/N-43
/98336
23 P

Data Management for Interdisciplinary Field Experiments:OTTER Project Support

Gary Angelici, Lidia Popovici, and J. W. Skiles

NASA

National Aeronautics and
Space Administration

Data Management for Interdisciplinary Field Experiments:OTTER Project Support

Gary Angelici and Lidia Popovici

Sterling Software, Inc.
1121 San Antonio Road
Palo Alto, CA 94303

J. W. Skiles

Johnson Controls World Services
7315 N. Atlantic Ave.
Cape Canaveral, FL 32920

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

ABSTRACT

The ability of investigators of an interdisciplinary science project to properly manage the data that are collected during the experiment is critical to the effective conduct of science. When the project becomes large, possibly including several scenes of large-format remotely sensed imagery shared by many investigators requiring several services, the data management effort can involve extensive staff and computerized data inventories. The OTTER (Oregon Transect Ecosystem Research) project was supported by the PLDS (Pilot Land Data System) with several data management services, such as data inventory, certification, and publication. After a brief description of these services, experiences in providing them are compared with earlier data management efforts and some conclusions regarding data management in support of interdisciplinary science are discussed. In addition to providing these services, a major goal of this data management capability was to adopt characteristics of a pro-active attitude, such as flexibility and responsiveness, believed to be crucial for the effective conduct of active, interdisciplinary science. These are also itemized and compared with previous data management support activities. Identifying and improving these services and characteristics can lead to the design and implementation of optimal data management support capabilities, which can result in higher quality science and data products from future interdisciplinary field experiments.

INTRODUCTION

Interdisciplinary earth science projects vary greatly in the manner in which they manage their data. Several project variables, including the number of and geographical distribution of investigators, the amount and type of data collected, and the number and difficulty of requirements, affect how the data management task is designed and implemented. The project may only require a modest effort, using only a few computer-literate graduate students, or it may require a formal system consisting of a large staff with computerized data inventories and sophisticated configuration management. A large number of geographically dispersed investigators sharing gigabytes of large-format data, including remotely sensed images gathered from numerous sensors aboard various platforms, contribute to the need for a larger data management effort. In addition, since many of these remote sensing instruments are unique in purpose, data collection technique, format, and analysis, knowledgeable staff familiar with these instruments must be available. Also, requirements stipulated by these projects, such as the ability of all project scientists and collaborators to receive these large and diverse data sets (and necessary documentation) in a timely fashion and the need to publish the data on permanent media after the project is complete, point to a larger data management effort able to satisfy all of the requirements. In general, larger projects require a fuller set of more complex data management services.

If the scientific goals of the large interdisciplinary projects are to be attained, appropriate techniques for providing these numerous and complex services must be designed and implemented, and the services must be provided in a manner which will enhance the science being performed. While each project is unique and the services must be tailored for each, many of the services provided are common among most projects. Techniques designed for one project may be applicable in a large degree to the requirements of another project. Likewise, other factors, such as the attitude of the staff and management of the data management team, can also affect the ability of the project to reach its objectives.

It is the goal of this paper to aid in the development and implementation of other data management capabilities by describing and evaluating the support of one such interdisciplinary field experiment. Beginning in late 1989, the Ames Research Center (ARC) site of the NASA PLDS (Pilot Land Data System) began the support of the OTTER (Oregon Transect Ecosystem Research) project with several data management services, such as data inventory, distribution, documentation, certification, and publication. Of the many possible services performed in support of projects, this paper focuses on the provision of three common services, data inventory, data use policy and certification, and data publication, and on four characteristics that are reflected in the attitude of the data management staff: flexibility, responsiveness, communication, and project focus. The OTTER experiences in providing the three services and applying the four characteristics are compared with experiences of previous support efforts and evaluated for effectiveness. The paper concludes with some of the longer term implications of the provision of appropriate services and the adoption of characteristics reflecting a pro-active attitude.

REVIEW OF EARLIER EFFORTS

Due perhaps to the impression that data management is a support activity that does not deserve attention apart from the projects it supports, there is little written in the professional literature about the provision of services or the attitude that is necessary for effective support of interdisciplinary field experiments. The authors are familiar with three major earth science interdisciplinary field experiments conducted recently which had data management support capabilities of various sizes and structures. In varying degrees, each of these projects experienced many of the same challenges and learned many of the same lessons learned by the OTTER project support effort. From available literature and personal communications, a review of the data management activities and lessons learned relative to the three services and four characteristics of a pro-active attitude are offered.

FIFE

The FIFE Information System (FIS) was created to serve the data management needs of the First International Land Surface Climatology Project (ISLSCP) Field Experiment (FIFE), a large project involving dozens of investigators and over 100 gigabytes of data. It is the best documented system relative to the problems involved in the configuration of data management capabilities for interdisciplinary field experiments. FIS developed "from the ground up" a data and inventory system with user interface software that were used heavily by investigators and collaborators in the period immediately following the field campaigns. FIS was also required to provide extensive geographic information, such as latitude and longitude, elevation, slope and aspect, for each FIFE site (Strebel, 1989) and this information has been associated with each data set.

An eight point data use policy was developed, including one stipulation that "FIFE investigators are expected to submit copies of updated or derived data sets which they personally distribute to other investigators in order to maintain a common level of data access and quality for all investigators" (FIS, 1989).

An elaborate scheme for the certification of data involving a two part coding scheme with the first part broken down into four categories: 1) EXM for example or test data, 2) PRE for preliminary or unchecked data, 3) CPI for checked by principal investigator and 4) CGR for checked by group and reconciled. The second part of the coding scheme is not constrained to a predetermined set with sample uses, including 1) ??? (sic) for data that may be questionable and 2) NFP for "not for publication" at the request of the investigator (FIS, 1989).

The approach taken in the publication of the 6000 ASCII data files on the FIFE prototype CD-ROM revolved around the goal to allow investigators to find all data for one day in the field campaign (Landis, 1992). User interface software was developed for MS-DOS computers to find the data of interest. One simple file naming convention, the date and time of data acquisition, was used for all image data. The approach toward the remotely sensed imagery was to re-format the data to extract the header information from the image data, which was left in a generic format with no header, to create a separate file. In addition,

separate ASCII files for latitude and longitude coordinate as well as azimuth and zenith angle data were created where applicable. Information on the data was provided in Structured Query Language (SQL) format to expedite the database ingest procedure. Software for the viewing of the image data files on MS-DOS computers was recorded. For several aspects of the CD-ROM, the philosophy was that most FIFE investigators had access to MS-DOS computers.

All four characteristics of a pro-active attitude identified as important in the support of active interdisciplinary research were demonstrated (and documented in the literature) by the FIS. The FIS exhibited flexibility in its ability to change its structures and procedures in the face of project requirements that changed frequently in response to the realities of data collection (Strebel, 1989). By reconfiguring their system in a timely fashion to remove critical bottlenecks, the FIS demonstrated responsiveness to the requirement of rapid data delivery (Strebel, 1990). FIS staff were in frequent communication with the FIFE project scientists as the system progressed through its various stages (Strebel, 1989) and participated in "multiple feedback loops", which included constant review and input from practicing scientists (Strebel, 1990). Finally, the FIS demonstrated a strong project focus by allowing the information system to be "under direct day-to-day control of scientist/users" (Strebel, 1989).

GRSFE

The data management support for the Geologic Remote Sensing Field Experiment (GRSFE) was performed mainly by scientists at the geosciences node of NASA's Planetary Data System (PDS) who worked closely with the central node of PDS on issues of standards and publication (Dale-Bannister, 1991).

Using techniques developed in the production of CD-ROMs for deep space missions, the GRSFE project produced a set of 9 CD-ROMs in cooperation with the central node of the PDS (Arvidson, 1991). Each of the image data files were accompanied by a PDS label file, which contained both data file structure information as well as general data identification information. These label files enabled the easy display of imagery on both MS-DOS and Macintosh computers using PDS-compatible software. Virtually all of the voluminous documentation for the experiment, including information on file formats and directory structure, is stored on a single text file. Another text file, ready for ingest into database systems, contains machine-readable templates describing the GRSFE project. Due to resource limitations, the image data sets were recorded on the CD-ROM in their native format instead of in separate bands per file (Guinness, 1991). Separate bands of imagery, some registered with data from other instruments, were provided for easy display and analysis. The GRSFE activities in the data inventory, data use policy and data certification areas have not been published in the open literature.

At the time of GRSFE data collection and processing, many of the data standards, which were applicable mainly to planetary data, had to be modified and expanded to handle earth science data (Dale-Bannister, 1993). The Planetary Data System was able to change its data dictionary to accommodate the differences inherent in earth science data, such as the remotely sensed imagery coming from earth-based multispectral instruments. The project

focus was achieved by the GRSFE project providing scientific liaison persons at the geosciences node to work with the central node of PDS. These persons provided guidance in the enhancement of standards and in the preparation of data sets for publication. Most of the responsiveness and communication were provided by staff at the geosciences node, including the PDS liaison in charge of managing the data as they were collected.

FEDMAC

The FEDMAC (Forest Ecosystem Dynamics Multisensor Aircraft Campaign) project was the smallest of the three projects, involving one data management person responsible for receiving data from investigators, processing data, and distributing data to a small group of investigators (Kim, 1993).

Because the effort was relatively small, there was no on-line investigator-accessible data inventory, as with the FIFE project. The inventory was maintained by the data management person who could be contacted for information about collected data.

Spectral data collected during the FEDMAC project were prepared for publication on high density floppy diskettes. The data were re-formatted to a common ASCII format, compressed and written to the diskettes with investigator-provided documentation. No CD-ROM production was anticipated.

Although the project was small, flexibility was demonstrated in the ability of the person to accept spectral data from investigators in a variety of ASCII formats (even though a specific format is requested) and change the data to a consistent format for distribution and publication. Communication with investigators was achieved mainly through the FEDMAC project manager. The personal involvement of the project manager was sufficient to provide a project focus revolving closely around the scientists' needs. A description of the FEDMAC data management and processing activities is offered in another paper in this issue of Remote Sensing of Environment.

OTTER DATA MANAGEMENT

Project and Data Profile

The OTTER project had the principal objective of estimating major fluxes of carbon, nitrogen, and water through forest ecosystems using remotely sensed image data (Peterson and Waring, submitted). More than 20 scientists from over 10 research institutions across the United States and in Canada were participating in the testing and validation of the predicted fluxes and the biological regulation of these fluxes as simulated by ecosystem processes models. Data were collected at six separate sites along an elevational and climatic gradient in west central Oregon during the spring, summer and fall of 1990 and in the spring of 1991.

The bulk of the data were remotely sensed imagery from instruments flown on satellites and high-altitude, medium-level, light, and ultralight aircraft. Satellite images for the project include registered composite AVHRR (Advanced Very High Resolution Radiometer) data generated by the EROS Data Center. NASA ER-2, C-130, and DC-8 aircraft, as well as light and ultralight aircraft, participated in the acquisition of data over the Oregon transect sites and were equipped with several sensors as listed in Table 1. All aircraft flights in this study were coordinated by the OTTER MAC (Multisensor Airborne Campaign). The MAC staff insured that aircraft overflew the OTTER sites and that scientists gathered field data at the same times during the study.

Spectral reflectance measurements, using a variety of field spectroradiometers, were also collected by OTTER investigators as ground truth for remotely sensed image data. Other ground data collected include base station meteorological, field sunphotometer, and ceptometer data as well as various biochemistry, biophysical, physiological, and nutrient cycling measurements. Results from several simulation runs of a forest ecology model (Running and Coughlan, 1988) were retained for future analyses, and data derived from mathematical calculations on raw data and from combinations of bands of raw data, such as leaf area index (LAI), were also generated. The data sets collected for the entire project, expected to total over 16 gigabytes in size, are listed with the number of each data set in the PLDS inventory in Table 2.

OTTER Support Hardware and Software

The data management support of the OTTER project was implemented using hardware and software at the Ames Research Center. During the OTTER project, this support organization was a part of the Pilot Land Data System (PLDS). PLDS (Meeson and Strebel, 1992) was comprised of staff at three NASA centers coordinating their efforts to support scientists in the land sciences community with a wide variety of data management services. All PLDS sites shared several coordinated capabilities, including software and data dictionaries, in order to present an identical interface to all scientists.

The hardware/software component of the data management capability at ARC, in place prior to the commencement of the OTTER project, featured a Sun 4/280 computer running the SunOS 4.1.1 variant of the UNIX operating system. The system had two gigabytes of main storage and a 650 megabyte erasable optical disk drive. The system was accessed via both Internet and Decnet as well as via 2400 and 9600 baud modems. A commercial relational database management system from ORACLE* Corporation was used to store the on-line data and inventory the off-line data. The TAE (Transportable Applications Executive) user interface package was used to organize and execute various data management services. User friendly character mode software written by the Goddard Space Flight Center component of PLDS was implemented under TAE to query the information in the database, to order off-line data for distribution, and to transfer on-line data. Pre-mastering software for use in the publishing of data on Compact Disc-Read Only Memory (CD-ROM) was available on the ARC computer system.

---

* Use of trade names in this paper is for convenience only and does not imply endorsement by NASA or the U. S. Government.

In the OTTER project science plan and in preliminary team meetings, data requirements, as envisioned at project commencement, were enumerated and discussed. In response to these requirements, a data management plan outlining services to satisfy these requirements was written by the PLDS data management coordinator and distributed to all OTTER scientists for comments. Throughout the duration of the project, other requirements emerged from various sources, including the project leaders and investigators, funding managers, and the data management team itself. The major project requirements of the OTTER project and the respective services provided by PLDS/ARC data management staff were:

1) **Data Inventory.** OTTER investigators, widely dispersed throughout the continent, needed to determine which data have been collected and to be able to receive necessary data for their research in a timely fashion. In response, the data management staff: a) worked with data providers to quickly place the data, or information about the data collected under the auspices of the OTTER project, into an on-line inventory on the NASA-network-accessible ARC computer system; b) provided OTTER scientists with access to the data and information via existing software and procedures on the ARC computer which allow database query and data ordering; c) wrote documentation about the use of the on-line system which is specific to OTTER project needs; and d) provided timely tape duplication and distribution services for all remote sensing image data, which are stored off-line on magnetic tape. A file transfer capability exists on the ARC system for the retrieval of data sets stored in on-line files.

2) **Data Certification.** OTTER investigators needed to know the quality of the data that have been collected. A data certification strategy designed to both determine quality and to identify candidate data sets for future publication was instituted.

3) **Data Use Policy.** Investigator-acquired data needed to be promptly processed and submitted to the data management staff. A data use policy designed to control data movement was developed and enforced.

4) **Data Format Coordination.** OTTER investigators needed to have a consistent, easy-to-load-into-a-spreadsheet format developed for investigator-collected data (such as ground-based spectrometer data) in order to better enable data use by OTTER and future investigators. Data management staff: a) coordinated the development and documentation of a usable format and b) enforced the submission of data in the proper format.

5) **Ancillary Data Access.** Investigators needed simple access to coincident meteorological and canopy chemistry data which were archived on-line at the Forest Science Data Bank (FSDB) at Oregon State University. An easy-to-use network file transfer capability from the ARC computer system was created.

6) **Data Publication.** To save this unique set of data for future use by the earth science community, OTTER scientists needed to have the data published on a permanent media. Selected portions of the OTTER data were processed through the ARC pre-mastering facility prior to the generation of CD-ROMs.

7) Services as Required. OTTER investigators needed assistance in various data-related issues that emerge during project implementation. The data management staff remained responsive and flexible throughout the duration of the project in providing support as needed.

Figure 1 displays most of these services in a simplified data flow chart.

DATA MANAGEMENT APPROACH

The support of the OTTER project by the data management staff was accomplished at the level of one full-time equivalent employee while exploiting the talents of the three ARC staff members. The coordinator oversaw the staff and support activities and performs general system design and configuration. The database programmer administered the database software, created and maintained the database structure, and wrote software for the implementation of various data and database procedures in support of OTTER. The third member of the staff is an ecosystem scientist who served as the liaison between the database management staff and the OTTER investigators providing services such as the gathering investigator-collected data and the duplication of data for distribution.

Using all of the skills of this small staff, our approach to data management was to provide the services required for OTTER, exploit the useful aspects of being a part of a major NASA data system, and exhibit the four important characteristics of flexibility, responsiveness, communication, and project focus toward investigators. A good example in the application of existing PLDS capabilities is described in the next section about the OTTER data inventory. Where expertise was lacking in certain areas, such as spectrometer data or data publication, the knowledge and experience of OTTER scientists and staff from other data systems, such as FIFE and GRSFE, were requested. By leveraging the capabilities of the PLDS with the funds provided for OTTER support, the costs of the support effort were actually quite low.

Although the support team is part of a large data system, it was a major goal of this support effort to exhibit a pro-active attitude in its relationship with the OTTER project. To help our staff best serve project needs, emphasis was placed on providing fast, effective, personalized service to each OTTER investigator. System development and other concerns were delayed while requests and problems presented by investigators were quickly but appropriately handled to conclusion. It was also the philosophy of the team not to consume the time of investigators with data management issues that had marginal value, such as the precise delineation of a data dictionary for the project. Also, excessive planning of data management activities was also not advocated. Plans were made to identify the activities that had to be accomplished, and these plans were coordinated with project activities, but their timetable of completion was left largely undefined due to expected changes in requirements and circumstances. While there would be competence in service provision, there would not be the excess of engineering activities which could suffocate the science being performed.

This approach was taken to facilitate good science and high quality data sets. Services which were seen as vital to the efficient accomplishment of science goals, such as timely data distribution, and to the production of a high quality data set at project completion, such as data certification, were given highest priority by the data management staff. Those tasks which did not directly affect those two goals, such as data dictionary agreement by all

investigators, were given lower priority. It was felt that this approach would aid the current science project and future science projects as well.

The three selected services as performed in the support of the OTTER project are described in some detail in the next sections. In the Discussion section, these three services and the four characteristics of a pro-active attitude are compared with the previous data management efforts discussed in the Review section.

Data Inventory

In the provision of the data inventory service, data management capabilities available from the PLDS project as a whole were used in the support of the OTTER project. The host computer system did not require any additional disk space or other capabilities, such as new system access connections, to support the requirements of the OTTER project. The menu front end, the user interface software, and the ordering software, along with all of their development tools, were adapted for use in the OTTER on-line data inventory system. The comments of the OTTER investigators were instrumental in driving some of the enhancements of the on-line system.

The PLDS data dictionary was used as a basis for the data dictionary for most of the OTTER data sets. Special fields, such as site name and number, were required by the project and added to the PLDS data dictionary for all OTTER data sets. Many of the PLDS attribute fields, such as country, were not useful for the project (since the project took place entirely in Oregon) and were therefore removed for access by investigators. The data dictionary was offered to investigators for review and revision, and few changes were recommended. The data dictionary as represented in the on-line system was adequate for the informational querying performed by the investigators. Information about the many data sets was entered into the inventory using proprietary database software tools and verified manually by staff other than the data entry staff. The accuracy and appearance of the information was verified by viewing it from the user's perspective using the user interface software.

Data Use Policy and Certification

A policy on the use of the OTTER data by OTTER investigators and potential collaborators was devised by the science support staff and adopted by OTTER investigators. A statement in this policy required that, under most circumstances, OTTER investigators were to submit data collected by themselves to the science support staff before they could receive data from the database. Another statement regulated the distribution of data to potential outside collaborators basically at the discretion of the project leaders. The science support staff coordinated the interchange of information, such as proposal documents, reviews, and data needs, between the potential collaborator and the project leaders.

To assess and document the quality of data being produced under the auspices of the OTTER project, and to inform all OTTER investigators of the assessment, a data certification strategy was instituted by the Ames science support staff. For remote sensing image data,

this strategy consisted of recording the quality of satellite and aircraft data imagery and ancillary information and placing these assessments into the on-line inventory system. Adopted by the OTTER scientists, a set of levels to indicate the stages of review that the data set entries have undergone was developed, and an attribute, certification level, was added to the database to relay this information to users of the on-line system. If data were provided by a data processing facility merely for testing of user software and were not intended for research purposes, the level is termed "TEST DATA" and that text would be stored in the certification level attribute. If data received from the processing facility were not reviewed by an OTTER investigator, the certification level would be called "UNCHECKED". If the investigator reviewed the data and was willing to vouch for its quality, "PI CHECKED" was placed in the certification level attribute in the on-line inventory system. If a group of OTTER scientists wished to review data together, a certification level of "GROUP CHECKED" would be recorded in the attribute.

The Ames science project support staff created hard copy forms for OTTER scientists to write comments on the quality of each scene, flight line, or flight run (depending upon the instrument). Information, such as the loss of image data (e.g., banding) and an estimate of the extent of atmospheric haziness during data acquisition, were recorded as the OTTER investigators reviewed the data for their research. These assessments of data quality were then placed into the on-line inventory system for each database entry in the text format attribute, certification comments. The date that the assessments were placed into the system was stored in the attribute, certification date. If these were previously unchecked entries, the certification level would be changed to "PI CHECKED". If new or updated information was received on any entry, the certification date was updated

Data Publication

In order to preserve useful data from the large volume of OTTER data collected, Ames science support staff published the wide range of OTTER data sets on CD-ROMs. With the assistance and guidance of OTTER investigators, the staff coordinated the selection, preparation, documentation and pre-mastering of a subset of the 16 gigabytes of OTTER data for the OTTER CD-ROM. Remote sensing image data comprised more that 90% of published OTTER data and required the most attention in the process.

CD-ROM Directory and File Structure- No NASA standard for the directory structure of CD-ROMs or for the format of earth science data files on CD-ROMs currently exists. Examination of the characteristics of existing CD-ROMs published by other projects, such as FIFE (Strebel, et. al., 1991), GRSFE (Arvidson, et.al., 1991), and Bonanza Creek (Way, et.al., 1992), and discussions with several earth science data publishing experts from the JPL Data Distribution Laboratory (Hyon, 1991 and Martin, 1992) led to decisions about the optimal structure of the OTTER CD-ROM.

For the structure of the directories on the OTTER CD-ROM, a top level directory was devoted to each OTTER data set. Within each data set was a set of directories for the OTTER sites and under that, a set of directories for the month and year of the data acquisition. A directory under all of these directories held all of the files associated with a particular remote sensing image or flight line.

For the structure of each of the remote sensing image files on the OTTER CD-ROMs, the Planetary Data System (PDS) labeling convention, which specifies that descriptive information about each file be recorded in an accompanying label file (Martin, 1988), was adopted. In addition to documentation about the image size and format of the data, information from the on-line inventory about remote sensing image data, including certification attributes, was recorded in these label files.

Where the data involved fewer bands and smaller images, the data were reformatted into a format consisting of a single band per file, which enabled the use of a wide variety of image display programs on several computing platforms. Ancillary files, such as calibration and housekeeping files, were offered in ASCII format separately from the data.

Where the data had a large number of bands or were complex in format, the data were stored on CD-ROM in their release format, with the assumption that researchers who are able to process these data already had processing software available for this format. Public domain software for the display of the image files on MS-DOS machines and Macintoshes were also recorded on the discs.

Data Preparation- The OTTER CD-ROMs contain only a selected subset of the remote sensing image data collected for OTTER. Because there were frequently many overflights with the same instrument of the same site at approximately the same time, a coordinated set of the best scenes and flight lines was selected for publication. Through on-line data display and consultations with OTTER investigators who were familiar with the data, the volume of data for publication was reduced while maintaining a representative and useful sample of OTTER data. Documentation on the characteristics of the data, prepared earlier by investigators and by the database management staff, was also included.

To prepare aircraft and satellite imagery for CD-ROM publication, software was written to support a procedure of remote sensing image display, subsetting, and verification. Each scan line of many of the aircraft data sets contained housekeeping information which recorded many data calibration values. Statistics on these values were calculated and, with other summary data, written to a separate ASCII file applicable to an entire flight line. This set of image files and supporting text files (including PDS label files) were then verified for spectral and spatial accuracy and general system compatibility using a variety of operating systems and image processing display devices.

After verification, these remote sensing image files, their label files and the other OTTER data were placed into the prescribed directory structure for the CD-ROM on a reserved hard disk partition on the PLDS computer. OTTER ASCII data in tabular format, such as meteorology and chemistry data, were verified by importing them into various database and spreadsheet programs on major microcomputer platforms and placed on disk. Once verified and moved to the reserved hard disk partition, the pre-mastering software on the Ames computer system created an ISO 9660 format CD-ROM image on magnetic tape from which preliminary "one-off" CD-ROM discs were generated by mastering facilities. These discs were tested on popular platforms by support staff and for accuracy by a group of OTTER scientists. Once the testing procedure was completed successfully, and modifications made to the disc contents, final discs were mastered and replications were made available for distribution to the members of the remote sensing and earth science community.

DISCUSSION

In this section, the experiences in providing the three selected services and applying the four characteristics reflecting a pro-active data management attitude, outlined in the Introduction, to the OTTER project support are compared with the experiences of the earlier science projects in providing similar services, where documentation is available in the literature. Examples of support service design and implementation decisions and the data management team actions in particular situations are described and evaluated by examining the impact upon science and data and the reactions of investigators. It is difficult to "prove" that a design decision is the best or that a characteristic reflecting an attitude is beneficial, but, if the OTTER experience reinforces the experiences of previous project support activity, the value of that characteristic has increased. If these experiences contradict previous experiences, there is perhaps a need to gather more evidence. Only selected aspects will be addressed, as it is impossible to list or discuss all aspects of these services in this forum.

Technical

Data Inventory- The provision of a data inventory capability can be the most time-and budget-consuming service offered by a data management effort supporting a science project. For the OTTER project, the time and money spent on data inventory was relatively small because of the existence of the capability prior to the project. FIS was faced with implementing the on-line system, including creation of a data dictionary, in a relatively short time frame. With the work of competent database management staff and the support of software development staff at the GSFC node of PLDS, the existing PLDS system at ARC could be adapted to adequately serve OTTER data inventory needs. The system was used successfully by OTTER investigators and their graduate students (Strahler, 1992) without complaint and without resorting to the telephone to make personal requests for data. The lower priority that was assigned to the data dictionary and the on-line system permitted more time to be spent with higher priority tasks, such as data distribution and data certification.

As an interesting sidelight, it was discovered that the on-line inventory system and the data distribution services have been most heavily used by OTTER scientists and their collaborators who use remotely sensed imagery. Their need to know exactly which flight lines were acquired on a given date and time and for which site and their need for copies of these large volume data sets for their research required the use of the on-line data inventory system and the data distribution capabilities. This fact underscores the usefulness of formal data management services for projects which collect large volumes of remotely sensed image data. Field ecologists on the project had little reason to use the on-line system as the few files they required for their research could be exchanged on floppy diskettes.

OTTER scientists dealing with remote sensing image data also found great utility in the OTTER on-line inventory in the search for ancillary data useful in the processing of remotely sensed image data. The OTTER investigators searched the database for field sunphotometer readings taken at the same date and time as aircraft overflights in order to atmospherically correct ASAS data. Also, investigators from Canada needed ground truth in the form of spectrometer readings taken by other investigators to compare with airborne spectrometers,

such as AVIRIS. Without the network-accessible, on-line inventory, such a thorough evaluation of data availability would not have been possible. And without a network file transfer capability, the access to the ancillary data would have been slower and more cumbersome.

Data Use Policy and Certification- The FIS data use policy was used as a basis for the OTTER data use policy and is similar in many of its rules. However, the OTTER project manager and the PLDS scientist liaison staff member felt that the one rule regarding the submittal of data to the data management staff should be strengthened. Whereas the FIS policy stated that data shared with other investigators should also be submitted to the database, the OTTER rule stipulated that no data is to be given to other scientists directly. It was discovered that scientists still shared data between themselves regardless of the rule. It appears that the FIS rule, while not as strict, turned out to be more realistic and may have resulted in more data being submitted to the data management staff.

OTTER certification strategy emulated the FIS', including the levels of certification and the use of the certification revision date. The OTTER scheme varied in that full English words were used for the level names instead of codes and that a separate field entitled "certification comments" was created and included in the data dictionary and on-line system for all data entries. As no criticisms or praise was expressed on these changes, their efficacy could not be measured.

Data Publication- In the publication of data on CD-ROM, the philosophy of the OTTER team was to make the data useful to as wide an audience as possible. Toward this end, we emulated the GRSFE structure and generated a CD-ROM that could be used on a wide range of computing platforms, from MS-DOS machines and Macintoshes to Suns and VAXes. Unfortunately, time was only available to prepare image display software for the microcomputers. OTTER coupled the strength of wide accessibility with the approach of FIS to generate separate files of headerless data for each band of remotely image data. This strategy opened the imagery on the CD-ROM to a wider market because most image display programs can handle this generic format, for which reading software can be relatively easily written. While these two major decisions created more work for the data management staff and extended the publication date, the benefits to the perceived audience warranted the implementation of the decision.

Characteristics of a Pro-Active Attitude

For all data management endeavors, the attitude of the staff managing and performing the data management can be very important to the accomplishment of the project goals, particularly for large active interdisciplinary experiments. While the data management services themselves may remain the same, the attitude with which they can be provided by the data management staff can vary significantly and have a direct effect on the science being conducted. For example, when requirements change, as with the unexpected addition of plots within a project site during data collection, the managers of the support team can react by requesting that a full requirement review be conducted with project managers to determine the workload and budgetary impact of this change, or by asking their own staff to determine the difficulty of the change and make the change quickly, if possible. Selecting the latter

option serves the project better by allowing the investigators to know what data were collected in all of the plots shortly after data collection.

The characteristics of a pro-active attitude are important, because, regardless of the project and its requirements, all interdisciplinary field experiments will benefit from the active application of these characteristics. In addition, the proper attitude enhances the provision of all data management services. Technical aspects, such as data inventory and data publication, are important, but these four characteristics are wide reaching in their applicability and effectiveness upon the success or failure of the management of data for active science projects.

Flexibility- A variety of factors change throughout the life of an active science project such as OTTER. Some changes are imposed from outside the project and others are the result of thoughtful re-consideration of project priorities. Regardless of the source, data systems must remain flexible to these changes and respond quickly and effectively (Strebel et al., 1990). As an example, we learned that some scientists were using 8mm magnetic tape drives and preferred to receive large remote sensing image files on that medium. Although PLDS/ARC did not have an 8mm tape drive on its system, the data management coordinator arranged to use an 8mm drive on a nearby computer system for data distribution. This was accomplished at no extra cost for the project and with no adverse impact to other services.

These sorts of efforts were witnessed in each of the three other projects reviewed. With investigators controlling the FIS effort, flexibility was enforced directly. With the large PDS central node playing chiefly a data archival role, flexibility toward the GRSFE project could be more easily accomplished at the geosciences node. Flexibility was demonstrated by the one data management person on the FEDMAC staff who was able to modify formats of spectral data, which were received in a variety of formats to a common format for distribution and publication.

Responsiveness- It was a goal of the OTTER data management staff to get magnetic tapes in the mail within two days of the order by investigators. All other work was halted to gather the data and documentation necessary to satisfy the request. When a request for PLDS/ARC to handle digitized aerial photography data was received, database development staff developed a straw man data dictionary, created the necessary tables in the database system, implemented the GenSQL user interface connections, and had the new data set in the PLDS on-line system within three days. On the day after implementation, an order for all of these data was received.

FIS also exhibited the same philosophy throughout its support of FIFE. The FIS staff was contacted frequently by phone with orders for data from investigators or with requests for assistance. With the responsibility for GRSFE project support at the local geosciences node of PDS, responsiveness is much more possible.

Communication- Communication between the project scientists and the information system is also critical for the support of active science projects. A good example of the level of communication that is necessary can be drawn from the collection, processing, and validation of field data. Scientists in active science projects like OTTER need to know what field data has been collected by their colleagues, and the condition of and methods used to collect those data.

ARC data management staff worked closely with the investigators to understand the various data collection techniques used for collecting spectrometer data, which involved ultralight and light aircraft-based collections as well as ground-based collection. A common format that could be used by all investigators was negotiated at an OTTER team meeting with science support staff leading the discussion. The approved spectrometer data format was written and distributed to the scientists. Guidelines for documenting the spectrometer data, including information about data collection methods and possible problems with the data, were also written and distributed. Data and documentation that have been received from the investigators were examined by ARC science support staff for compliance with the approved format, for general quality, and for completeness of the documentation. Most investigators provided data in the exact format, which will provide a consistent presentation when data are published on CD-ROM. There have been several such instances in the OTTER project where basic and close communication between the science project investigators and the information management staff was critical to the project objectives and, most importantly, to the production of a high quality data set.

FIS staff, partly due to their high level of scientific expertise, were constantly in communication with FIFE scientists throughout the project. Likewise, the GRSFE data management staff also consisted of scientists and could easily communicate in the language of the scientist. The communication in the FEDMAC project was achieved through the close professional relationship between the project leader and the data management staff person.

Project Focus- The basic activities of the support of OTTER have demonstrated that information systems must be closely associated with the project in order to provide optimum service and produce high quality data products. To enable this capability within a large information system, it is important to have knowledgeable and responsive staff as an interface between the project and the information system. The ecosystem scientist on the data management staff focuses on the requirements of the OTTER project and communicates them to the development staff for implementation in the data system. For example, the scientific perspective and active OTTER scientist interaction on the part of the ecosystem scientist on the data management staff was indispensable in the specification of data certification attributes, which were subsequently incorporated into the information system. The liaison also was the focus for the knowledgeable servicing of project participant data needs, such as data distribution and documentation.

The FIS went a step farther than OTTER in that the information system was under direct control of the scientists/users. The scientist liaison in GRSFE could filter the inputs from the central node of PDS to ensure that the project needs were being met by PDS standards and procedures. The small size of the FEDMAC project coupled with the close relationship between the data management staff and the project leader provided a more intimate focus on project requirements.

CONCLUSIONS

Global change research will be highly interdisciplinary for some time to come, requiring a large degree of collaboration and scientist-to-scientist communication (Skole, et.al., 1992).

Much of this research will be conducted as part of interdisciplinary field experiments. These projects depend upon the data management capability to aid in the accomplishment of the scientific goals of the project. The analysis of different techniques of providing common data management services and the identification of beneficial characteristics of a pro-active attitude is important if data management is to improve. As we show in this paper, the OTTER data management staff evaluated and selectively applied and adapted techniques developed by previous data management efforts in planning and implementing project support. In addition, new procedures, policies, and approaches were established to meet the unique demands of the OTTER project.

Some positive and some negative results of the techniques used and attitudes displayed in providing services to the OTTER project were stated in this paper. The active use of the OTTER on-line system by scientists using remotely sensed imagery points to the importance of such a capability to sizable field experiments. All of the characteristics of a pro-active attitude, flexibility, responsiveness, communication, and project focus, appear to be important in the support of field experiments. While the effectiveness of services and proper attitude in project support cannot be "proven", the thoughtful discussion of these techniques and characteristics can promote greater sharing of expertise toward the goal of the improvement of similar efforts in the future. More documentation must be produced so that techniques and characteristics can be evaluated and enhanced toward the advancement of the body of knowledge in the field of active interdisciplinary field experiment support.

Through the work of talented and adaptive data management staffs, the maximum scientific benefit from these active science projects can be realized. In addition, the highly coordinated, widely varying suites of high quality scientific data, including large volumes of remote sensing image data sets, can be preserved as a legacy of the projects.

ACKNOWLEDGEMENTS

LITERATURE CITED

Arvidson, R., Dale-Bannister, M., Guinness, E., Slavney, S., and Stein, T. GRSFE 9-Disc CD-ROM Set, 1991. Washington University, PDS Geosciences Node.

Dale-Bannister, M., Arvidson, R., Guinness, E., Slavney, S., and Stein, T. 1991. The Geologic Remote Sensing Field Experiment: The Archive and Its Use. *Geoscience Information Society Proceedings.* Vol. 22.

Dale-Bannister, M., Personal Communication, 1993.

FIFE Information System (FIS) Internal Document, 1989.

Guinness, E., Personal Communication, 1991.

Hyon, J., Personal Communication, 1992.

Kim, M., Personal Communication, 1993.

Landis, D., Strebel, D., Newcomer, J., and Meeson, B. 1992. Archiving the FIFE Data on CD-ROM. *International Geoscience and Remote Sensing Symposium (IGARSS) Proceedings,* Vol. 1, pp. 65-67.

Martin, T., Martin, M., Davis, R.L., Mehlman, R., Braun, M., and M. Johnson, 1988, Standards for the Preparation and Interchange of Data Sets, Version 1.1, D-4683, Jet Propulsion Laboratory, October 3, 1988.

Martin, M, Personal Communication, 1992.

Meeson, B. Strebel, D., and Paylor, E. 1992. Pilot Land Data System: Science Needs and Implications. International Space Year Conference on Earth and Space Science Information Systems, in press.

Peterson, D. and R. Waring, R. (submitted). Overview of the Oregon Transect Ecosystem Research Project. *Ecological Applications.*

Running, S. and Coughlan, J. 1988. A General Model of Forest Ecosystem Processes for Regional Applications. I. Hydrologic balance, canopy gas exchange and primary production processes. *Ecological Modelling* 42:125-154.

Skole, D., Chomentowski, W., Ding, B., and Moore, B. 1992. Data and Information System Requirements for Global Change Research., *AIAA 30th Aerospace Sciences Meeting.*

Strahler, A, Personal Memorandum, 1992.

Strebel, D., Newcomer, J., Ormsby, J., Hall, F., and Sellers, P. 1989. Data Management In The FIFE Information System. *International Geoscience and Remote Sensing Symposium (IGARSS) Proceedings,* Vol. 1, pp.42-45.

Strebel, D., Newcomer, J., Ormsby, J., Hall, F., and Sellers, P. 1990. The FIFE Information System. *IEEE Transactions in Geosciences and Remote Sensing*, Volume 28, No. 4.

Strebel, D., Landis, D., Newcomer, J., DeCampo, J., and Meeson, B. Collected Data of the First ISLSCP Field Experiment, Prototype Volume. Published on CD-ROM by NASA Goddard Space Flight Center, 1991. Goddard Space Flight Center. Pilot Land Data System.

Way, J.B., McDonald, K., Paylor, E., Karas, G., Chernobieff, and Biery, N. Bonanza Creek 3-Disc CD-ROM Set, 1992. Jet Propulsion Laboratory, Pilot Land Data System.

TABLE 1. Airborne remote sensing instruments used for OTTER data acquisition.

Daedalus Thematic Mapper Simulator (TMS)

Airborne Visible Infrared Imaging Spectrometer (AVIRIS)

Thermal Infrared Multispectral Scanner (TIMS)

Large-format color infrared cameras (RC-10)

Advanced Solid-state Array Spectrometer (ASAS)

NS001 Thematic Mapper Simulator

Airborne tracking sunphotometer

Synthetic Aperture Radar (SAR) instrument

Fluorescence Line Imager (FLI)

Compact Airborne Spectrographic Imager (CASI)

Spectron Engineering (SE) 590

Barnes MMR (Modular Multiband Radiometer)

Surface temperature measurements, and video tapes

TABLE 2. Status of the OTTER data inventory as of July 8, 1992.

| Data Set	Entries	Size[a]
Satellite		
AVHRR Scenes	40	131
Aircraft		
Aerial Photographs	300	N/A
Airborne Sunphotometer Days	2	1.8
Aircraft SAR	6	47
ASAS Tilt Angles	362	5560
AVIRIS Scenes	30	6600
Daedalus TMS Flight Lines	95	499
Digitized Aerial Photographs	7	5
NS001 Flight Lines	68	1224
TIMS Flight Lines	71	596
Field, Laboratory		
Field Sunphotometer Observations	414	.1
Meteorology	10	4
Canopy Chemistry	70	.03
Spectron SE590 Spectra	512	2.5
Timber Measurements	5	.01
Derived		
Forest-BGC Model Runs	4	0.18
Leaf Area Index	7	0.06
TOTAL	**2003**	**14,670.7**

[a] megabytes
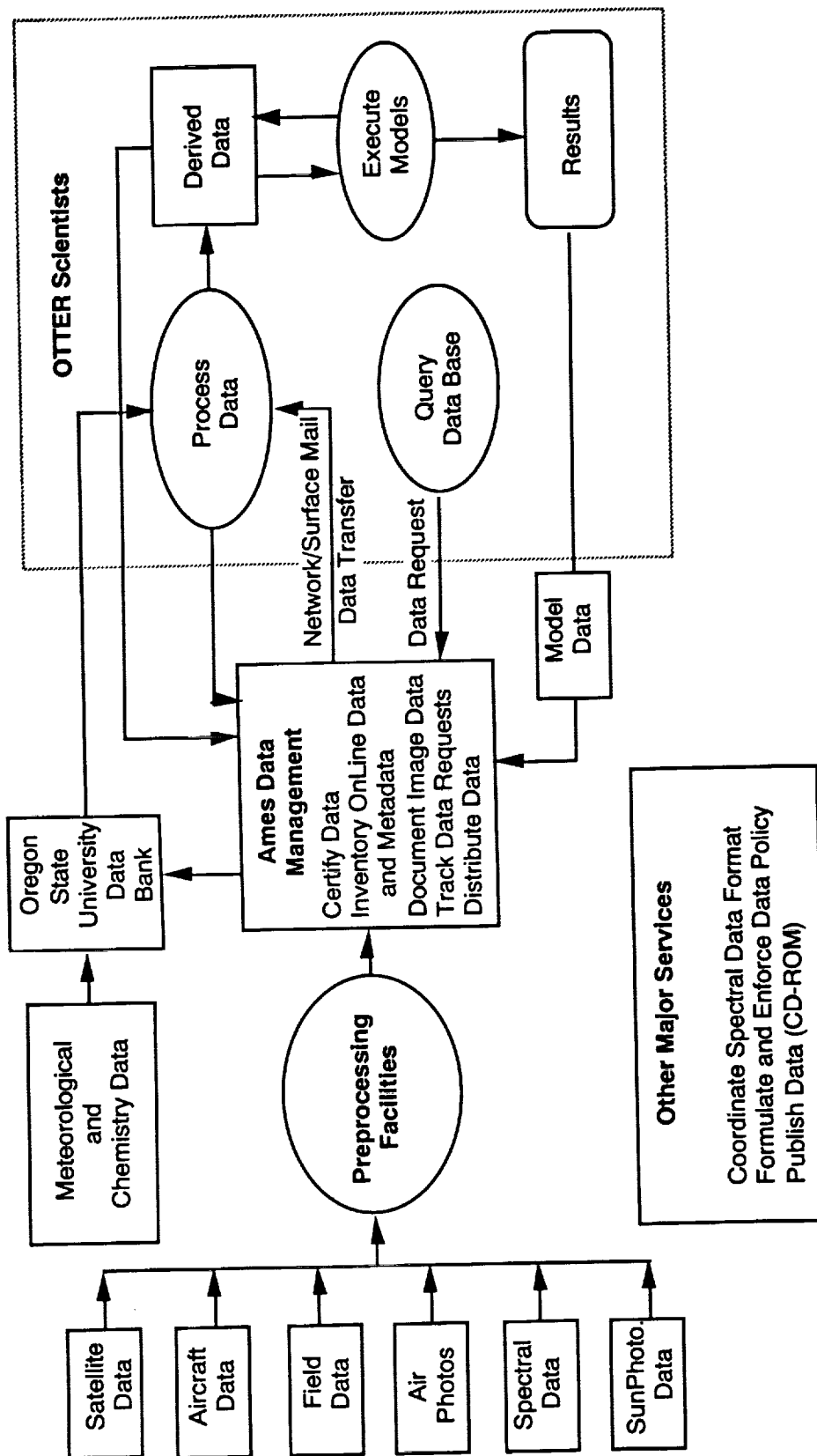
FIGURE 1.

FIGURE CAPTIONS

FIGURE 1.
The flow of data plus some data management services provided to OTTER. Data collected by the OTTER project, from image data to sunphotometer data (shown at left), are preprocessed by various facilities and sent to the ARC data management staff. Information about the data are entered into the on-line inventory and small text files are placed on-line. OTTER scientists, whose activities are included within the dashed-line box, can query the database at ARC, order data, and have data transferred via network or surface mail. Scientists then use these data, as well as meteorological and other field data held at the Oregon State University data bank, in their research to derive new data and execute models. The derived data and the results of these models (lower right) are then returned to the data management facility for inventory and distribution.

REPORT DOCUMENTATION PAGE

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE November 1993	3. REPORT TYPE AND DATES COVERED Contractor Report	

4. TITLE AND SUBTITLE Data Management for Interdisciplinary Field Experiments: OTTER Project Support	5. FUNDING NUMBERS NAS2-13210
6. AUTHOR(S) Gary Angelici,* Lidia Popovici,* and J. W. Skiles**	

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) *Sterling Software, Inc. **Johnson Controls World Services 1121 San Antonio Road 7315 N. Atlantic Ave. Palo Alto, CA 94303 Cape Canaveral, FL 32920	8. PERFORMING ORGANIZATION REPORT NUMBER A-94023

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) National Aeronautics and Space Administration Washington, DC 20546-0001	10. SPONSORING/MONITORING AGENCY REPORT NUMBER NASA CR-4559

11. SUPPLEMENTARY NOTES

Point of Contact: Gary Angelici, Ames Research Center, MS 242-B, Moffett Field, CA 94035-1000
(415) 604-5947

12a. DISTRIBUTION/AVAILABILITY STATEMENT Unclassified-Unlimited Subject Category – 43	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The ability of investigators of an interdisciplinary science project to properly manage the data that are collected during the experiment is critical to the effective conduct of science. When the project becomes large, possibly including several scenes of large-format remotely sensed imagery shared by many investigators requiring several services, the data management effort can involve extensive staff and computerized data inventories. The OTTER (Oregon Transect Ecosystem Research) project was supported by the PLDS (Pilot Land Data System) with several data management services, such as data inventory, certification, and publication. After a brief description of these services, experiences in providing them are compared with earlier data management efforts and some conclusions regarding data management in support of interdisciplinary science are discussed. In addition to providing these services, a major goal of this data management capability was to adopt characteristics of a pro-active attitude, such as flexibility and responsiveness, believed to be crucial for the effective conduct of active, interdisciplinary science. These are also itemized and compared with previous data management support activities. Identifying and improving these services and characteristics can lead to the design and implementation of optimal data management support capabilities, which can result in higher quality science and data products from future interdisciplinary field experiments.

14. SUBJECT TERMS Data management document, Inventory, Ecosystem project management			15. NUMBER OF PAGES 25
			16. PRICE CODE A03

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

NSN 7540-01-280-5500