respondent cooperation declines with increasing fears of publicity.

Worden emphasized that the OFSS&P seeks to address trade-offs in data collection, handling, and respondent burden for the entire Federal statistical system. The office has developed an Information Collection Budget which will be used to evaluate statistical reporting burdens on the public. Remote sensing systems may provide observational means that minimize respondent reporting burdens, but the issues of privacy, verification, accuracy, and cost remain important problems. Worden observed that the criteria used by the OFSS&P are appropriate to statistical systems for natural resources, science, and technology, but neither concepts nor institutions are sufficiently developed to permit trade-off and benefit-cost analyses comparable to those undertaken for economic and other Federal statistics.

Data and Information from the SRS System  
by Raymond Jessen, University of California, Los Angeles

In Professor Jessen's talk he dealt mainly with procedures for the estimation of cereal crops. He began with a historical perspective. In the early days SRS type estimates depended upon reports from farmers. Errors could crop up in this type of reporting so adjustments were made and a structure developed called the Crop Reporting Board. This board examines information produced by state officers to make its estimates of production. As this board looks at prices of crops that are often highly speculative the integrity of their data is of primary importance and has created some problems in the past. This area of data integrity is still being worked on today.

In the 1930's some experimentation was done on the use of airphoto interpretation to look into cereal grain areage for:

- o mapping; and
- o crop type identification.

Each field area was visited on the ground. This data collection procedure was not adopted but was still useful in looking at variability of field sizes - this was good for research in sampling.

Next, came research on yield. This was done in small plots in a number of states. These plot data were then aggregated up to field size statistics. Cotton ball counts were the first really valid statistical estimates of yield. Corn was next looked at for yield. Corn was the first crop where a big difference between the plot data and the field harvest data began to show up. (Biological yield - different than - harvested field yield). These research harvests developed in the 1930's but the procedures developed were not adopted until the 1950's.

We should remember that it is easy to sample in Indiana and Illinois but in other areas without regular field boundaries this is difficult so a natural boundary scheme was worked out. But SRS is not just interested in yield or just in acreages. All this led to the current SRS systems.

In this system farmers are visited and shown airphotos of their fields with boundaries overlayed. Field sizes are used as a surrogate for planted areas. These photos are then changed and annotated in the field by the farmer and the reporter. Crop and acres are the type of information gathered but other data are taken as well (e.g., number of livestock).

This information is taken by SRS during their survey in about 400 samples sites per state. Some states have more sites - California

has some 800+, New England may have states lumped together.

Now in June not all crops are planted (e.g., corn, soybeans). These data are gathered by asking what will be planted. These data then form the sample units for yield data. These plots are then looked at monthly and detailed data are forwarded to the Crop Reporting Board; along with other information. This other information survey data sent to farmers who are asked how is your crop compared to the norm. Later farmers are asked how much yield does it look like. In this type of reporting you have early predictions which are more forecasts (crop not made yet); later these can be referred to as estimates. Again these data are reported to the state office, then to the Crop Reporting Board. These people use ancillary data too! Such as car loadings, (that is, number of cars loaded with a specific crop and the total tonage of that load)., etc. Then they make essentially very judgmental decisions.

Next, we asked Dr. Jessen why this decision system is not more systematized? Jessen replied that he is not sure of how subjective judgments like this can be made more objective. After all, if we can do more valid sampling, we could get a more unbiased estimate. If so, we should be able to get right on. What we have though is a series of:

- o sampling problems;
- o reporting error problems; and, .
- o non-response type problems (not too much of a problem because you can drive around fields and get data).

SRS and USDA are getting some accuracy checks on specific crops. The orange crop is one that has more objective limb counts (e.g., oranges-squeezed or shipped, cotton-gined, sugar beets-milled). These can be relatively accurately verified. In crops where on farm use is made of the crop such as corn for cattle feed estimate data is less reliable.

Jessen was then asked - Why do we keep this system going? Why not change over to a new system? Jessen stated that in his opinion it was because they don't know the magnitude and sources of error and the problems associated with other schemes. He also stated that we must remember that SRS/USDA want to keep farm visits as they are collecting more than just the acreage and yield data.

Another problem area pointed out by Jessen involved the determination of objective yields (biological yields are higher than reported yields). This problem is well known within the agency (USDA).

Although remote sensing is currently being used in Landuse mapping to reduce sampling error, more work and documentation is needed in this area. In the future they may use Remote Sensing to put out more local statistics using Landsat as a base.

In conclusion, Jessen restated the problem that biological yield may be higher than reported yield and how do we get a handle on this. The agency is aware of it, he says, but it is not a straightforward problem to handle. He also stated that he was aware that remote sensing is being used in Landuse mapping to reduce sampling errors and that in the future USDA/SRS may use remote sensing to put out more local statistics. For this Landsat might serve as an acceptable base.

Assessment of Statistical Accuracy of Estimates Made from Remotely Sensed Data  
Patrick L. O'Dell, University of Texas at Dallas

By its nature, resource inventory from satellite remote sensing involves sampling, applications of decision theory, and most importantly the processing of enormous amounts of data. This last element in resource inventory schemes such as LACIE or AgRISTARS means that classical sampling techniques are nearly applicable. The sensor data have a large number of potential errors



associated with registration, atmospheric, etc., which require statistical methods that are robust or entirely new. When considering accuracy assessment, the LACIE or AgRISTARS procedure involves three kinds of estimating conditions of crop yield: a) where a true value is made available to compare modeled estimates, b) where estimates of true value generated from other sources are available, c) where no grounds for comparison of modeled estimate exist. Condition (a) is always desirable, condition (b) can in part be overcome by comparing the variances of estimates generated by both methods, but for (c) the only practical alternative is to develop a modeled hypothesis and verify it on a surrogate. The verification for (c) can come from the use of analogue areas or use of a test set of verified data.

Much of the AgRISTARS work will involve category (c) conditions, where no check is available for verification. Therefore, fundamental research needs to be undertaken in looking at: understanding some radiance to sensor model parameter effects, improved statistical theory to incorporate spatial attributes e.g., BLOB and AMOEBA, and improved sampling strategies which are both more robust and sensitive to differentiating crop types. These could not practically evolve through using ~~stan~~ standard data sets to test models and applying systematic constraints (covariances) to the test one. A final need is to document the research procedures as well as the methodology adopted in each application to focus on commonality for concern of future adoption as procedures.

What follows is my list of research topics which I feel should be pursued. If resolved these should lend significantly to our understanding and hopefully increase the accuracies of our estimates.

- (a) Develop a better technique for estimating the number of distinct

crops in a region. We have used cluster analysis methods in the past but other techniques have not been applied to this problem (mode estimation using statistical theory of extremes).

(b) Investigate algorithm for finding boundaries which may be applicable for adjusting estimates for mixed pixels, especially for small field application.

(c) Perform a major study to devise various sampling plans and methods for different application using remotely sensed data. Sampling plans where probabilities of multiclassification are incorporated.

I am interested in systematic random sampling plans in which a systematic methodology is incorporated to stratify with respect to weather (agrophysical regions or agrometeorological regions as it were).

Compute the bias term for all estimators. Bias remains the dominant problem of all production estimates. Develop what I will call attribute vector observation vectors (composed of 0's and 1's) which will allow better use of temporal information. That is, if an observation over time has attributes that qualify it to be that of a crop class of interest the vector would contain component of only ones (zero's imply that the pixel lacks an attribute of the particular crop class. Classification should follow using these vectors.

Finally, I have had interest in evaluating information content of remotely sensed data to determine whether or not the data sets produced by remote sensing can do a specified job. My concern has been whether or not the data contains

- (a) sufficient information to estimate acreages accurately, and
- (b) sufficient information to separate similar crop classes such as wheat and barley

The research, so far, has been primarily concerned with showing that

verified data whose accuracy parameters are supposedly known contain an order of magnitude more information for estimating proportions in a mixture probability density than does classified data (data processed with some mixture of man and machine) or unlabeled "raw" data. This research tends to confirm the old adage "In bridge, one peak is worth a hundred finesses!"

Now I will get on my "soap box"

1. Document, Document, Document
2. I still would like to see a formal Book of Test Data Sets
3. I would like to see a formal book of models
4. I would like to see a formal book of sampling plans.

In the early days of NASA, I recall several books summerizing significant results. Books on remote sensing has started to appear but more are needed to complete the record.

Summary of remarks\* by Professor Donald Clough, Feb. 23 & 24, 1981 at Houston meeting of Working Group on Utilization and Evaluation.

Certain elements in a data base require careful scrutiny in the evaluation process. A possible list would include:

- (a) integration
- (b) uniformity and standardization
- (c) time requirements of the users (e.g. real or archival)
- (d) Elusiveness of specific user needs, -- a distinction being made here between requirements, as put forward in answer to questionnaires, and the purposes to which information is put.
- (e) credibility -- in the way of acceptable justification for the information system. In Canada, sovereignty imperative carried great weight. Cost/benefit analysis served only as the crutch leaned on for policy purposes. Language is a very important factor, and quantitative arguments are not nearly as potent as the political ones.



- (f) Public good. Issues relating to involvement with private industry, matters relating to proprietary information and confidentiality have to be faced.
- (g) Policies and standards. While the "data bank" concept has appeal, one-step shopping for all kinds of information for all kinds of purposes seems unrealistic especially in view of the lack of coordination among agencies as to definitions, classification, formats, map scale, and so on. Different agencies not only have their own specific areas of interest, but their own favored methods and standards.

(On this point, Gaylord Worden offered the cogent observation that purpose considerations cannot be emphasized enough. What, he asks, are the agencies trying to achieve? If there is no acceptable answer to this basic question, there can be no bona fide analysis, cost/benefit or any other.)

---

\* Professor Clough's contribution included many observations in addition to his "presentation." Because these were extremely pertinent, I have tried to capture them along with the gist of his statement. IRH

In the case of technology push, which characterizes remote sensing applications, most of the experiments in agriculture are not well controlled; internal logic is not well constructed; and it is virtually impossible for "captive" scientific and technical people to be objective and uncover the linkages.

The history of satellite technology is important and should not be overlooked as a source of clues to understanding how we got where we are and where we are heading. Canada developed magnetometers. Larry Morley established the center as receiving station for Landsat I. Then came cabinet authorization, criteria for evaluation at the political as compared with bureaucratic levels far different one from the other. The main desideratum for Canada was "sovereignty protection," and the Cabinet agreed and approved. Cost/benefit analysis came later, but it was only the window-dressing following the political process. In some known cases, the claims embodied in such analyses were so unrealistic as to prove embarrassing.

At present, Canada is interested in the Seasat concept. (Incidentally, satellite technology has been good for Canada's export market, the building of receiving stations all over the world being the stock-in-trade of a Canadian Company.) The arguments in favor of SAR (Synthetic Aperture Radar) systems are in the political arena and not based on economics, since proponents are well aware that 15 to 20 years will be needed for proper development of the system. Not only is it premature to try to measure benefits but at this time it is known that there is not market. "Users" are, in fact, scientists associated with the government. There being significant lack of a commercial market, Canada is skeptical as to whether its space program will ever "pay" in the conventional sense. This is not, however, to say that they value it less. On the contrary, the Treasury Board looks not only to its spinoff (such as the construction of ground stations worldwide) but also to certain kinds of needed technical capability. Canada needs SAR because of its problems with cloud cover; it needs a sensor able to provide illumination through clouds and night and day, ice surveillance in its northern regions being a matter of paramount concern. It would appear that at this stage, clear focus on a recognized area of concern could be a more promising way of demonstrating remote-sensing capability than diffusion through efforts to reach many users. Clough's advice: find a winner.

It was Clough's considered opinion that neither NASA nor NOAA was equipped yet to deliver an information system and he asked how the message about the technology's present inadequacies and shortcomings (cf. Billingsley's presentation) could get to the users. This is not a trivial question when one recalls the amount of "selling" and "showing-and-telling" that has accompanied Landsat and that we see happening in the case of the shuttle. Clough says that optimization cannot occur before we understand and address the structure of the market and for this a structural, theoretical framework is needed, but unlikely because

ORIGINAL PAGE IS  
OF POOR QUALITY

ORIGINAL PAGE IS  
OF POOR QUALITY

of the bureaucratic, managerial approach which has been used. There can be no proper evaluation until we find some one who is willing to pay a price. Otherwise, we are all operating on faith.

MODELING MISREGISTRATION AND RELATED EFFECTS  
ON MULTISPECTRAL CLASSIFICATION\*

Fred C. Billingsley

Jet Propulsion Laboratory, California Institute of Technology  
Pasadena, California 91109

Spectral analysis generally takes the form of multispectral classification in which the classification is done by comparing the sample measurement vector to the statistics of the set of known material vectors (training statistics) representing all possible classes, and by using one of several decision methods, determining which of the knowns it most nearly matches.

The problem pursued will be the effects of misregistration on the accuracy of multispectral classification in answer to the question:

What are the effects on multispectral classification accuracy of relaxing the overall scene registration accuracy from 0.3 to 0.5 pixel?

The misregistration is but one of a group of parameters (noise, class separability, spatial transient response, field size) which must all be considered simultaneously. The thread of the argument (which will be discussed in detail below) is this: any noise in the measurements (due to the scene, sensor, or the analog to digital process) causes a finite fraction of measurements to fall outside of the classification limits. For field boundaries, where the misregistration effects are felt, the misregistration causes the border in a given (set of) band(s) to be closer than expected to a given pixel, so that the mixed materials in the pixels causes additional pixels to fall outside of the class limits. Considerations of the transient distance involved in the difference in brightness between adjacent fields, when scaled to "per pixel", allows the estimation of the width of the border zones. The entire problem is then scaled to field sizes to allow estimation of the global effects.

This approach allows the estimation of the accuracy of multispectral classification which might be expected for field interiors, the useful number of quantization bits, and one set of criteria for an unbiased classifier.

ORIGINAL PAGE IS  
OF POOR QUALITY

The following conclusions were elaborated in detail in the presentation:

- The difference between 0.3 and 0.5 pixel misregistration is in the noise for multispectral classification.
- Precision users may have to re-register image segments anyway, making extreme registration precision by the system of less importance.
- Interpolation algorithm choice is relatively unimportant, provided a higher-order interpolator is used.
- If small field are important, small pixels are more important than sensor noise contributions.

In addition, several observations result:

- System registration to 1-2 pixels should satisfy users of film products.
- There is a grey area of 0.5 to 1-2 pixels in which the requirements for high precision are not well justified.

#### THE BASIC MODEL

The expected effect of misclassification may be estimated by a simple first-order approach, because the differences in classification accuracy between the many classification schemes and conditions that have been tested are overshadowed by the vagaries in the data and assumptions in the classification process, so that higher order analysis will contribute little additional understanding.

Consider first the probability of correct identification of a field interior pixel. Field interiors are nonuniform because of the combined effects of sensor noise, scaled to equivalent reflectivity ( $NE\Delta\rho$ ) and inherent nonuniformities in the field itself. The overall brightness distribution is considered to be Gaussian - this is approximately true for field interiors, although the distribution deviates considerably toward bimodal for mixed materials at field borders.

The combined effect of these various noise sources produces a finite probability of misclassification. (Figure S-1) The first-order estimate considers the total variance caused by the scene, sensor and quantization as compared to the defined class size limits, however these are determined. Similar, but relatively second-order, effect may be expected with a higher order analysis. Proper classifier training, resulting in accurate limits, is essential (Hixson et al, 1980).

For simplicity, and because of the later desire to misregister one (or more) of the bands, the discussion will assume that spectral bands as sensed will be used, and that for recognition, the unknown pixel must fall between



appropriate limits in every band tested. Therefore, brightness outside of a limit in any one band is sufficient for rejection, so that we need to consider only one band at a time.

The probability of a sample being within the class limits can be derived by assuming that an ensemble of clean signals from a series of areas of the same material can be anywhere within the quantizing range with uniform probability, but that individual samples are perturbed by the Gaussian noise with a distribution equal to  $\sigma$ . The probability distribution of the signal plus noise is found by convolving the probability distribution of the signal with that of the noise. The probability of correct class assignment (i.e., the pixel is within the class limits) is then found by integrating the probability distribution between appropriate class limits (Friedman 1965). The result of this calculation is shown in Figure S-2. In the useful range of  $\beta$  ( $3 < \beta < 7$ ), the curve can be approximated by

$$\beta \log P = -0.40$$

where  $P$  = probability of correct classification, and

$$\beta = \frac{\text{class size}}{\sigma_{\text{scene}}}, \text{ with class size and } \sigma_{\text{scene}} \text{ in the same units.}$$

Sources of noise will be the scene itself and the sensor, both assumed to be random for this analysis. The root mean square (rms) sum is taken to give the total effective noise. A number of pixel measurements may be averaged together to reduce the noise before classification. This final noise figure may be compared to the width of the class to give  $\beta$ , from which the probability  $P$  of correct classification may be estimated. This leads to the Classification Error Estimator, Fig. S-3.

As an example, consider a scene having a field-interior variation of 3%, to be viewed with a sensor having a total noise figure of 1%. The total effective noise seen by the classifier (upper left) will be the rms sum of these, or 3.16%, which for a total 0-255 digital number (dn) range, would be 8.1 dn. If the class width (determined by the classifier algorithm) is 25 dn (right center) the  $\beta = 3.1$ , giving  $P = 0.742$  (right lower). If this  $P$  is not accurate enough for the analysis, several pixels must be averaged (right upper): a 2x2 averaging will raise  $\beta$  to 6.2, giving a new  $P = 0.86$ .

Considering  $\beta$  in this way allows an estimation of the total noise permissible as it affects the attainable classification accuracy. If the amount of scene noise to be encountered in a given classification task can be estimated, the allowable extra noise from the sensor and quantization can be specified by estimating the loss of accuracy of the classification caused by quantization error. This leads to an estimate of the number of bits which will be useful.

Define the perfect sensor as having no random noise nor quantization error (i.e., an infinite number of bits). This will define (for  $n \times n$  pixels averaged)

$$\beta_0 = \frac{\text{class size} \cdot n}{\sigma_{\text{scene}}} \quad \text{and} \quad P_0 = 10^{-0.4/\beta_0}$$

ORIGINAL PAGE IS  
OF POOR QUALITY.

For the real sensor,  $\beta/\beta_0$  because of the finite  $\sigma_{\text{sensor}}$  and  $\sigma_{\text{quantization}}$ .

The new probability of correct classification  $P$  is related to  $P_0$  by:

$$P = P_0 (\beta/\beta_0)$$

A plot of the loss in classification accuracy vs.  $P_0$  is given in Figure S-4, for the parameter families  $\beta_0/\beta$  and  $\sigma_{\text{sensor}}/\sigma_{\text{scene}}$ . Noise allocation starts with defining the desired  $P_0$  and ascertaining that the required  $\beta_0$  can be obtained. Definition of the allowed  $\Delta P$  determines (e.g., from the graph) the allowed  $\sigma_{\text{sensor}}/\sigma_{\text{scene}}$ . An estimation of the scene noise for which the other conditions apply allows the calculation of the total sensor noise allowed. The final step is to partition this noise between sensor random noise and quantization noise.

For example, let the desired  $P_0 = 85\%$  and allow no more than 2% loss due to the total sensor noise. The no-sensor-noise  $\beta_0$  must be  $\geq 5.7$  to give  $P_0$ . Then, from Figure S-4, the allowed  $\sigma_{\text{sensor}} = 0.6 \times \sigma_{\text{scene}}$ . If the scene has a  $\sigma_{\text{scene}} = 2\%$ , the allowable  $\sigma_{\text{sensor}} = 0.6 \times 2\% = 1.2\%$ , which must be partitioned between  $NE\Delta\rho$  and the quantization noise. For  $NE\Delta\rho = 1\%$ , the allowable  $\sigma_{\text{quant}} = \sqrt{1.2^2 - 1^2} = 0.66\%$ , which can be met by 6-bit quantization.

Two observations are important here: (1) Increasing the number of bits of quantization produces improvements which asymptotically approach zero, as each successive bit reduces the step size by a factor of 1/2. (2) A scene having as little as 2% variation is a very uniform scene. Since this noise is rms'd with the sensor noise, it will overwhelm any but a very noisy sensor. Therefore, for purposes of multi-spectral classification, more than six bits would seem to be unnecessary.

#### EDGE EFFECTS

To this point, the analysis is based on pixels well inside uniform fields and well away from field boundaries. A number of experimenters have spent appreciable time discovering that classification accuracy falls off at boundaries due to what has become known as the mixed-pixel effect. We will start at that point and attempt to model the effect to allow us to quantify our expectations.

We assume as a starting point that all the spectral bands used in classification, whether obtained from one date or series of dates, are in perfect registration. This means that when the pixel grids from each band are aligned the data contents (field borders, roads, all features) are also aligned - note that this is more than simply having all internal distortions removed, which is all that most geometric rectifications accomplish. Misregistration will (later) be considered as the lack of alignment of the pixel grids; because the computer can only work with pixel grids, aligning these pixel grids because appear to the computer as a shift in the boundaries. We will assume that training samples are accurate and that class limits have been set from these by the classifier chosen. The classification is modelled as follows: signature shifting in any individual band will tend to cause misclassification, so that the situation may be treated one band at a

ORIGINAL PAGE IS  
OF POOR QUALITY

time. The effects of pixel mixture in all bands may then be rms'd together if desired. The entire analysis simplifies to the consideration of the transient intensity shift across field boundaries as compared to the class limits and the noise components of the measurement.

The first step in analyzing the spatial extent of pixel mixing across borders is to estimate the shape and extent of the transient intensity shift. If the impulse response functions or the modulation transfer functions (MTFs) of the various components (and, hence, the entire system) are known, a precise transient response may be calculated. For example, the specifications for the Thematic Mapper for Landsat D call for a 2% to 98% time equivalent of about 2 pixels implying a 10%-90% transient response of about 1.3 pixel. The practical result of this is that the "infinitely sharp" edges of the real scene will be softened by the filtering effect of the scanning aperture (assumed to be rectangular and having uniform response) and it is this softened transient response which is sampled. Interpolation required for registration will cause some further softening, and the use of any of the competent higher-order interpolation functions ( $\sin x/x$ , TRW cubic convolution, modified cubic convolution, other splines) will have minor effects of the rise time. A total  $\tau_{10-90}$  (transient response from 10% to 90%) of 1.5 pixels with no ringing will be used as a surrogate global value.

The transient situation across a border is sketched in Fig. S-5. We are concerned here with the decrease in probability that a given pixel will have a value within the class limits as that pixel moves toward the boundary, as shown in Figure S-6. The analysis only needs to determine the area under the normal curve (assuming the noise is Gaussian) between the limits as determined by the classification class size and the offset from the "field interior value" caused by the mixture. The important scaling involved is the amount of signal shift caused by the transient total shift  $T$ , as related to the desired class size  $S$ , for a given  $\beta$ . The left portion of Figure S-7 reflects this shift in brightness (vertical axis) as it affects the area within the class (the probability of recognition).

The transient rise distance estimated for the Thematic Mapper has very close to a Gaussian shape and a  $\tau_{10-90} = 1.5$  pixel. The amount of brightness shift is the difference between the brightness of the field under consideration and the adjacent field which is causing the shift. The important intensity relation is the magnitude of this shift,  $T$ , as related to the size  $S$  of the class being tested by the ratio  $T/S$ . These curves, for various  $T/S$ , are combined with the probability curves of the previous discussion in Figure S-7. From this may be estimated the loss in probability in classification of pixels near borders.

#### BIAS IN FIELD SIZE ESTIMATION

It can be appreciated that several things are happening simultaneously: If the lower limit of field B and the upper limit of field A have a gap between, pixels "lost" by field B will not be picked up by field A, and will be considered unknowns and not be counted in either field. The lost pixels will be some interior pixels, due to insufficient  $\beta$ , and a large number of near-border pixels, resulting in apparent field size loss. Only if the lower limit of field B and the upper limit of field A are coincident will pixels lost from one field be picked up by the other, and vice versa, to give

ORIGINAL PAGE IS  
OF POOR QUALITY



complete account of all pixels. For the field size estimator to be unbiased, the loss-and-pickup in both directions must cancel; that is, on the average the true border must be located. The total effect will depend on the ratio of the number of border pixels to the number of field-interior pixels, and hence is a function of the field shape and size.

This leads directly to the required algorithm for field size estimation: First divide the scene into blobs, each of which is sufficiently uniform, and with closed boundaries. Then for each blob (field) determine the average brightness for all the interior pixels which are safely away from the border. For each segment of the border, the correct field edge decision level is midway (in  $\sigma$ 's) between the average brightness of the two fields on either side. After the borders are located using this criterion, the field interiors may be reclassified using the classification limits as determined from the training samples.

#### EFFECTS OF MISREGISTRATION

In preparation for estimation of the misregistration effects, an analysis will first be made of the expectations of registered data and the sensitivity to the various parameters estimated. The starting model used has rectangular fields aligned with the pixel grid. Pixels are grouped into four zones: 1) Interior (i)-those with centers 2 or more pixels inside borders, 2) Inner border (ib)-pixels with centers 1-1/2 pixel inside borders, 3) Outer border (ob)-pixels with centers 1/2 pixel inside borders, 4) Exterior border (xb)-pixels outside the borders, with centers 1/2 pixel outside. Estimates of classification accuracy for each zone are obtained from Figure S-7. The total estimate of classification accuracy is the sum of pixels in each zone multiplied by the corresponding zone accuracy estimate. Later, the field will be misregistered, changes in the number of pixels in each zone calculated, and the probabilities again summed. The following parameters are required:

- $r$  - the field shape ratio, length of long side/length of short side
- $T$  - transient brightness difference between field being considered and its neighbor
- $S$  - decision class size
- $\tau$  - transient distance for 10% to 90% response
- $\beta$  - class size  $S/\sigma$  of Gaussian noise

The following global values selected for the parameters are considered to be representative:

- $r$  = 2
- $T/S$  = 1 to 5
- $\tau$  = 1.5 pixels
- $\beta$  = 3 to 5

After the parameters  $r$ ,  $T/S$ ,  $\tau$ , and  $\beta$  are selected, the resultant (from Fig. S-7) probabilities are substituted for the brightnesses in the various zones to produce a "probability image" aligned with the desired output pixel grid. The probability assigned to a pixel at a given location represents the probability that that pixel will have a brightness falling within the classification limit determined by the classifier, for the given spectral

ORIGINAL PAGE IS  
OF POOR QUALITY

$$P = \frac{1}{n_1^2} \left[ p_{11}n_{11} + p_{1b}n_{1b} + p_{ob}n_{ob} + p_{xb}n_{xb} \right]$$

where  $n_1$  is the field width (short side) in pixels, and  $n_{11}$ ,  $n_{1b}$ ,  $n_{ob}$ ,  $n_{xb}$  are the number of pixels in the various zones. Using these values, the global estimate of the probability of correct classification with no misregistration is given Figure S-8 for three values of T/S. The predominant effect is the pixel mixture (the effect of T/S). As expected, this is worst for small fields ( $n_1$  small) because of the larger percentage of border pixels for these fields. Note that for T/S = 1, decision level midway between brightnesses of adjacent fields, no probability loss occurs, even with small fields. Unfortunately, this desirable condition cannot be systematically obtained.

#### MISREGISTRATION OF CONGRUENT FIELDS

The initial model for misregistration is a displacement of  $d$  pixels, equal in both  $x$  and  $y$ . The result of this misregistration is that some area is lost from the external border, causing a further classification accuracy decrease. The misregistration loss as seen by the external border loss is given by

$$\Delta P = p_{xb} \left[ d \frac{\tau + 1}{\tau} \frac{1}{n_1} + (4d - d^2) \frac{1}{n_1^2} \right]$$

The basic character of this misregistration loss term is  $1/n_1$ , so that it will have a slope approximately equal to  $-1$  on a log-log plot vs  $n_1$ . The precise results depend critically on the values of  $p_{xb}$  estimated for the  $p_{xb}$  from Figure S-7:

T/S	$\beta$	$\tau = 1$	$\tau = 1.5$	$\tau = 2$
1	3	.10	.14	.20
	5	.02	.025	.07
	7	0	.01	.04
2	3	0	0	0
	5	0	0	0
	7	0	0	0

Using these values, the loss  $\Delta P$  due to displacement misregistration is plotted in Figure S-9 for various parameter combinations.

#### MISREGISTRATION DUE TO NON-CONGRUENCE

##### 1.) SIZE AND RATIO (ASPECT) CHANGES

Size and aspect ratio changes can come about from several causes such as scan velocity or altitude changes, and if uncompensated can cause additional

ORIGINAL PAGE IS  
OF POOR QUALITY

misregistration errors. Progressive misregistration from a point of accurate registration will be caused by both causes (Figure S-10a); the modeling of this effect considers first that size changes  $N = n'/n$  will cause a shift in points  $n$  to points  $n'$  both vertically and horizontally, and then that changes in aspect ratio will cause further shifts in the horizontal position of vertical borders by changing the field shape ratios by the factor  $R = r'/r$ . The resulting shifts are:

$$\Delta n_v = (N - 1) n_v \quad \text{and} \quad \Delta n_h = (NR - 1) r n_v$$

For analysis, this shift will be divided around the borders symmetrically as optimum field registration is accomplished (Figure S-10b). Two cases must be distinguished (using scan velocity as a surrogate cause):

Case I: A slow scan decreases pixel spacing and puts more pixels into a given field. When these are placed into the output grid, the field appears stretched. The field as defined by the other (correct) bands now covers only part of the stretched field, so that the classification tends to see only interior pixels, and the accuracy will increase, ultimately reaching the field-interior accuracy. The sizes of the border errors are:

$$e_1 = \frac{1}{2} (N - 1) n_1 \quad \text{and} \quad e_2 = \frac{1}{2} (NR - 1) r n_1$$

Case II: A fast scan has the opposite effect, causing the field to appear smaller and the analysis pixels defined by the other bands now include more exterior pixels. The classification accuracy will decrease.

For fast scan, the smaller apparent field covers an area expressed as a fraction  $f_i$  of the total:

$$\text{Fractional Areas: } \begin{cases} f_1 = \frac{r'(n'_1)^2}{r n_1^2} = RN^2 & \text{(Interior)} \\ f_{xb} = \frac{2Nn_1 + 2NRn_1r + 4}{r n_1^2} & \text{(External Border)} \end{cases}$$

The total expected probability is

$$P_{tot} = f_i P_i + f_{xb} P_{xb}$$

Since the external border pixels are now included within the analyzed field, but with a low probability, the fractional area  $RN^2$  represents approximately the fraction of the basic field-interior accuracy to be expected. Since the total size shrinkage (in pixels) is small for small  $n_1$ , only larger  $n_1$  need be considered, and the  $1/n_1^2$  term may be dropped.

ORIGINAL PAGE IS  
OF POOR QUALITY

$$P_{tot} \approx RN^2 p_i + \frac{3}{n_1} p_{xb}$$

For large fields, the probability is seen to be independent of field size, and only weakly dependent (because of low  $p_{xb}$ ) for small sizes.

## 2.) WAVY BORDERS AND MULTIPLE ACQUISITIONS

For single-band analysis, with borders distorted so that there are pixels both inside and outside of the analyzed area, some pixels will have increased probabilities of correct classification and some will have less. The decrease in probability across border is (very) approximately linear, so that the (signed) average displacement will model the effect.

For multi-band analysis, those pixels having a low probability of classification will have the largest effect as the net probability at each pixel location is the product of the probabilities obtained for each acquisition (band). In this case the rms displacement will produce a better model of the effects.

## SOME OBSERVATIONS

### I. ON BASIC CLASSIFICATION

- The total noise figure (compared to the class size in a given determination) controls  $\beta$ , and in turn controls the maximum attainable classification accuracy. However, for practical range of  $3 < \beta < 7$ , increasing  $\beta$  has only a moderate effect.
- Because of this, if small fields are most important, the reflected energy might be more profitably be divided into smaller pixels, even at the expense of  $NE\Delta\rho$ . As this will cause an increase in data rate, optimum coding should be investigated. The possible noise introduced in reconstructing the data will cause some further decrease in the overall effective  $NE\Delta\rho$  and so decreases  $\beta$ . But since there is smaller sensitivity to  $\beta$  than to  $1/n_1$ , there should be a net gain in utility.
- Increasing the number of bits of quantization produces improvements which asymptotically approach zero, as each successive bit reduces the step size by a factor of  $1/2$ .
- A scene having as little as 2% variation is a very uniform scene. Since this noise is rms'd with the sensor noise, it will overwhelm any but a very noisy sensor. Therefore, for purposes of multi-spectral classification, an extreme number of bits would seem to be unnecessary.

### II. ON EDGE EFFECTS

- For accurate field size estimation, the decision brightness must be halfway between the brightnesses of the fields on either side of a given boundary. This means that classifiers set for material identification will in general produce errors in field size. But the

ORIGINAL PAGE IS  
OF POOR QUALITY



field-interior brightness is increasingly hard to estimate for small fields because of the fewer interior pixels.

- It is important to keep the transient response distance and the accompanying sample spacing small, to get as many pixels into a given ground distance as possible. Field area errors become large at  $n_1 = 5$  or less. The transient distance must also be matched between spectral bands.
- At the resolution expected for the Thematic Mapper, the atmospheric point spread function may become more dominant than the Thematic Mapper point spread function. If this is determined to be true, the registration requirements may be relaxed since scene-dependent registration will be required anyway.

### III. ON MISREGISTRATION

- For large T/S (i.e., 2 or more) the edge effects are so great that the base probability is drastically affected, and the external border pixels have zero probability of being within the class limits. For this reason, there is no misregistration effect for large T/S.
- Square fields show the most misregistration loss, when scaled to  $n_1$ .
- A shape ratio  $r=2$  is believed to be representative.
- Misregistration loss decreases with higher  $\beta$ . However, these losses in general are small to begin with, and the discussion calling for sacrifice of  $\beta$  to gain smaller IFOV (more pixels  $n_1$  into a given field) would seem to override.
- Increase in  $\tau$  decreases the basic accuracy of edge pixels and also increases the misregistration losses.
- Geometric rectification and registration procedures must not only remove the internal distortions but must also produce pixels on a defined (preferably ground-referenced) grid. Current procedures do not do this. Without this reference grid, users will have to re-interpolate before multi-temporal data can be compared.
- Scale and aspect ratio errors will have only minor effects on moderate-area problems. But they will cause problems in correlating over large distances.
- Altitude relief displacement will require users to use many control points to register images in areas of high relief.
- Unless standard reference grids are established, users requiring registration will have to interpolate every image, even in low relief areas.
- For single-band analysis, the algebraic average of the displacement may be used. For multi-band analysis, with erratic errors in location among the bands, the lowest probability of correct classification holds and the rms of the displacements is appropriate.

ORIGINAL PAGE IS  
OF POOR QUALITY

# AN UNANSWERED QUESTION

This report models the potential misregistration effects on multispectral classification accuracy. It may allow the comparison of the various tests and simulations, and points out the variables which must be reported for those simulations to allow their validation. It does not answer the following question: Given a certain loss in accuracy due to misregistration, how does that damage the ability to use the data analysis results? These evaluations will be discipline dependent, and must be sought separately.

## REFERENCE

Friedman, H.D., On the Expected Error in the Probability of Misclassification, Proc IEEE Vol 53, p. 658.

Hixson, M., Scholz, D., Fuhs, N., Akiyama, T., Evaluation of Several Schemes for Classification of Remotely Sensed Data, Photogrammetric Engineering and Remote Sensing, Vol 46 # 12, Dec 1980, pp 1547-1553.

ORIGINAL PAGE IS  
OF POOR QUALITY

Summary of a Heuristic Method for the Comparison of Related Structures  
by Golledge and Hubert as presented by Strahler.  
Department of Geography, University of California, Santa Barbara, CA

Basically this work involves a strategy to evaluate data analysis schemes that are supposed to clarify the structure underlying a set of proximity measures.

As Alan sees it, there are four categories of possible applications:

- (1) Given two different analyses based on a single data source, does either represent the data significantly better than the other?
- (2) Given 2 analyses from a single data source, which is closer to a specific theoretical structure assumed to underlie the data?
- (3) Given 2 theoretical structures and one analysis does either represent the analysis better than the other?
- (4) Given a single analysis based on one data source, is the information present in the data accounted for satisfactorily by the analysis?

The original basis for this type of work was laid out in a method developed by Wolfe (1976, 1977)

$$P_{u,v-w} = \frac{P_{uv} - P_{uw}}{2(1-P_{vw})}, P_{vw} \neq 1$$

$u, v, w = 3$  random variables with some joint distribution, in which

$$\text{Var}(V) = \text{Var}(w)$$

Thus,  $P_{uv} = P_{uw}$  if and only if  $P_{u,v-w} = 0$ .

This means that a test of equality between  $P_{uv}$  &  $P_{uw}$  reduces to a test of  $P_{u,v-w} = 0$ .



ORIGINAL PAGE IS  
OF POOR QUALITY

- This test can be performed under normality.
- But, we can also use RANDOMIZATION to test this nonparametrically.
- To do this, we estimate  $P_{u,v-w}$  as  $r_{u,v-w}$

$$r_{u,v-w} = \frac{r_{u,v} - r_{u,w}}{\sqrt{2(1 - r_{v,w})}}, \quad r_{u,w} \neq 1$$

- we form a vector of values  $V-W$ , and permute randomly, forming many  $r_{u,v-w}$ 's
- Then we compare  $r_{u,v-w}$  with the distribution obtained by permuting  $U$ , and attach a significance level based on these permutations.
- If significant, conclude that  $r_{uv}$  and  $r_{uw}$  are different.

#### Details:

- There are  $N!$  permutations of  $U$ . ( $N$  is number of samples). Instead of enumerating  $r_{u,v-w}$  for all permutations, just do a reference sample.
- $V$  and  $W$  must have equal variance. Therefore, standardize them.

#### Extension of Matrix Concept

- Assume set  $S$  of  $n$  objects, and there proximity matrices  $A$ ,  $B$ , and  $C$  defined on  $S \times S$ . Assume main diagonal  $a_{ii} = 0$ .
- Define  $r_{A,B}$  = simple correlations between elements of  $A$  or elements of  $B$ .
- Replace elements of  $B$  and  $C$  by their Z-sums.
- To estimate distributions of  $r_{A,B-C}$  permute more \_\_ columns simultaneously in  $A$ .

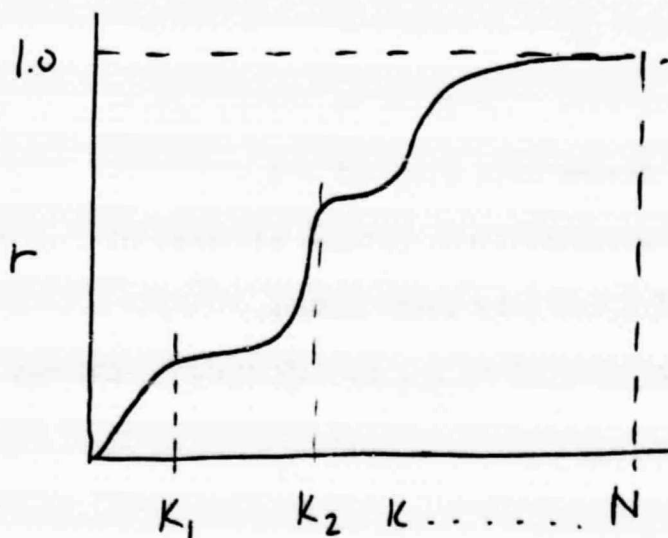
### REMOTE SENSING EXAMPLES

#### (1) Divergence: (A.H.S.)

- Let  $V$  be a classification of training data based on  $K$  channels; let  $W$  be a classification based on  $K+1$  channels.
- Assign  $V-W$  = class no. when classes agree; assign  $V-W$  when classes do not agree.
- Use  $\phi$  coefficient instead of  $r$  between  $U$  and  $W-W$ ; permute  $U$  to establish reference distribution.
- Problem of degrees of freedom--difference may be significant due to degrees of freedom--ability to fit a higher-order model. However, if difference not significant, this is very helpful.

#### (2) Classification (A.H.S.):

- How many classes are there in these data?
- Let  $K$  classes number from 10 to 50. Find "best"  $K$  using randomization for  $r_{A, D_{i-1}, i}$
- "D" is distance matrix between each point and class of which it is supposed to be a member



*E.g.,  $K_1$  and  $K_2$  are possible choices.*

- Compare  $U-W$  with  $U-r_{u,u-w}$ .
- Permute  $U$  to get distribution of  $r_{u,u-w}$ .
- Assign confidence level to  $r_{u,u-w}$  based in sampled distribution.
- (Note that this could be used to compare alternative models as well).

The Sections which follow, summarize the discussions of important research issue growing out of these workshops and present our working groups conclusions concerning area of research need. This is followed by a section listing specific prioritized research areas demanding attention.

ORIGINAL PAGE IS  
OF POOR QUALITY

WHAT WE LEARNED ABOUT YOUR INFORMATION NEEDS

Josef Cihlar\*  
Canada Centre for Remote Sensing  
Ottawa, Ontario.

1. INTRODUCTION

The procurement of accurate and representative intelligence about information needs of remote sensing data users is a difficult task. A major reason for this difficulty is the high diversity of the user community and its information requirements, both between disciplines and among users/user agencies within disciplines. Secondly, the needs are not constant but evolve with time in quantitative as well as qualitative terms. Thirdly, the information requirements are often difficult to quantify in the manner appropriate for remote sensing techniques.

Discussions with managers and specialists involved in resource management are a very effective means of obtaining reliable information concerning information needs. It is for this reason that the Planning Committee of this conference decided to give the participants an opportunity for feedback about their views and concerns. A questionnaire (see Appendix) was prepared prior to the conference to ensure uniformity in discussions of various groups. The questionnaire was distributed following the banquet on 29 October to 16 tables and each group was asked to answer the questions based on their knowledge and experience. The discussion lasted about one hour. A precedent for this format has been set at two recent annual meetings of the Soil Conservation Society of America.

---

\*Member of the Planning Committee of the Conference

## 2. RESULTS

### 2.1 Participants' Backgrounds and Experience

According to their professional orientation, the 132 discussion participants could be divided into nine groups:

Agriculture, Rangeland, Farming .....	12 %
Forestry .....	9 %
Soils .....	14 %
Hydrology, Water Quality, Civil Engineering .....	5 %
Ecology, Wildlife .....	9 %
Geography, Land Use, Urban Planning .....	17 %
Geology, Geomorphology .....	10 %
Photogrammetry, Education, Remote Sensing .....	18 %
Other (Business, Economics, Meteorology, Oceanography) .....	6 %

The participants indicated the following degrees of experience in using remote sensing products:

Satellite Products:	no experience .....	13 %
	small experience .....	39 %
	large experience .....	48 %
Airborne Products:	no experience .....	4 %
	small experience .....	28 %
	large experience .....	68 %

It was also evident that the types of remote sensing products used differed, digital satellite data being used much more frequently than digital airborne data (note that one person could have worked with both types):

Satellite data used:	digital form .....	57 %
	photographic form .....	69 %
Airborne data used:	digital form .....	26 %
	photographic form .....	91 %

From the above statistics, it appears that this conference attracted largely professionals with some degree of remote sensing expertise. Individual resource disciplines were well represented. Since the type of work performed by the participants was not addressed in the questionnaire, it is not possible to determine the relative proportions of management personnel and technical specialists.

## 2.2 Resource Information Requirements

Question 1 concerned information which are not provided by existing methods but could be supplied through satellite and airborne remote sensing. As expected, a wide variety of information needs were listed. The lists were subsequently reviewed to summarize common requirements under separate headings. Although an attempt was made to obtain clear and specific descriptions of information needs, the answers were vague in numerous cases. The results are reproduced below in a form consistent with the original responses. A slash (/) is used to indicate that the requirement was stated by more than one group but not all groups used the modifier following the slash.

**ORIGINAL PAGE IS  
OF POOR QUALITY**

### 2.2.1 Agriculture, Rangeland, Farming:

- (i) real or near-real time information (preferably at pre-visual times) on weeds, insects, defoliation, disease, moisture stress, herbicide and fertilizer mis-application;



- (ii) accurate, timely/annual crop (all types) estimates by production areas relevant for private and public decisions;
- (iii) information required for planting decisions (planting rates, fertilizer applications, field conditions, nearby weed infestations);
- (iv) range forage production and grazing capacity, inventory, trend assessment over large areas and location of brush in pastures;
- (v) crop yield prediction through modelling for both domestic and international crops;
- (vi) pipeline construction effect on long-term crop production;
- (vii) quantitative information on residue amounts;
- (viii) more accurate weather prediction.

2.2.2 Forestry:

- (i) infestation and plant stress;
- (ii) statewide/annual timber volume monitoring including forest depletion;
- (iii) rate of tropical deforestation;
- (iv) forest energy assessment;
- (v) fuel type condition monitoring;
- (vi) better subdivision of forest land cover categories.

2.2.3 Soils and Geology:

- (i) erosion: critical areas, statewide erosion class changes, effectiveness of conservation measures;

ORIGINAL PAGE IS  
OF POOR QUALITY

- (ii) depth to bedrock and to water table;
- (iii) surficial and shallow subsurface drainage;
- (iv) soil moisture over large areas;
- (v) rapid soil mapping in poorly accessible areas, and mapping vegetation and rocks in support of soil mapping;
- (vi) location and mapping of subsurface minerals.

2.2.4 Hydrology, Water Quality, Civil Engineering:

- (i) surface water parameters: depth, suspended solids, pH, alkalinity, heavy metals, P<sub>2</sub>O<sub>5</sub> content, and lake surface temperatures over large areas;
- (ii) snowpack water equivalent, and snow depth;
- (iii) irrigated acreage by crop, type of irrigation system, and source of water;
- (iv) regional hydrology monitoring (Wyoming, Montana);
- (v) accurate flood plain delineation, and areal coverage for flood events;
- (vi) ice hazard detection (oceans);
- (vii) land water resources data.

2.2.5 Ecology and Wildlife:

- (i) biomass estimation;
- (ii) vegetation mapping over large, remote areas and mapping over small areas at large scale;
- (iii) habitat cover maps for large areas; monitoring habitat changes;
- (iv) mapping age classes in brush communities;

- (v) wetlands plant types inventory;
- (vi) changes in critical coastal zone areas;
- (vii) success of mines reclamation.

2.2.6 Geography, Land Use, Urban Planning:

- (i) statewide land cover maps;
- (ii) better land mapping methods;
- (iii) level II land use maps of urban areas and change detection in urban areas;
- (iv) land cover (at 5 acre resolution) of crop types and crop rotation for soil erosion assessment.

2.3 Technology Requirements

Numerous items listed in response to Question 1 are more appropriately characterized as technology requirements. The first five were mentioned by more than one group:

- (i) all weather data for timely applications;
- (ii) improved spatial resolution;
- (iii) increased coverage frequency;
- (iv) better timeliness, near-real or real time including fast data delivery;
- (v) continuity of data;
- (vi) satellite data reception in the field;
- (vii) worldwide, consistent coverage;
- (viii) reliable data formats and better distribution of remotely sensed data (a library of images desirable), and better information on image quality before purchase;

- (ix) integration of existing ground and aircraft data;
- (x) whole-frame registration, and better registration of ocean scenes;
- (xi) other spectral bands (mid-IR, thermal IR);
- (xii) interface Landsat data with geographic information systems, and improve transferability of data;
- (xiii) package remote sensing techniques;
- (xiv) make available user-oriented products;
- (xv) nationwide data base for resource applications;
- (xvi) inexpensive, common format mini-processing systems;
- (xvii) regional environmental monitoring centres.

#### 2.4 Major Obstacles

Question 2 requested a list of "major obstacles which must be overcome in order to make remote sensing a practical and useful tool in your work". The first 10 items listed below were mentioned by more than one discussion group:

- (i) continuity of data;
- (ii) lower costs;
- (iii) communication gap between researchers and users, including better publicity for applications of digital data;
- (iv) cheap, subsidized data that the user can afford;
- (v) recognition that remote sensing is a supplement to existing systems and rarely a substitute;
- (vi) long-term, active federal commitment to support an operational system, and better technology transfer procedures;

- (vii) understanding and acceptance of the need for remote sensing methods by people in key positions and the provision of support;
- (viii) institutional problems: disagreements concerning leading role and lack of coordination among agencies;
- (ix) locally available expertise in remote sensing;
- (x) recognition that remote sensing offers products which need promotion (currently non-existent) to be funded;
- (xi) end user must know how to use remotely sensed data as well as their advantages and limitations;
- (xii) better education at universities and (?) high schools, and general public awareness;
- (xiii) opportunities for training and for users to work with remotely sensed data on their own problems and in their own environments;
- (xiv) lack of funds and time;
- (xv) technology not directed at the grass roots level and too much emphasis on high technology;
- (xvi) unwillingness of the private sector to commit to new technology, and unwillingness of the government to explore technology transfer to industry;
- (xvii) user clout in setting priority for remote sensing data availability;
- (xviii) technology transfer at minimum cost and low cost, timely data for all users.

ORIGINAL PAGE IS  
OF POOR QUALITY

3. CONCLUDING REMARKS

An accurate and comprehensive specification of information requirements cannot successfully be provided in a brief discussion such as reported here. Nevertheless, the list of information requirements given above has two unique aspects. First, the requirements represent problems of current concern to resource managers. Secondly, the list was compiled by a broad spectrum of resource managers and is therefore relevant for a range of disciplines and resources.

The major obstacles and many technological requirements were identified quite clearly and consistently. It seems evident that although remote sensing methods have shown considerable potential for resource management applications, a lot more work must be done in making the appropriate methods available to agencies and individuals responsible for operational work. Only when this is accomplished will remote sensing yield the anticipated benefits from improved resource management.

Many participants were also concerned about the cost of future remotely sensed data. The requirement for inexpensive data was echoed by the majority of discussion groups. This is an important consideration, especially now when plans for operational satellite remote sensing are being finalized in the United States.



DRAFT LETTER FROM IDA HOOS

REC'D 2/27/81

Understandably, scientists and engineers charged with the responsibility of devising and designing satellite and other technology capable of responding to future needs would like a scientific basis for evaluating present data systems and assessing the value of the information they can deliver. At first glance, this might appear to be a relatively simple research task; one would have only go to the users, ascertain their needs, and learn about their present and prospective requirements.

(a) The above may look reasonable but is entirely unrealistic. It is based on nonexistent premises and untenable assumptions, not least among which is the one that might best be called technical optimism. Most of the "capability" of Landsat, for example, to serve as a unique source of information superior to any other is still in the q.e.d. category; it is still to be demonstrated. The "test cases" do not yet stand up well under test. The most ardent advocates are not users willing to pay for systems and service but "brokers" and technical middle people who have a vital role in the process of technology transfer but who are pushers rather than users.

(b) Approaching the matter from the viewpoint of political reality, one might go so far as to state that we cannot honestly evaluate the systems in place for fear of perturbing precarious relationships or offending participating agencies. When, for example, NASA has worked out a difficult cooperative arrangement with the Veteran's Administration for the promotion of some device, "research and evaluation" are regarded with trepidation for fear something other than a valentine be produced. The "user needs surveys" conducted in conjunction with the NASA-NOAA transfer, the presentations at our workshops, and our cumulative experience as professionals and academics should convince us that in order to perform a proper analysis we must establish social ground

truth as painstakingly and meticulously as do the technical people in their assessments. We also know that neither time nor circumstance allows for such a course of action.

(c) In a "management environment" where cost/effectiveness is the iron law, justification for present and especially future generations of satellite technology by reference to value and returns is tantamount to quantifying pie-in-the-sky. It was difficult enough to compile creditable numbers to satisfy Senator Proxmire. Nor was it an easy task with the traditionally axe wielding OMB. What we can expect from not only Mr. Stockman but even from the public-at-large is already the handwriting on the wall. In an era where NASA's stock is low, it is unrealistic to count on big investments and quixotic to make claims that cannot be \_\_\_\_\_ substantiated. Technology \_\_\_\_\_ cannot be rationalized by non-existent demand. The public has its own perceptions of punitives and in the present political climate not all premises carry the same amount of credibility.

(d) Information on the international scene appears to carry more threat than promises. Books like The Geopolitics of Information (Geoffrey Smith, Oxford University Press, 1980), reports like "Information and the New World Order," and sessions of world organizations attest to the enormous concerns being generated by the gathering and dissemination of information about resources. Nations are concerned about the possibility of "economic espionage." New alliances and interdependencies are foreseen. Simply to offer all countries equal access at the same cost as is to act in a discriminating manner. These are factors that must be taken into account in the calculation of cost/benefit ratios of future generation of satellite technology.

But this brings us back into the morass of factors impeding proper assessment of the use and value of information. And this probably explains the

dearth of such evaluations. The social context renders them mutually impossible. But, is this to say then that none should be done? Quite the contrary. Starting with the premise that the social context is crucial, we might do well to consider another social context -- not the one that serves as a kind of gridlock but one that provides us with what seems to be a generally agreed upon paradigm for the future. The world new of the future as envisioned in the Global 2000 report might be a good framework. It is the global situation needs of the year 2000 and on to which tomorrow's technology must respond, and it is clear that remote sensing and the information it can provide would be a key factor in meeting the challenges which transcend all national borders and internicine warfare among government agencies. Because it is the far horizon that has more certainty and for which there is greater agreement than the present, we might consider it as the framework and thus unfetter ourselves from preoccupation with fabricating a story designed to please rather than to enlighten.



Energy, Mines and  
Resources Canada

Énergie, Mines et  
Ressources Canada

Science and Technology Science et Technologie  
CANADA CENTRE FOR REMOTE SENSING

December 9th, 1980

Your file Votre référence

Our file Notre référence  
5072

Dr. J. Estes,  
University of California,  
Santa Barbara, California,  
93106, U.S.A.

Dear Jack,

Attached is a manuscript "What We Learned About Your Information Needs" which was written on the basis of the group discussions at the Kansas City Conference. Given your current interest in information requirements and processing, you may find this useful.

Best regards,

Josef Cihlar, Head,  
Applications Development Section.

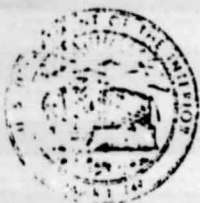
Attachment (1)

JC/ma

2464 Sheffield Road  
Ottawa Canada  
K1A 0Y7

2464, rue Sheffield  
Ottawa Canada  
K1A 0Y7





# United States Department of the Interior

GEOLOGICAL SURVEY  
RESTON, VA 22092

In Reply Refer To:  
EGS-Mail Stop 710

April 7, 1981

## Memorandum for the Record

Subject: Meeting report

Visitor: Professor John E. Estes, Department of Geography, University of California, Santa Barbara, California

Time of visit: April 3, 1981, 3:00 to 4:00 p.m.

Purpose of visit: To discuss topics for basic research in accuracy of digital classification.

Jack Estes introduced our conversation with the information that he is involved with NASA on the process of information, utilization, and evaluation of digital classification of Landsat data. He is looking for research projects that will carry into the 1990's.

I first discussed that there is no standard for defining accuracy and that it means different things to different people. Hard and Brooner, 1976, PE&RS, first documented an approach to accuracy as the lower of the two-sided confidence limits about the observed sample proportion correct. I consider accuracy to be the observed sample proportion correct, with associated confidence limits. Brian Berry recommended to Fitzpatrick-Lins, 1980, PE&RS, that the lower one-sided confidence limit be established about the observed sample proportion correct when the errors of omission are not considered. Berry (1979), in his review of my draft paper, did not necessarily agree with the point-in-polygon accuracy test, and stated that his local computer systems should be used.

I then recommended, as a long range research project, that he propose bringing digital classification to the status of a mature mensurational system. I stated my concept that a mature mensurational system is one where the mathematics and statistics are completely rigorous and general in application; where the error budget is completely known and understood; systematic errors are calibrational and correctable; random errors are known and their effect can be propagated; and finally, that the system is completely documented as to theory, analysis, and computer program. I referred him to my paper, "The Role of Software in Photogrammetry," 1967, PE&RS, for the philosophy involved for a mature mensurational system. That paper concerns positional statistics, but the concept is applicable to thematic remote sensing. I also referred him to my paper, "The Ballistic Camera Accuracy Review Project," 1964, PE&RS, for consideration of the massive effort involved in upgrading a mensurational system, and bringing it to the mature stage.

ORIGINAL PAGE IS  
OF POOR QUALITY

ORIGINAL PAGE IS  
OF POOR QUALITY

2

I then discussed other smaller research problems: 1) investigating the probability distribution for systematic sampling since the binomial distribution is based on simple random sampling; and also to consider sampling in two frames, as we do in the Ling sampling algorithm. 2) considering the entire classification error matrix, instead of only the diagonal elements as most people do. Roy Mead and staff at VPI&SU use one approach to studying the entire matrix, and I am looking at another approach. Both a single matrix from a given accuracy test should be studied, as well as several matrices from an experiment to study different factors. A rigorous approach is needed to interpret the entire content of the classification error matrix on a probability basis.

Estes requested a copy of the paper "Summary Tables for Selected Digital Image Processing Systems," V. Carter and others, 1977, USGS Open-File Report 77-414.

He showed me in a document he had prepared for NASA (he will send me a copy), an illustration of a concept for analyzing a digital classification system. He stated that that section of the document discussed what we had talked about, but did not put it as succinctly as the concept of a mature mensurational system.

I wished him luck on his project.

*George Hf*  
George H. Rosenfield

cc: Witmer  
Guptill  
Place  
Wray  
Rosenfield  
Estes  
RGR



International Business Machines Corporation

555 Bailey Avenue  
P.O. Box 50020  
San Jose, California 95150

June 2, 1981

Dr. John Estes  
Geography Department  
University of California, Santa Barbara  
Santa Barbara, California 93106

Reference: Your Memo Dated May 18, 1981

Dear Jack,

If Nevin Bryant agrees, I believe that the material which I sent him on the referenced letter should be incorporated in the summary notes on the meeting at Santa Teresa Lab. A particular point to be made is that NASA should not spend its precious funds on research in data base management systems since there are so many other forces which will move these. On the other hand, I feel very strongly that demonstration systems need to be funded, both for the purpose of making data available, and secondarily, to provide a test bed for resolving practical (non-basic research) issues in data base management.

The following applies to the preliminary summary of proposed research areas:

1. Information and decision processes:

I believe the most likely approach and therefore the first priority is that associated with expert systems. In order for the general public or laymen "government agencies" to obtain value from satellite data, a great deal of highly technical assistance must be given. I believe the expert systems approach would be ideal and would make a very useful demonstration project. At the same time, this area is considered to be one of the more advanced areas of computer science discipline.

Dr. John Estes

-2-

June 2, 1981

I would be very doubtful that work to build an economic theory which treats information as a commodity will yield anything useful. Information is an entity which allows the better management of other more tangible and valuable assets. The value of information thus depends very much on what assets one is able to manipulate. I believe NASA would be very unlikely to obtain payoff from such work.

2. Data and Information Attributes:

I am concerned that all of the areas of research here are very dependent upon the particular context and application. I would strongly recommend that these areas be pursued within the scope of a demonstration project such that concrete examples can be found. I am not hopeful that much of general value can be established.

3. Data Base Management and Use:

Here, as said in my letter, I believe that NASA money should not be spent on basic computer science work since this field has so much activity in that area. Rather, the money should be focused on demonstration projects in which the research element would be investigations into the particular data type that comes from satellites. I believe all of the elements listed in this category would be good topics within a particular demonstration project.

4. Data and Information Systems Performance:

All of the areas listed here look useful and I have no sense of priority about them. As you can tell from my multiple suggestions of it, I believe the key to progress in this area is well-chosen demonstration projects

Dr. John Estes

-3-

June 2, 1981

which attempt to advance the state of practice, if not research, in reducing and providing data to end users. Since the future behavior of end users is so uncertain, I believe it is important to have demonstration projects to "do market research". I believe it is very unlikely that academic research can really isolate where the values are in such a broad and complex future context.

Sincerely,

Glenn Bacon

GCB:cap

# SUMMARY BREAKDOWN OF TOTAL PARTICIPANTS BY EMPLOYMENT

	PL1	PL2	AW	KC	SJW	PL3	HW	OTHER	TOTALS
PRIVATE INDUSTRY		2	2	3	7			2	16
FEDERAL GOVERNMENT	2	3	3	2	7	2	5	1	25
STATE AND LOCAL GOVERNMENT			3	1					4
UNIVERSITY	8	5	8	4	3	4	7	4	43
TOTALS	10	10	16	10	17	6	12	7	88

# SUMMARY BREAKDOWN OF INDIVIDUAL PARTICIPANTS BY EMPLOYMENT

PRIVATE INDUSTRY	14
FEDERAL GOVERNMENT	16
STATE AND LOCAL GOVERNMENT	4
UNIVERSITY	22
TOTAL INDIVIDUALS PARTICIPATING	56

INFORMATION AND DECISION PROCESSES WORKSHOP  
ASILOMAR, CALIFORNIA  
SEPTEMBER 1980

ATTENDEES

GLEN BACON, IBM Corporation, Santa Teresa  
ROBINSON BARKER, St. Regis Paper Company, Jacksonville, Florida  
JAMES T. BONNEN, Michigan State University  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
CHRISTOPHER CLAYTON, University of California, Santa Barbara  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California, Santa Barbara  
OSWALD GANLEY, Harvard University  
ROBERT MACDONALD, NASA Johnson Space Center  
MICHAEL J. MCCORMICK, Planning and Community Affairs, State of Washington  
C. BART MCGUIRE, University of California, Berkeley  
ROBERT POTTER, Water Resources, Sacramento, California  
LEONARD SLOSKY, Science and Technology, Colorado  
CHARLES VARS, Oregon State University  
DIANA WATTS, Space Science Laboratory, University of California, Berkeley  
DARRELL WILLIAMS, NASA Goddard Space Flight Center

REMOTE SENSING FOR RESOURCE MANAGEMENT  
SPECIAL SESSION MEETING  
KANSAS CITY, KANSAS  
OCTOBER 1980

ATTENDEES

WILLIAM ANDERSON, Technicolor Graphics, Sioux Falls, South Dakota  
MARION BAUMGARDNER, Laboratory for Applications of Remote Sensing, Purdue University  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
JOSEF CIHLAR, Canada Center for Remote Sensing, Ottawa, Ontario  
JOHN ESTES, University of California, Santa Barbara  
PEGGY HARWOOD, National Governors Conference, Washington, D.C.  
WILLIAM MACFARLAND, University of Missouri, Columbia  
EARL MERRITT, Earth Satellite Corporation, Washington, D.C.  
GENE THOMPSON, Missouri Farm Association, Jefferson City  
CHARLES VARS, Oregon State University



PLANNING SESSION #1  
SANTA BARBARA, CALIFORNIA  
APRIL 1980

ATTENDEES

HUGH CALKINS, State University of New York, Buffalo  
CHRISTOPHER CLAYTON, University of California, Santa Barbara  
GEOFFREY DUTTON, Harvard University, Computer Graphics Laboratory  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California, Santa Barbara  
ROBERT MACDONALD, NASA Johnson Space Center  
DENNISON PARKER, United States Department of the Interior  
RONALD SHELTON, Michigan State University  
DIANA WATTS, Space Science Laboratory, University of California, Berkeley  
CHARLES VARS, Oregon State University

PLANNING SESSION #2  
CLEAR LAKE, TEXAS  
JULY 1980

ATTENDEES

GLEN BACON, IBM Corporation, Santa Teresa  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
CHRISTOPHER CLAYTON, University of California, Santa Barbara  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California, Santa Barbara  
FORREST HALL, NASA Johnson Space Center  
IDA HOOS, Space Sciences Laboratory, University of California, Berkeley  
ROBERT MACDONALD, NASA Johnson Space Center  
BRUCE SCHEER, The Planning Economic Group, Boston, Massachusetts  
CHARLES VARS, Oregon State University



DATA BASE USE AND MANAGEMENT WORKSHOP  
SAN JOSE, CALIFORNIA  
JANUARY 1981

ATTENDEES

DELMAR ANDERSON, Central Intelligence Agency  
GLEN BACON, IBM Corporation, Santa Teresa  
RALPH BERNSTEIN, IBM Palo Alto Scientific Center  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
JOHN ESTES, University of California, Santa Barbara  
STAN HANSEN, Boeing, Seattle, Washington  
ADRIAN HOOKE, NASA Jet Propulsion Laboratory  
IDA HOOS, Space Science Laboratory, University of California, Berkeley  
RAY LORRIE, IBM Research, San Jose  
ROBERT MACDONALD, NASA Johnson Space Center  
BOB MYERS, IBM, Yorktown Research Center  
GENE RICE, NASA Johnson Space Center  
EVE SCHWARTZ, Fleet Numerical U.S. Navy, Monterey, California  
WILLIAM SHARPLEY, ESL/TRW, Mountain View, California  
DAVID SINNOTT, NASA Research Center  
DON WALKETT, Terra-Mar, Palo Alto, California

PLANNING SESSION #3  
HOUSTON, TEXAS  
FEBRUARY 1981

ATTENDEES

NEVIN BRYANT, NASA Jet Propulsion Laboratory  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California, Santa Barbara  
IDA HOOS, Space Science Laboratory, University of California, Berkeley  
ROBERT MACDONALD, NASA Johnson Space Center  
CHARLES VARS, Oregon State University

DATA AND INFORMATION PERFORMANCE WORKSHOP  
HOUSTON, TEXAS  
FEBRUARY 1981

ATTENDEES

FREDERIC BILLINGSLEY, NASA Jet Propulsion Laboratory  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
DONALD J. CLOUGH, University of Waterloo, Ontario, Canada  
WILLIAM COBBERLY, University of Tulsa, Tulsa, Oklahoma  
JOHN ESTES, University of California, Santa Barbara  
CECIL HALLUM, NASA Johnson Space Center  
IDA HOOS, Space Science Laboratory, University of California, Berkeley  
GLENN HOUSTON, NASA Johnson Space Center  
RAYMOND JESSEN, University of California, Los Angeles  
PATRICK O'DELL, University of Texas, Dallas, Texas  
ALAN H. STRAHLER, University of California, Santa Barbara  
CHARLES VARS, Oregon State University  
GAYLORD WORDEN, Department of Commerce, Washington, D.C.

INDIVIDUAL CONTACTS OUTSIDE OF MEETINGS AND WORKSHOPS

CARL BRONTHAVER, Sparks Commodities  
REGINALD GOLLEDGE, University of California, Santa Barbara  
ALAN MURPHY, Oregon State University  
ALAN OETTINGER, Harvard University  
GEORGE ROSENFELD, Department of the Interior/USGS  
TERRANCE SMITH, UCSB/Carnegie Mellon  
JOHN ZUMBRUNN, Commodities Corporation

FUNDAMENTAL RESEARCH PROGRAM  
INFORMATION UTILIZATION AND EVALUATION

PRESENTATION

NASA HEADQUARTERS

MAY 1, 1981

BY

JOHN ESTES

CHARLES VARS

FUNDAMENTAL RESEARCH PROGRAM  
INFORMATION UTILIZATION AND EVALUATION

PRESENTATION

NASA HEADQUARTERS

MAY 1, 1981

BY

JOHN ESTES

CHARLES VARS

## STEERING COMMITTEE

GLEN BACON, IBM Corporation

NEVIN BRYANT, NASA Jet Propulsion Laboratory

CHRISTOPHER CLAYTON, University of California

LUDWIG EISGRUBER, Oregon State University

JOHN ESTES, University of California

FORREST HALL, NASA Johnson Space Center

IDA HOOS, Space Science Laboratory, University of California

ROBERT MACDONALD, NASA Johnson Space Center

BRUCE SCHEER, The Planning Economic Group, Boston

RONALD SHELTON, Michigan State University

CHARLES VARS, Oregon State University

## BREAKDOWN OF TOTAL PARTICIPANTS BY EMPLOYMENT

ORIGINAL PAGE IS  
OF POOR QUALITY

	PL1	PL2	AW	KC	SJW	PL3	HW	OTHER	TOTALS
PRIVATE INDUSTRY		2	2	3	7			2	16
FEDERAL GOVERNMENT	2	3	3	2	7	2	5	1	25
STATE AND LOCAL GOVERNMENT			3	1					4
UNIVERSITY	8	5	8	4	3	4	7	4	43
TOTALS	10	10	16	10	17	6	12	7	88

## BREAKDOWN OF INDIVIDUAL PARTICIPANTS BY EMPLOYMENT

PRIVATE INDUSTRY	14
FEDERAL GOVERNMENT	16
STATE AND LOCAL GOVERNMENT	4
UNIVERSITY	22
TOTAL INDIVIDUALS PARTICIPATING	56



## AREAS OF EMPHASIS

- AS SHOWN ON HEADQUARTERS VIEWGRAPH

Data Base Structures

Identification of Data/Information of Significant  
Economic/Social Value

Alternative Methods of Using Remote Sensing  
and Performance Evaluation

- AS DEFINED BY INFORMATION UTILIZATION AND EVALUATION  
STEERING COMMITTEE

Information and Decision Processes

Data and Information Attributes

Data Base Management and Use

Data and Information Systems Performance

## AREAS OF EMPHASIS (con,t)

- INFORMATION AND DECISION PROCESSES

- Current and Probable Future Institutional Arrangements

- Informational Components of Decision Systems

- Decision Processes

- DATA AND INFORMATION ATTRIBUTES

- Develop Methods for Assessing Data Attributes

- Interaction of Data Attributes

- Effects on Future Decision Processes

## AREAS OF EMPHASIS(con't)

- DATA BASE MANAGEMENT AND USE

Input of Remote Sensor and Other Data

Storage, Retrieval and Management Issues

Interaction and Transmission of Stored Data

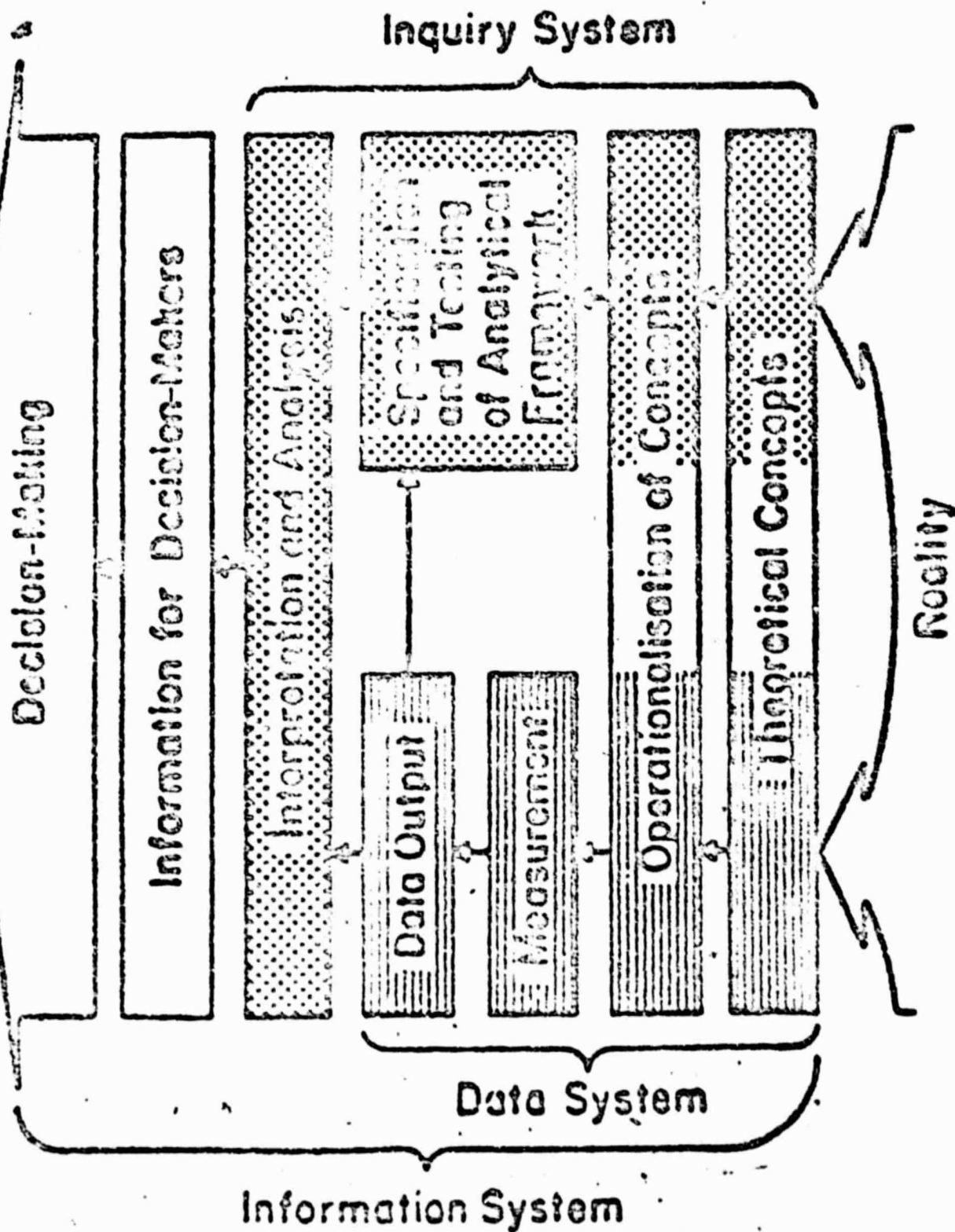
Types of Products Demanded of Future Systems

- DATA AND INFORMATION SYSTEMS PERFORMANCE

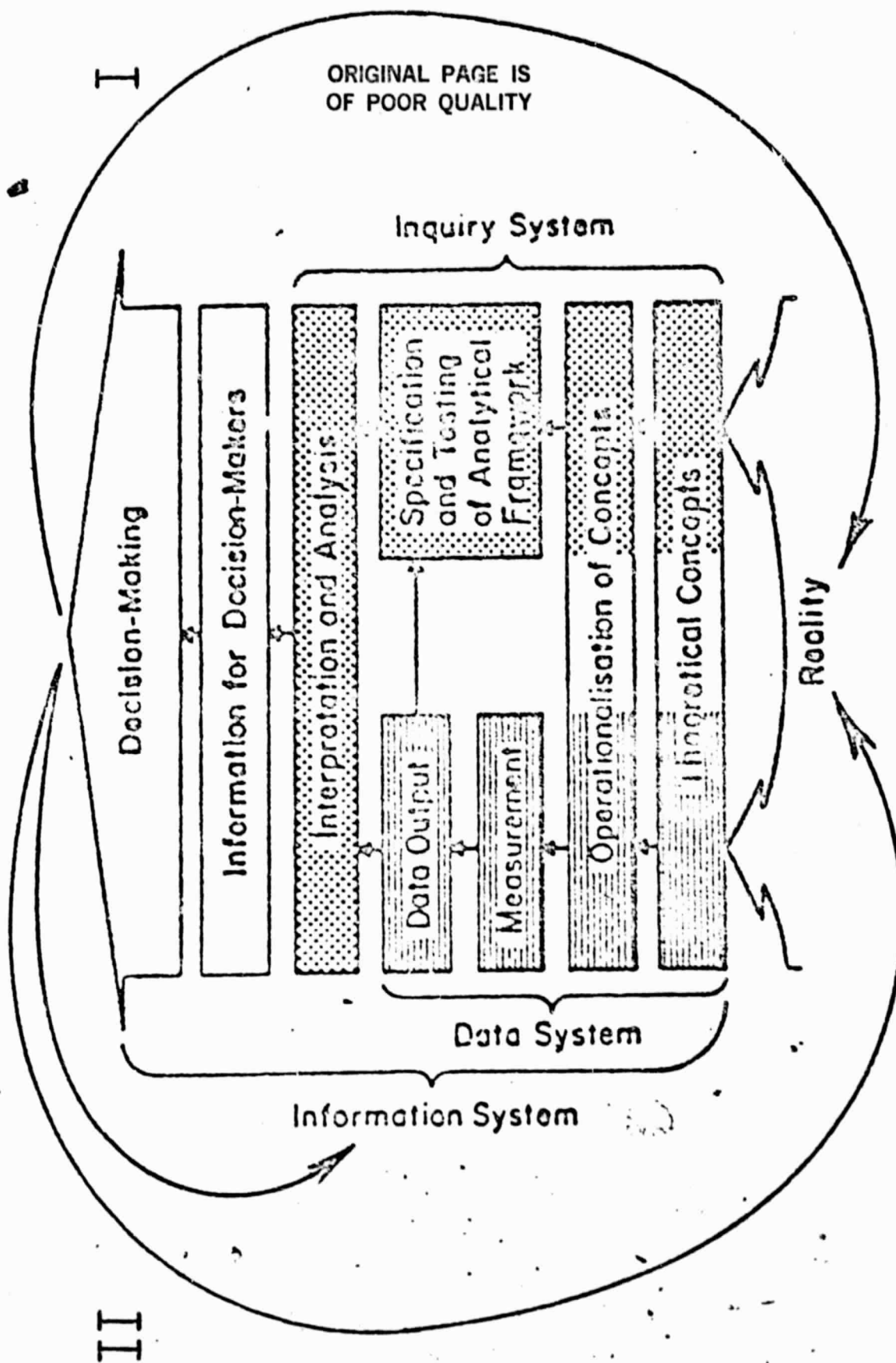
Performance Parameter Definition

Performance Parameter Assessment

Estimation of Performance Parameters in Constrained Situations



# TWO MODES OF INFORMATION USE



## MAJOR RECOMMENDATIONS FROM KANSAS CITY DISCUSSIONS

- EXISTING INFORMATION SYSTEMS: THE USE, USERS, AND RELATED PROCESSES NEED TO BE REVIEWED AND DOCUMENTED BEFORE FUTURE SYSTEMS CAN BE EVALUATED
  - Review and Documentation Should Include Systems with Other Than Obvious Remote Sensing Applications
  - Review Process Should Place Special Emphasis on Why, How, and What the Effects of Multiple Data Sources are on Operation of Existing Information Systems



## PRELIMINARY SUMMARY OF PROPOSED RESEARCH AREAS

### • INFORMATION AND DECISION PROCESSES

- Development of a General Theory of the Economics of Resource Management Information Systems
- Develop Models at Varying Levels of Aggregation of Information Systems with Public, Private and International Components
- Investigate Analytic Systems and Processing for Information Extraction from Data Bases on Disparate Concepts/Data Explore Potential Linkages Between Artificial Intelligence
- Oriented "Expert Systems", Image Analysis Logic and Data Bases for Resource Management Decision Making
- Augment Current Research in Economic Theory which Treats Information as a Commodity

## PRELIMINARY SUMMARY OF PROPOSED RESEARCH AREAS

(con't)

- DATA AND INFORMATION ATTRIBUTES

- Develop Improved Techniques for Measuring Tradeoffs Between Use of Existing and New Information Systems for Renewable Resource Decision Making
- Investigate Impacts of Timeliness, Reliability, Accuracy and Assured Product Delivery on Cost Benefit Potential
- Examine Alternative Means of Assessing the Public Good and Multiple Use Aspects of Data/Information
- Explore Use of Existing Collateral Data to Reduce Requirements for Sensor and Ground Truth Data Acquisition

## PRELIMINARY SUMMARY OF PROPOSED RESEARCH AREAS (con't)

- DATA BASE MANAGEMENT AND USE

- Factors Influencing Data Input and Interaction Potential of Remotely Sensed Data. Examine Impacts of:

- data base structure on input and archiving formats
    - improve query capabilities processing rates on
    - systems development and use

- data availability, archiving and the opportunity costs of data storage

- decentralized data bases/information systems on cost/benefits ratios

- Improvement Of Processing Strategies For Integration Of Disparate Data Types. Key Issues Include:

- understanding generic functions in spatially oriented data manipulation

- effects of positional accuracy on estimates derived from multiple data planes

- investigate performance and capacity requirements for the large record size and special purpose processing required for imagery and geographic applications

## PRELIMINARY SUMMARY OF PROPOSED RESEARCH AREAS

(con't)

- DATA BASE MANAGEMENT AND USE (con't)

- Examination of Possible and Probable Future Environments  
which will Impact Renewable Resource Data Base  
Management and Use

## PRELIMINARY SUMMARY OF PROPOSED RESEARCH AREAS

(con't)

### • DATA AND INFORMATION SYSTEMS PERFORMANCE

- Examine Fundamental Aspects of the Accuracy of Products of Remote Sensing
- Develop Relevant Measures which Characterize Data/Information (e.g., Minimally Sufficient Statistic)
- Develop Procedures Necessary for Renewable Resources Satellite Sensor Systems to Achieve "Mensuration Systems" Status for Given Applications
- Continued Examination of Fundamental Aspects of Sampling Theory as They Apply to Spatial Distributed, Temporally Varying, Renewable Resources Parameters; at Local, Regional, National, and International Scales. Key Issues Include:
  - greater sensitivity to deviations from "normal conditions"
  - identification of key parameters and determination of correlation coefficients between remote sensor scene derived data and ground conditions for use when no independent source of verification exists
  - explore the potential of nonparametric test of data

BACKUP VIEWGRAPHS



PLANNING SESSION #1  
SANTA BARBARA, CALIFORNIA  
APRIL 1980

HUGH CALKINS, State University of New York, Buffalo

CHRISTOPHER CLAYTON, University of California

GEOFFREY DUTTON, Harvard University  
Computer Graphics Laboratory

LUDWIG EISGRUBER, Oregon State University

JOHN ESTES, University of California

ROBERT MACDONALD, NASA Johnson Space Center

DENNISON PARKER, United States Department of the Interior

RONALD SHELTON, Michigan State University

DIANA WATTS, Space Science Laboratory,  
University of California

CHARLES VARS, Oregon State University

PLANNING SESSION #2  
CLEAR LAKE, TEXAS  
JULY 1980

GLEN BACON, IBM Corporation  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
CHRISTOPHER CLAYTON, University of California  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California  
FORREST HALL, NASA Johnson Space Center  
IDA HOOS, Space Sciences Laboratory, University of California  
ROBERT MACDONALD, NASA Johnson Space Center  
BRUCE SCHEER, The Planning Economic Group, Boston  
CHARLES VARS, Oregon State University

INFORMATION AND DECISION PROCESSES WORKSHOP  
ASILOMAR, CALIFORNIA  
SEPTEMBER 1980

GLEN BACON, IBM Corporation  
ROBINSON BARKER, St. Regis Paper Company  
JAMES T. BONNEN, Michigan State University  
NEVIN BRYANT, NASA Jet Propulsion Laboratory  
CHRISTOPHER CLAYTON, University of California  
LUDWIG EISGRUBER, Oregon State University  
JOHN ESTES, University of California  
OSWALD GANLEY, Harvard University  
ROBERT MACDONALD, NASA Johnson Space Center  
MICHAEL J. MCCORMICK, Planning and Community Affairs,  
State of Washington  
C. BART MCGUIRE, University of California  
ROBERT POTTER, Water Resources, Sacramento, California  
LEONARD SLOSKY, Science and Technology, Colorado  
CHARLES VARS, Oregon State University  
DIANA WATTS, Space Science Laboratory ,  
University of California  
DARRELL WILLIAMS, NASA Goddard Space Flight Center

INFORMATION AND DECISION PROCESSES WORKSHOP  
ASILOMAR, CALIFORNIA  
SEPTEMBER 1980

THEORY OF INFORMATION SYSTEMS: PUBLIC AND PRIVATE  
DIMENSIONS -

C. Bart McGuire

IMPLICATIONS OF GLOBAL REMOTE SENSING SYSTEMS:

A DISCUSSION -

Ludwig Eisgruber, leader

INFORMATION SYSTEMS USED IN PUBLIC AND  
PRIVATE DECISION MAKING -

Bob Barker, Bob Potter, Leonard Sloski

ARE THERE COMMON RESEARCH NEEDS?: A DISCUSSION -

Charles Vars, leader

CRITICAL ANALYSIS OF EXISTING INFORMATION SYSTEMS -

James Bonnen

CULTURAL, LEGAL AND POLITICAL IMPLICATIONS OF GLOBAL  
INFORMATION SYSTEMS -

Oswald Ganley

CRITICAL ISSUES RESEARCH: A DISCUSSION -

Ludwig Eisgruber, leader

REMOTE SENSING FOR RESOURCE MANAGEMENT  
SPECIAL SESSION MEETING  
KANSAS CITY, KANSAS  
OCTOBER 1980

WILLIAM ANDERSON, Technicolor Graphics

MARION BAUMGARDNER, Laboratory for Applications  
of Remote Sensing

NEVIN BRYANT, NASA Jet Propulsion Laboratory

JOSEF CIHLAR, Canada Center for Remote Sensing

JOHN ESTES, University of California

PEGGY HARWOOD, National Governors Conference

WILLIAM MACFARLAND, University of Missouri

EARL MERRITT, Earth Satellite Corporation

GENE THOMPSON, Missouri Farm Association

CHARLES VARS, Oregon State University

DATA BASE USE AND MANAGEMENT WORKSHOP  
SAN JOSE, CALIFORNIA  
JANUARY 1981

DELMAR ANDERSON, NASA Jet Propulsion Laboratory

GLEN BACON, IBM Corporation

RALPH BERNSTEIN, IBM Palo Alto Scientific Center

NEVIN BRYANT, NASA Jet Propulsion Laboratory

JOHN ESTES, University of California

STAN HANSEN, Boeing, Seattle, Washington

ADRIAN HOOKE, NASA Jet Propulsion Laboratory

IDA HOOS, Space Science Laboratory,  
University of California

RAY LORRIE, IBM Research

ROBERT MACDONALD, NASA Johnson Space Center

BOB MYERS, IBM Yorktown Research Center

GENE RICE, NASA Johnson Space Center

EVE SCHWARTZ, Fleet Numerical U.S. Navy

WILLIAM SHARPLEY, ESL/TRW

DAVID SINNOTT, NASA Research Center

DON WALKETT, Terra-Mar



DATA BASE USE AND MANAGEMENT WORKSHOP  
SAN JOSE, CALIFORNIA  
JANUARY 1981

SPACE APPLICATIONS DATA SYSTEMS PROGRAM.  
Nevin Bryant for Peter Bracken

PLANETARY SCIENCE DATA ACQUISITION -  
Adrian Hooke

REQUIREMENTS AND CHARACTERISTICS OF SCIENTIFIC DATA BASES -  
Stan Hansen

CARTOGRAPHIC DATA BASES -  
Delmar Anderson

COMMITTEE ON DATA MANAGEMENT AND COMPUTATION  
REPORT REVIEW -  
Ralph Bernstein

DATA BASE ISSUES IN GEOGRAPHIC APPLICATIONS -  
Ray Lorrie

PROBLEMS OF HIGH RATE COMPUTATION -  
William Sharpley

DATA BASE CONSIDERATIONS FOR NAVY WEATHER FORECASTING -  
Eve Schwartz

PLANNING SESSION #3  
HOUSTON, TEXAS  
FEBRUARY 1981

NEVIN BRYANT, NASA Jet Propulsion Laboratory

LUDWIG EISGRUBER, Oregon State University

JOHN ESTES, University of California

IDA HOOS, Space Science Laboratory , University of California

ROBERT MACDONALD, NASA Johnson Space Center

CHARLES VARS, Oregon State University

DATA AND INFORMATION PERFORMANCE WORKSHOP  
HOUSTON, TEXAS  
FEBRUARY 1981

FREDERIC BILLINGSLEY, NASA Jet Propulsion Laboratory

NEVIN BRYANT, NASA Jet Propulsion Laboratory

DONALD J. CLOUGH, University of Waterloo

WILLIAM COBBERLY, University of Tulsa

JOHN ESTES, University of California

CECIL HALLUM, NASA Johnson Space Center

IDA HOOS, Space Science Laboratory,  
University of California

GLENN HOUSTON, NASA Johnson Space Center

RAYMOND JESSEN, University of California

PATRICK O'DELL, University of Texas

ALAN H. STRAHLER, University of California

CHARLES VARS, Oregon State University

GAYLORD WORDEN, Department of Commerce

DATA AND INFORMATION PERFORMANCE WORKSHOP  
HOUSTON, TEXAS  
FEBRUARY 1981

PERFORMANCE PARAMETERS AND CRITERIA FOR JUDGING FEDERAL  
INFORMATION SYSTEMS -  
Gaylord Worden

DATA AND INFORMATION FROM THE SRS SYSTEM -  
Raymond Jessen

ASSESSMENT OF THE STATISTICAL ACCURACY OF ESTIMATES MADE  
FROM REMOTELY SENSED DATA -  
Patrick Odell

MAINTAINING THE SPATIAL COMPONENT IN STATISTICAL  
AGGREGATION -  
Alan Strahler for Reginald Colledge and Terrence Smith

FEDERAL-PROVINCIAL AND INTERNATIONAL DIMENSIONS OF  
CANADIAN REMOTE SENSING PROBLEMS -  
Donald Clough

MODELLING THE EFFECTS OF MISREGISTRATION ON MULTI-SPECTRAL  
CLASSIFICATION -  
Fred Billingsley

THE DATA REPRESENTATION ACCURACY QUESTION IN REMOTE  
SENSING -  
Bill Cobberly

## INDIVIDUAL CONTACTS OUTSIDE OF MEETINGS AND WORKSHOPS

CARL BRONTHAVER, Sparks Commodities

REGINALD GOLLEDGE, Univ. of Calif., Santa Barbara

ALAN MURPHY, Oregon State University

ALAN OETTINGER, Harvard University

GEORGE ROSENFELD, Department of the Interior / USGS

TERRANCE SMITH, UCSB / Carnegie Mellon

JOHN ZUMBRUNN, Commodities Corporation

## ASILOMAR WORKSHOP CONCLUSIONS CONCERNING RESEARCH AREAS

- NO GENERAL THEORY OR METHODOLOGICAL FRAMEWORK EXISTS FOR ESTIMATING THE VALUE OF INFORMATION SYSTEMS
- FACTORS AFFECTING GENERAL THEORY DEVELOPMENT INCLUDE:
  - Lack of a Real Market Value for the Output of Most Public Informations Systems
  - Quality is Based on Multiple Characteristics
  - Some Information is a Public Good, and Private Value may be Different from Social Value



## ASILOMAR WORKSHOP CONCLUSIONS CONCERNING RESEARCH AREAS (con't)

- LIMITED HIGH QUALITY RESEARCH HAS BEEN CONDUCTED IN EACH AREA
- ONLY RUDIMENTARY MODELS HAVE BEEN DEVELOPED, TESTED, AND EMPIRICALLY ESTIMATED
- SEVERAL APPROACHES HAVE BEEN EMPLOYED BUT CONTROVERSY CONTINUES CONCERNING OPTIMUM METHODOLOGIES
- NO METHODOLOGY HAS PROVEN SUCCESSFUL IN GENERAL APPLICATION TO PUBLIC RESOURCE DATA SYSTEMS
- DIFFERENT APPLICATIONS CONCERNING APPROPRIATE METHODOLOGIES RAISE MORE QUESTIONS THAN THEY ANSWER
- MANY SOURCES OF DATA ARE AVAILABLE TO DECISION MAKERS FEW REGARDED AS CERTAIN, THEREFORE CHOICES AMONG DATA TYPES MUST OFTEN BE MADE WITHIN UNCERTAIN ENVIRONMENTS

## ASILOMAR WORKSHOP CONCLUSIONS CONCERNING RESEARCH AREAS (con't)

- DEVELOP A GENERAL THEORY OF THE ECONOMICS OF  
RESOURCE MANAGEMENT INFORMATION SYSTEMS AND THEIR  
MODIFICATION
- DEVELOP RELEVANT MEASURES OF DATA AND INFORMATION  
AND EMPIRICAL METHODS FOR INVESTIGATING INFORMATION  
SYSTEM PERFORMANCE
- DEVELOP INFORMATION SYSTEM MODELS COMPRISED OF PUBLIC,  
PRIVATE AND INTERNATIONAL COMPONENTS
- INVESTIGATE ANALYTIC SYSTEMS AND PROCESSING STRATEGIES  
FOR EXTRACTION OF DISPARATE INFORMATION FROM DATA BASES
- ANALYZE DATA AND INFORMATION DISSEMINATION SYSTEMS  
INCLUDING STUDIES OF
  - Accessibility
  - Confidentiality and Property Rights
  - Decentralization
  - Public Good
  - Economies of Scale
  - Joint Use of Data

## SAN JOSE WORKSHOP

### KEY POINTS FROM STAN HANSEN'S PRESENTATION

#### "Requirements and Characteristics of Scientific Data Bases"

- WE ARE NOT JUST COMPUTERIZING EXISTING DATA AND PROCEDURES: WE ARE ACTUALLY CREATING NEW DATA TYPES
- MOST COMPUTER SYSTEMS ARE VERTICALLY INTEGRATED TO PERFORM SPECIALIZED TASKS. WE NEED A GLOBAL TRANSFER CAPABILITY TO FACILITATE TRANSFER THROUGH HORIZONTAL INTEGRATION

- RESEARCH IS NEEDED ON:

Decentralized Data Base Structures

Distributed Data Processing

Common Carriers

IF USERS ARE TO HAVE FLEXIBLE RESEARCH AND DESIGN ENVIRONMENTS

- USERS DO NOT HAVE REQUIREMENTS  
THEY HAVE REACTIONS TO SITUATIONS IMPOSED UPON THEM

## SAN JOSE WORKSHOP SUMMARY AND CONCLUSIONS

- DATA BASE MANAGEMENT AND USE IS DEVELOPING COMMESURATE WITH OVERALL EXPANSION OF CAPABILITIES IN THE COMPUTER AND COMMUNICATIONS FIELDS
- THIS DOES NOT MEAN THE FIELD IS ACTIVELY SOLVING PROBLEMS CONFRONTED BY REMOTE SENSING TECHNOLOGY IN THE RENEWABLE RESOURCES DISCIPLINE AREA
- CASE STUDIES ARE A PRACTICAL MEANS FOR DETERMINING ESSENTIAL COMPONENTS OF DATA BASE USE AND MANAGEMENT FOR EARTH RESOURCE APPLICATIONS
- NASA MUST MAKE RENEWABLE RESOURCES NEEDS UNDERSTOOD TO THE COMPUTER SCIENCE COMMUNITY

## SAN JOSE WORKSHOP SUMMARY AND CONCLUSIONS

(con't)

- COMMERCIAL MARKET HAS MOVED RAPIDLY IN THE AREA OF DATA BASE AND QUERY AND WILL CONTINUE TO EXPAND THE SPEED AND EFFICIENCY OF CATALOGUE AND QUERY CAPABILITIES
- NASA DATA ADMINISTRATION NEEDS SHOULD BE MADE KNOWN TO STANDARDS BOARDS AND COMMITTEES

## SAN JOSE WORKSHOP PRELIMINARY RESEARCH AREAS

- IMPROVED METHODS FOR LOCATING AND ACCESSING LARGE VOLUMES AND LARGE RECORDS OF DATA WITHIN EXISTING DATA BASES
- EFFICIENT METHODS FOR COMBINING DIVERSE DATA SETS
- SYSTEMS LEVEL COMMUNICATION (e.g., TRANSPORTABLE LANGUAGES/UNIVERSAL TRANSLATORS)
- INTEGRATION OF HIGH DATA RATE SYSTEMS AND DATA BASES
- MANAGEMENT OF GEOREFERENCED DATA BASES
- STANDARDIZATION OF DATA. IN PARTICULAR DEVELOPMENT OF A COMPREHENSIVE HEIRARCHY OF LAYERED STANDS PERMITTING GLOBAL INTERCOMMUNICATION OF PRODUCTS WITHIN AN OPEN SYSTEM



## SAN JOSE WORKSHOP PRELIMINARY RESEARCH AREAS (con't)

- ARCHIVING AND PURGING OF DATA BASES
- DATA AGGREGATION
- DATA VALIDATION
- INTERACTIVE MAN-MACHINE ENVIRONMENTS AND THE ASTHETICS OF DISPLAY
- EXAMINE USER REQUIREMENTS AND CONSTRAINTS WITHIN THE CONTEXT OF CURRENT SYSTEMS ENVIRONMENTS TO PROJECT FUTURE NEEDS AND USES

## SAN JOSE WORKSHOP PRIMARY RESEARCH AREAS

- PERFORMANCE AND CAPACITY IMPROVEMENT

- Ability to Handle Longer Records and Special Purpose Processing Associated with Imagery and Geographic Applications
- Basic Structures of Data Base Management Systems

- DATA ADMINISTRATION

- Managment Strategies and Data Ownership

- DATA BASE AND QUERY

- Improved Cataloging Capabilities
- Improved Query Capabilities (e.g., Employing Hueristic Logic)

- SYSTEMS STUDIES

- Development of Models for Base Generation and Use
- Problems with Data Handling and Trend Analyses (e.g., Logic)

## PRELIMINARY SUMMARY AND CONCLUSIONS DRAWN FROM HOUSTON WORKSHOP

- QUALITY AND POLICY RELEVANCE OF FEDERAL LEVEL STATISTICS VARY WIDELY
- FAILURE TO ANTICIPATE DATA NEEDS HAS LED TO POLITICAL PRESSURE ON STATISTICAL AGENCIES
- PRIVACY PROTECTION AND BURDEN MINIMIZATION ARE PROBLEMS OF INCREASING SIGNIFICANCE TODAY
- OFFICE OF FEDERAL STATISTICAL STANDARDS AND POLICY (OFFSS&P) CRITERIA ARE APPROPRIATE TO NATURAL RESOURCES STATISTICAL SYSTEMS, BUT NEITHER CONCEPTS NOR INSTITUTIONS ARE SUFFICIENTLY DEVELOPED TO PERMIT TRADEOFF AND COST BENEFIT ANALYSIS COMPARABLE TO THOSE UNDERTAKEN FOR ECONOMIC AND OTHER FEDERAL STATISTICS
- REMOTE SENSING MAY BE A USEFUL DATA SOURCE, BUT ANY CHANGE IN EXISTING DATA DATA/INFORMATION GENERATION SYSTEMS WILL ONLY OCCUR IF THE SOURCES AND MAGNITUDES OF THE ERRORS ASSOCIATED WITH THE NEW PROCEDURES ARE FULLY CALIBRATED
- REMOTE SENSING CAN SERVE IN SOME CASES TO MAKE LOCAL STATISTICS MORE ACCURATE
- FOR MANY RENEWABLE RESOURCES CLASSICAL STATISTICAL TECHNIQUES ARE "NEARLY APPLICABLE". "NEARLY APPLICABLE", AS MOST ASSUMPTIONS UPON WHICH CLASSICAL TECHNIQUES ARE BASED, ARE VIOLATED IN APPLICATION

PRELIMINARY SUMMARY AND CONCLUSIONS FROM  
HOUSTON WORKSHOP (con't)

- WE MUST CONTINUE WORK ON DEVELOPMENT OF ROBUST METHODS FOR STATISTICAL ASSESSMENT OF RENEWABLE RESOURCES APPLICATIONS
- CASE STUDIES EMPLOYING THOROUGHLY VERIFIED TEST DATA SETS WOULD BE OF GREAT VALUE IN NEW SAMPLING PROCEDURE DEVELOPMENT
- WE MUST UNDERSTAND AGENCYS' OBJECTIVES OR THERE CAN BE NO BONA FIDE ANALYSIS
- IN MANY CASES IT MAY BE PREMATURE TO ASSESS BENEFITS AS USERS ARE SCIENTISTS ASSOCIATED WITH GOVERNMENTS

## PRELIMINARY RESEARCH AREAS FROM HOUSTON WORKSHOP

- DEVELOP CONCEPTS TO FACILITATE TRADEOFF AND COST BENEFIT ANALYSES OF ECONOMIC AND OTHER FEDERAL STATISTICAL SYSTEMS. THESE STUDIES SHOULD ADDRESS QUESTIONS OF:
  - Privacy
  - Accuracy Verification
  - Costs
- ANALYSIS ORIENTED TOWARDS MAKING SATELLITE SENSORS "MATURE MEASUREMENT SYSTEMS" (AFTER ROSENFELD)
- USE OF REMOTE SENSING TO IMPROVE THE ACCURACY OF LOCAL STATISTICS
- NEW APPROACHES TO SAMPLING MORE RESPONSIVE TO VARIATIONS FROM NORMAL CONDITIONS
- NEED FOR BASIC UNDERSTANDING OF RESOLUTIONS NECESSARY FOR THE PRODUCTION OF MINIMALLY SUFFICIENT STATISTICS FOR GIVEN APPLICATIONS

PRELIMINARY RESEARCH AREAS FROM HOUSTON WORKSHOP  
(con't)

- NEED FOR CASE STUDIES TO IMPROVE OUR UNDERSTANDING OF TODAY'S STATE-OF-THE-ART-ESTIMATION SYSTEMS. THIS APPROACH MAY REVEAL GAPS IN OUR KNOWLEDGE
- NEED FOR CONTINUED FUNDAMENTAL EXAMINATION OF SAMPLING THEORY AS APPLIED TO RENEWABLE RESOURCES TYPE INFORMATION PROBLEMS



## RESEARCH AREAS AS DEFINED BY DISCUSSIONS WITH OTHER INDIVIDUALS

- WORK ON THE FUNDAMENTAL ASPECTS OF ACCURACY
- METHODS FOR MAKING SATELLITE SENSORS MATURE MEASUREMENT SYSTEMS FOR GIVEN APPLICATIONS AREAS
- DEVELOPMENT OF LINKS BETWEEN RESEARCH, ARTIFICIAL INTELLIGENCE, DATA BASE, AND IMAGE ANALYSIS LOGIC AND THEIR IMPACTS ON "EXPERT SYSTEMS"