DATA MANAGEMENT FOR EARTH OBSERVATIONS

by

George F. McDonough
Deputy Director, Data Systems Laboratory

George C. Marshall Space Flight Center
Marshall Space Flight Center, Alabama 35812

The Earth Observations efforts of NASA and other Federal agencies have matured into near-operational programs which require large amounts of data on a time-constrained basis. To meet these needs considerable efforts have been undertaken by the agencies as well as by users of the data. Potential problems are discussed along with an approach to reducing costs and providing needed information. User participation in the process of data management is one key to its success. Some available means for the user to become involved are considered.

With increasing maturity of the acquisition systems for remote sensing, especially the weather and Earth resources satellites such as Nimbus and Landsat, emphasis has shifted toward the application of the data from these satellites. For example, the extensive use of Landsat data for a wide variety of applications has established the utility of this kind of remotely sensed data (Fig. 1). As a consequence, emphasis is shifting beyond initial utilization and approaching the operational use of remote sensing data. By operational use is meant the day-to-day use of remotely sensed data as a part of the process of decision making regarding natural resources, for example. As the applications of such data increase, so do the requirements for management of the data, which is the subject of this presentation.

- BASIC REMOTE SENSING SCIENCE IS MATURE
  - SIGNATURES
  - APPLICATIONS
- EMPHASIS SHIFTING TO OPERATIONAL SYSTEMS
  - AUTOMATED PROCESSES
  - MULTIPLE SYSTEM INTERACTIONS
  - TIMELINESS
  - COST
- UTILIZATION BY PUBLIC BROADENING
  - WEATHER
  - MAPPING

Figure 1. Evaluation of data situation.

In each of the areas presented in Figure 2, there is substantial effort underway to establish the required systems, technology, and analysis techniques to meet the data management needs of the 1980's and beyond.

- AGRICULTURE
  - INVENTORY
  - CROP FORECASTING
- WEATHER
  - GLOBAL MARINE
  - SEVERE STORMS
- WATER AVAILABILITY
- POLLUTION MONITORING
- LANDUSE

Figure 2. Some areas of major data system emphasis.

Taking Global Marine Weather as an example, a substantial number of users can be identified. Some of the more important users are listed in Figure 3. The Navy and Coast Guard are served by DoD weather programs, fishing fleets, oil companies (off-shore rigs), etc., are served by the weather programs included under the "Local and Large Scale Weather" programs. Commercial shipping companies, which range worldwide, currently have potential requirements beyond the goals of these programs.

• U.S. NAVY • U.S. COAST GUARD	DoD WEATHER PROGRAMS
• FISHING FLEETS • OIL COMPANIES • COASTAL CITIES • RECREATION ACTIVITIES • WEATHER RESEARCHERS • COMMERCIAL SHIPPING COMPANIES	LOCAL AND LARGE SCALE WEATHER

Figure 3. Users of marine weather information.

Figure 4 shows requirements for a Global Marine Weather forecasting system which measures temperatures, winds, etc., by means of instrumented buoys, stations, ships, and satellites. The data are inputs to an analysis and prediction model which produce 1, 3, and 15 day forecasts. A 1 day forecast is produced in 12 h from acquisition, a 3 day forecast in 24 h, and a 15 day forecast in 3 days.

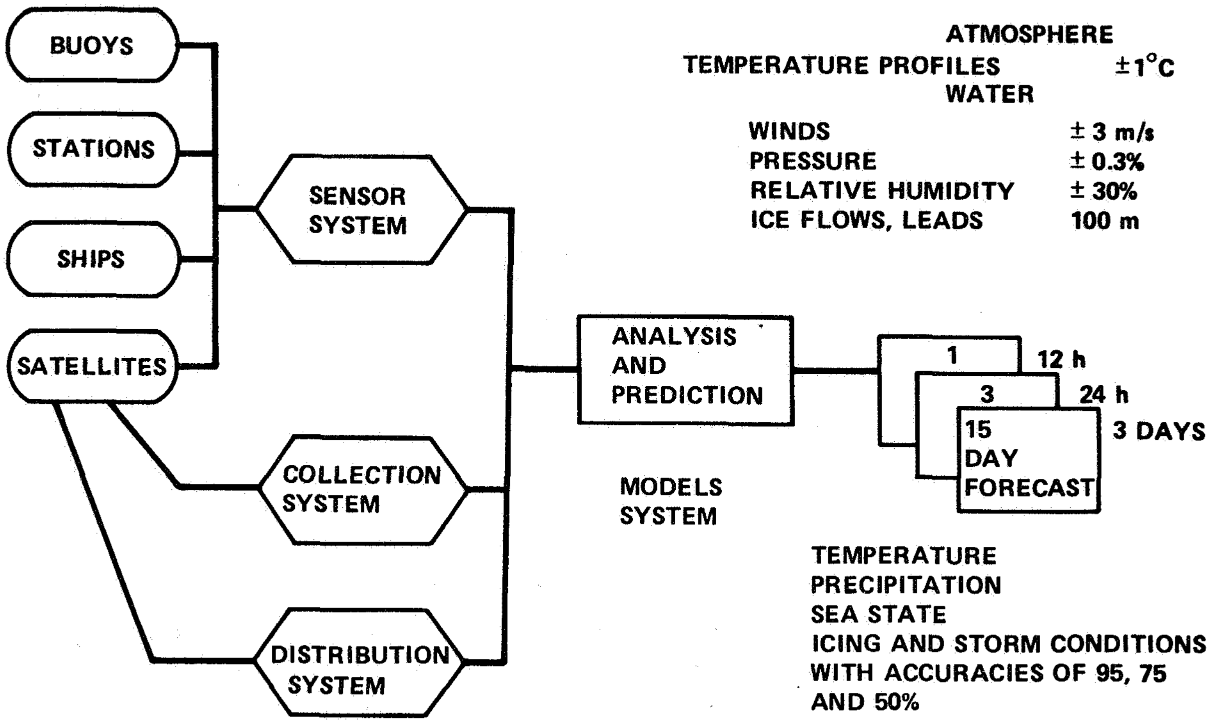

Figure 4. Functional system requirement.

The flow of data for the Navy system is shown in Figure 5. Satellites such as Nimbus, which service the National Weather Service (NOAA) system supplement the military acquisition systems. Seasat A, currently under development by NASA, also will feed data to the system in the 1980's. Fleet Numerical Weather Central possesses the models to transform these data into forecasts which are fed back to NOAA and also, through Fleet Weather Centers (FWC) to ships at sea. Information also is supplied to commercial forecasting companies who, in turn, supply it to commercial shipping and other users.

Data management often is perceived as a problem both by originators and/or suppliers and by users, largely because of the potential for high costs, delays, and other detrimental results which can occur if data management is not properly used (Fig. 6). The Global Marine Weather situation shows the complexity of interfaces as well as the use of multiple data sources. Many of the sensors used are capable of acquiring data at high rates (megabits per second) which could lead to high costs.

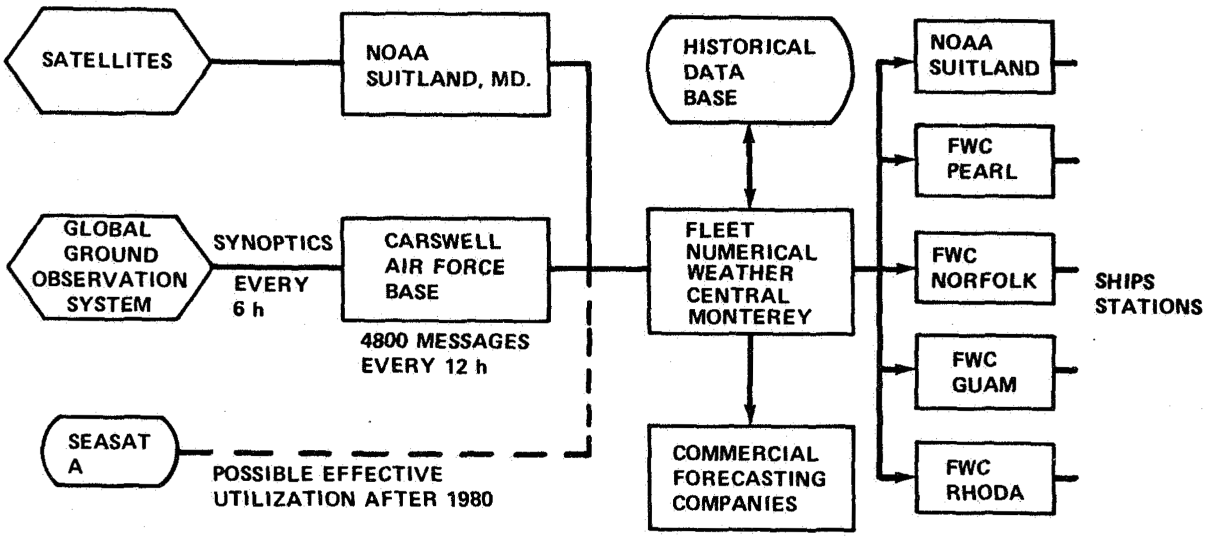

Figure 5. Data flow for the Navy system.

WHY IS REMOTE SENSING DATA A PROBLEM TO NASA?

- HIGH DATA RATES
- MULTIPLE DATA SOURCES
  - MULTIPLE SPACECRAFT
  - AIRCRAFT
  - GROUND DATA COLLECTION
  - ANCILLARY DATA
- PROCESSING, HANDLING, AND STORAGE COSTS
- COMPLEX INTERFACES WITH USERS

Figure 6. The data management problem.

To establish what the possible data acquisition rates mean, Figure 7 shows the meaning of 10^{15} bits/year (a bit is the basic element of input to a computer which represents essentially the basic, or smallest, element of data). This amount of data can be acquired by the sensors under planning for the 1980's. Roughly it equals 10 times all the data acquired for whatever purpose in the year 1972. If NASA were to process these data using our Skylab system, which was a substantial one, it would take 100 years.

VOLUME:
10^{15} BITS/YEAR
100 LIBRARIES OF CONGRESS
100 MILLION SEARS CATALOGS
30 ERTS-1
10 DATA TOTALS FOR 1972
PROCESSING TIME:
100 YEARS AT SKYLAB RATES

Figure 7. Significance of estimated yearly data output for the 1980's.

This problem can be solved either by vastly increasing our capacity to process and handle data or by reducing acquisition to match information needs (Fig. 8). Clearly, reducing acquisition is preferable, and using this approach does not prevent improving or adding capabilities where required. It does present the problem of understanding the use to which the data are being put well enough to judge what data are required. This involves intimate association with the user and his application.

TWO APPROACHES SHOW PROMISE FOR SOLUTION:

1. INCREASE THROUGHPUT TO MATCH SENSOR CAPABILITIES
   - OPTIMIZE UTILIZATION OF FACILITIES
   - ADD FACILITIES
   - ADVANCE STATE-OF-THE-ART: IMPROVE FACILITIES
2. REDUCE ACQUISITION TO MATCH INFORMATION REQUIREMENTS

ADVANTAGES OF REDUCED ACQUISITION

- MOST EFFECTIVE WAY TO REDUCE COSTS
- IMPROVES EFFICIENCY OF INFORMATION APPLICATION

Figure 8. Approaches to problem solution.

The user also has problems with remote sensed data since some of the handling and processing is his responsibility and since the ultimate application is his job (Fig. 9). If data can be put into formats similar to those of the products with which he normally works, his problems can be diminished appreciably. He must be informed as to availability and location of data before he can use the data, and he must understand automated retrieval systems if that is his only access to the data. This requires extensive training and familiarization.

WHY IS REMOTE SENSING DATA A PROBLEM TO THE USER?
• UNFAMILIAR FORMATS
• CHANGES IN ESTABLISHED PROCEDURES
• LACK OF INFORMATION ON WHAT IS AVAILABLE AND WHERE
• DATA INVOLVES HIGH TECHNOLOGY
• TRAINING/FAMILIARIZATION REQUIRED
• COSTS OF PROCESSING

Figure 9. The data management problem.

The NASA approach to data management includes, in addition to technology development and selectivity, concepts of end-to-end optimization of data systems — from sensor to user, coupled with an early consideration of data problems (Fig. 10). Too often data problems are considered only after acquisition systems and other major elements of the data system have been defined. This limits the gains to be made from efficient data management.

• EARLY CONSIDERATION OF DATA PROBLEMS
• END-TO-END OPTIMIZATION
• COMPRESSION
• TECHNOLOGY DEVELOPMENT
• SELECTIVITY
– HIGH INFORMATION/DATA RATIO
– CAREFUL TAILORING TO USER NEEDS
– UTILIZATION OF USER KNOWLEDGE

Figure 10. An approach to data management.

Some potential reductions in data processing requirements are given in Table 1. The reductions are based on knowing enough about the application to judge what is required to solve the problem. For example, Landsat acquires data over the test site every 18 days or approximately 20 times per year, but the application, in this case landcover analysis, only requires two images per year. Similarly two spectral channels are sufficient, although four are available and the test site covers only a small portion of the 100 × 100 n.mi. image. Finally, data compression techniques can further reduce the throughput until total reduction is from 9000:1 to 60 000:1. Even greater decreases are feasible, as will be seen, so that it will be possible to handle all the required data in an affordable way.

TABLE 1. SOME POTENTIAL DATA STREAM REDUCTIONS (LANDSAT EXAMPLE FOR A QUADRANGLE TEST SITE)

Acquisition Information	Temporal (images/ year)	Spectral (channels/ image)	Spatial (pixels/ channel/ image)	Amplitudinal (bits/pixel/ channel/image)	Information Content (bits)
	20	4	7.5×10^6	6	3.6×10^9
Application Information	2-Growing and Nongrowing Season	2-Sufficient for Gross Feature Extraction	5×10^4-Test Site Size	2-Information Preserving 0.3 Non-Info Preserving	4×10^5 6×10^4
Compression Ratio	10:1	2:1	150:1	3:1 20:1	9×10^3:1 6×10^4:1
Information Extraction Method	Selectivity			Data Compression Analysis →	Answer-Ultimate Compression or Information Extraction

A further development of data compression, based on work at the Jet Propulsion Laboratories shows promise of reductions of more than 3000:1. It involves the preprocessing of Landsat data in a "black-box" to reduce the effort done by the user. Cost savings to the user are shown in Figure 11. Savings such as these can make remotely sensed satellite data very competitive with other types of data.

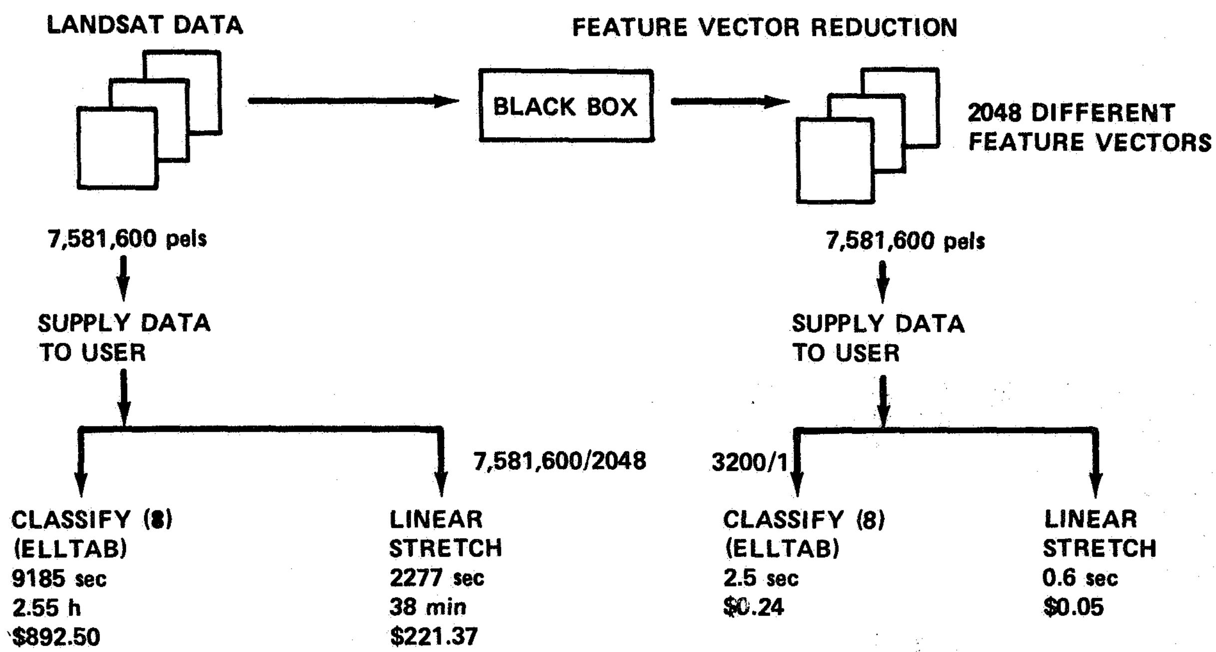

Figure 11. User cost savings.

Users have access to considerable help from NASA, as well as other agencies such as USDA who utilize remotely sensed data in their activities (Fig. 12). This help is available through the Technology Utilization Office and other channels. Universities and industry participating in NASA remote sensing and data management programs also can provide help.

One of the tools available to users is the Data Systems Dynamic Simulator at MSFC, which enables users and data systems designers to evaluate alternate methods of handling and processing data (Fig. 13). It also serves as a training and familiarization device for users who are not knowledgeable of Earth resources data systems.

- DOCUMENTS DESCRIBING
  - DATA SYSTEMS
  - AVAILABLE DATA AND APPLICATIONS
- COMPUTER – INTERACTIVE ANALYSIS CONSOLES
- SIMULATIONS
- DATA BASES
  - AVAILABLE IMAGERY
  - DATA SYSTEM TECHNOLOGY
- TRAINING

Figure 12. Some available tools to help the user.

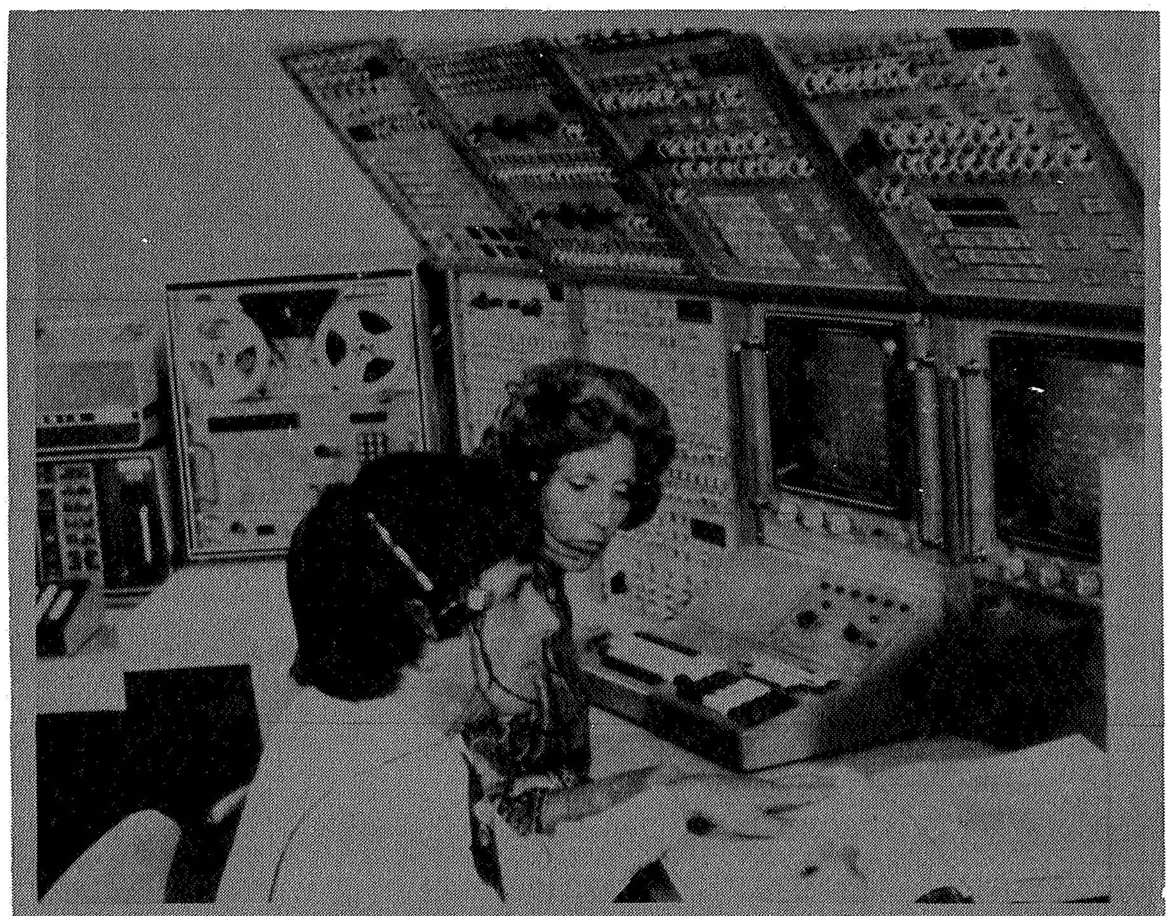

Figure 13. Data Systems Dynamic Simulator at MSFC.