SESSION 5: DATA MANAGEMENT
CHAIR: BOB COOK, ORNL

5.1 EOS VALIDATION DATA ARCHIVAL POLICY
Tim Suttles, NASA GSFC

In outlining the following EOS Validation Data Archival Policy, it was emphasized that the policy was only applicable to people within the NASA EOS programme:

- Measurements must be publicly accessible though a team SCF, and EOS DAAC or a functionally comparable alternative (DAAC = distributed active archive centre);

- Access should be expedited, delays beyond 6 months from data collection is only allowed for the most exceptional circumstances;

- Whenever possible, public on-line access to validation data should be provided - the "homepage" approach is desirable;

- For non-DAAC approaches, minimum configuration should be anonymous ftp access with catalogue information and a 5-year commitment.

5.2 DATA AND INFORMATION SYSTEM FOR SAFARI 2000
Bob Cook, ORNL

The presentation aimed at informing the meeting of how data and information generated by the Safari 2000 would be managed. Data from Safari-92 was only collated 4 years after the programme had been completed. It is intended that the data management process being designed for the Safari 2000 campaign will facilitated more efficient data collation and the timely completion of an end product. Although the data and information system described essentially indicates what is in place in the US, it is intended that the system be mirrored within the region.

Topics covered in the presentation included: (i) lessons Learned from other field campaigns, (ii) SAFARI 2000 Data Policies, (iii) goals for a Data and Information System (DIS), (iv) description of Project DIS Activities, and (v) roles and responsibilities for investigators and the DIS.

Data Lessons Learned from SAFARI 92:

- The absence of a data management plan during SAFARI-92 hindered data availability for synthesis activities. This indicates the need to get the PI data into a project data and information system quickly, so that investigators are able to analyze and synthesize the data and so make best use of the data.

- Packaging of critical data sets for specific days or areas of interest is needed.

- Experience during SAFARI-92 indicated the benefit of keeping remote sensing data grouped with ground-based observations.

- Awareness of regional sensitivity to data access, both during and after the campaign. There was a desire by the region to have a copy of processed data at an in-region facility for future use.

- Training to regional scientists needs to be provided, so that data and data center can be used effectively.

Benefits of a Project Data and Information System

Benefits to be gained through the establishment of the system include: (i) promotion of common methodologies, (ii) encouragement of PIs to distribute and document their data, (iii) facilitates proper credit being given to data originators, (iv) allows for synthesis, modeling, and validation of data sets, (v) facilitates inter-comparisons among sites and across environmental gradients, (vi) provides support to data requesters, and (vii) it gives visibility to both the project and data.

Elements of the SAFARI 2000 Data Policy

A Data Policy was established at the SAFARI 2000 Regional Implementation Workshop held at Gaborone, Botswana, in July 1999. The main aspects of the policy included:

- Open access to and sharing of data - reference was made to the ICSU/IGBP data policy and the START-DIS model.

- No privileged data access period beyond essential data preparation (up to 12 months), but will the assurance of (i) the protection of intellectual property rights, and (ii) the protection of key data sets for graduate students.

- Need for data sources to be recognized and for co-authorship, particularly for regional scientists.

Goals and Activities of the Data and Information System

The SAFARI 2000 Science Team will establish a project data and information system for implementing the data policy. The *short-term* **goal** of the DIS is to:

- make data readily available to the project team both in the United States and southern Africa (data will be available to the SAFARI 2000 participants within 12 months and to the public within 18 months); and

- prepare selected data sets for synthesis and analysis.

The *long-term* **goal** of the DIS is to assist in preparing selected data for archive and distribution after the SAFARI 2000 project.

Activities required to meet the *short-term* goal were identified as:

(1) Establishing SAFARI 2000 Data Centers in US and southern Africa, using the Mercury System

The mercury is an easy-to-use search system which enables relevant data to be found. In using the Mercury System, data will be distributed via the Internet or hard media. The system includes a metadata and documentation editor.

(2) Metadata, data, and data center software are to be shared so that both the US and within-region centres have identical data holdings.

(3) Estimating number and volume of expected data sets, and identifying "Golden Days" remote sensing data, regional data, field site data, tower flux data, (etc.).

(4) Registration of data sets from PIs, sites, and networks in Mercury in a timely manner.

(5) Compiling of general information about SAFARI 2000 sites (different PIs, multiple sources, published reports).

(6) Compiling regional data of key characteristics for use in models and synthesis. Such data sets may include data from various disciplines such as climate, soil, vegetation, land cover, land use history, hydrology (etc.).

Activities to meet the *long-term* goal of assisting in the preparation of selected data for archive and distribution include:

(1) Application of the 'Twenty Year Rule', i.e. to ensure that 20 years from now, any scientist will be able to understand the data that were collected and archived. This will ensure that the who, what, where, when, how, why questions are answered with regard to the data.

(2) Collect selected data from investigators and data centers.

(3) Quality assure the metadata and documentation.

(4) Prepare the selected project data for archive at the ORNL DAAC and the southern Africa Data Center.

The activities of the data centres is outlined in Figure 52 at data flows from the investigators to the archive. Data flows from investigators to the mirror Data Centers are illustrated in Figure 53.

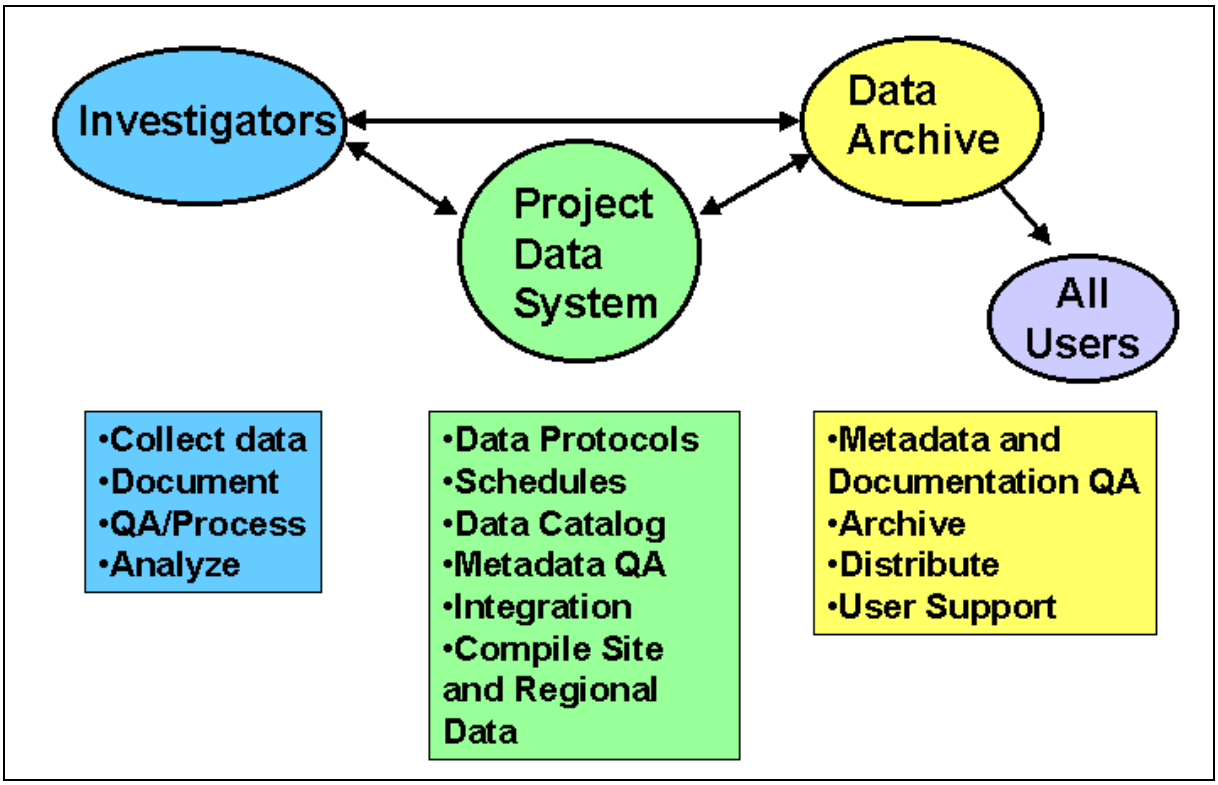

Figure 52. Activities as data flows from the investigators to the archive.

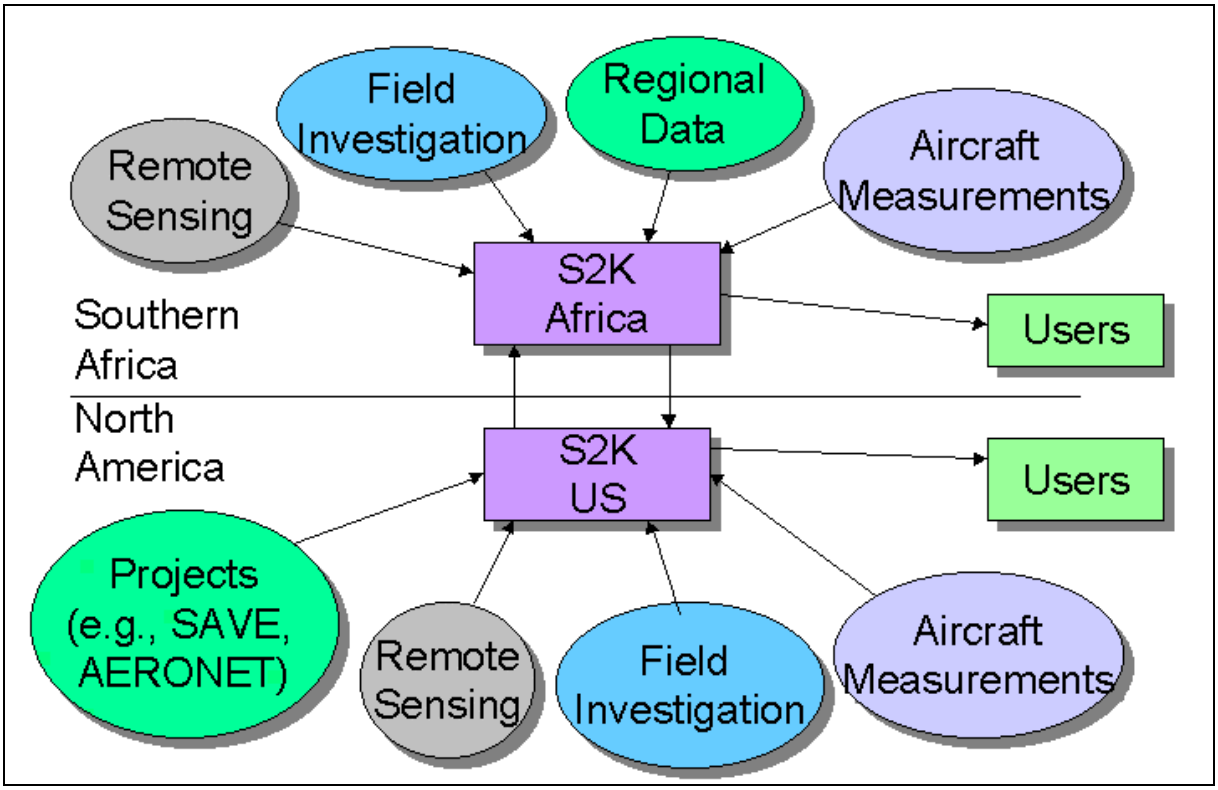

Figure 53. Data Flow from Investigators to Mirror Data Centers.

Critical Data Sets and their Management

Critical data sets to be generated during SAFARI 2000 will include:

- Satellite derived products:
  - Fire distribution and Burn Scar (MODIS; AVHRR)
  - Aerosol Optical Thickness and Size Distribution (MODIS; MISR; AVHRR)
  - fPAR/LAI (MODIS)
  - Land Cover Regional Distribution (% tree cover - AVHRR/MODIS/L7)
  - Vegetation characteristics (MISR; ASTER; IKONOS)
  - Carbon Monoxide and Methane profiles (MOPITT)

- Airborne activity products:
  - ER-2 Data (MAS, MOPITT-A, CLS, AirMISR transects)
  - CV580, SAWB 690, MRF C130 data (aerosol and trace gas chemistry, aerosol physical, optical and radiative properties)

- Ground-based activity products:
  - Aerosol optical thickness and size distribution - AERONET

- Vertical Structure of the atmosphere - Micro Pulse LIDAR data (Ozonesonde Profiles from SHADOZ, Regional Rawindsonde Observations and wind field analyses)
- Plot Level Vegetation Composition, Canopy Structure, Land Cover Distribution (% tree cover)
- Aerosol and Trace Gas Chemistry and Fluxes
- Soil Characteristics

Critical data products for SAFARI 2000 should be kept together for easy long-term access by the community. This includes both field data and selected remote sensing data. The SAFARI-2000 Steering Committee / Science Team should select the data sets required for archive. These data sets will be archived at a data center in southern Africa and at the ORNL DAAC for Biogeochemical Dynamics.

The raw satellite and airborne remote sensing data (source data) will be archived at appropriate data center, e.g. data collected under NASA sponsorship will be archived at responsible DAACs.

Roles and Responsibilities for the SAFARI 2000 Data System

- Responsibility for *acquisition and submittal of data*:
  - PIs are to acquire data, perform initial QA and register with the DIS with appropriate metadata.
  - EOS supported investigators should submit their raw data to the DAAC for their discipline.

- *Timely data distribution* will be done by the data centres using the Mercury System.

- *US-Regional Data Liaison* will be driven by the ORNL DAAC. Responsibilities in this regard will include quick and long-term data distribution; transporting data sets to the region; and regional training on Mercury Data System.

- *Defining data for focused studies and synthesis* is to be the responsibility of the SAFARI 2000 Science Team, to be achieved through Data Workshop(s).

- *Metadata quality assurance and packaging of synthesis data sets* will be done by the Goddard Space Flight Center (Dave Landis and Jamie Nickeson).

- *Long-term data archive* will be the responsibility of the regional data centre and the ORNL DAAC. Raw data will be stored at the appropriate archive.

Discussions

Participants at the meeting made the following recommendations:

* Regional meteorological departments should get together to product a comprehensive description of the regional meteorology.

* In terms of the data flow process illustrated, the importance of the Weather Bureau as a regional data source should be noted.

* Scientists should identify colleagues in the region and integrate them into the process, not just offer co-authorships.

* The SAWB and the other climatological data archives in the region (including Zambia, Zimbabwe, Botswana, Namibia) should keep as comprehensive data sets as possible for the duration of the campaign including: forecast data, and Meteosat imagery (30 min intervals), (etc.). Such data may be held by the bureaus in the short term, but should thereafter be transferred to the Regional Data Centre.

* Data from ground-based activities should also be submitted to the Regional Data Centre.

* The forecast trajectories and the data used to run the trajectories and the model itself is to be archived.

5.3 DATA AND INFORMATION SYSTEM SUPPORT - ORNL AND GSFC
David Landis, NASA GSFC

David Landis acknowledged his colleague from GSFC, Jaime Nickeson. GSFC data management activities will include the packaging of synthesized data for distribution to the science team. It was emphasized that the content of the data set must be determined based on science team inputs. Distribution of the synthesized data is done through the preparation of CD-ROMS or DVDs. DVDs are preferable since they are able to accommodate a much greater volume of data (each DVD can hold 5 Gbytes, whereas CDs only store 600 Mbytes).

In synthesizing and packaging the data for broad distribution, the need for the standardization parameter names, measurement units, data format/content (dates, times, coordinates, sites), and data documentation was highlighted. The aim being to facilitate the implementation of the "20-year Rule". Standardization is crucial particularly considering the multi-disciplinary nature of the data streams.

Partnership between Scientists and Information Managers

It is important that scientists and information managers work together to define the science objectives and to establish priorities with regard to data to be acquired and processed, website development (etc.). Data needs of projects need therefore to be determined in order to identify where most of the resources

should be concentrated. Science teams must drive this process. Although the information managers have the tools available to do the job, they need feedback and direction from the science teams.

It is recommended that a Working Group be established to interact with the Information System Staff. This working group should be organized either by disciplines, by sites or by both. The chairperson of the working group will be interacted with by the Information System Staff and will be expected to contact the necessary PIs. The Working Group should facilitate decision making and consensus on key data issues and provide feed back to the information managers.

Data Documentation

It is crucial that data documentation be complete! Experience has shown that it is best completed near the time of data collection. Where the gap between documentation and data collection is too large, information will be lost. The data documentation ensures usability of the data for long-term studies yet to be defined and is the key to the "20-year Rule". Information System staff will work with the science teams to develop the documentation.

In generating the data documentation a standardized 20 page information sheet was established for the BOREAS experiment. It is recommended that information on instruments, methodologies (etc.) could be 'cut and pasted' from the BOREAS data sets, so that time is spent only on entering information of actual activities specific to the site/campaign, field team details, data collected (etc.).

Data Flow and Data Maturation

The maturation of data can be perceived to occur in a three stage process:

STAGE 1	STAGE 2	STAGE 3
Data/doc status Raw data / doc	**Data/doc status** Refined data/doc	**Data/doc status** Finalized data/doc
Who Individual science team	**Who** Information managers	**Who** Scientific community
Data access Individual science team	**Data access** Project members (peers)	**Data access** Public

GSFC and ORNL Responsibilities and Science Team Responsibilities

GSFC and ORNL will find external data sources, process data ensuring both standardization and procedural quality assurance, and will make data available to the science team and community. A web page will be established for the dissemination of information and for the provision of data links for teams.

The Science Team needs to clearly define all requirements (e.g. content of data sets, manner in which data are synthesized, priorities for archiving) and to ensure that data are scientifically acceptable. The individual investigators will be responsible for compiling the data documentation for their respective data sets.

5.4 DATA MANAGEMENT WITHIN THE REGION
Harold Annegarn, Wits University on behalf of Luanne Otter (CSIR)

Delays have been experienced in setting up a Regional Data Centre, as recommended during the SAFARI 2000 Workshop in Gaborone, Botswana in July 1999. (At the Gaborone meeting it was envisaged that a Regional Data and Training Centre be established at the University of Botswana). In the absence of such a larger structure, it has been necessary to make interim arrangements. These arrangements currently involve the contracting of a professional person, namely Lance Coetzee from Environmental Management Services, as the within-region data manager. He will be employed on a retainer basis, with additional fees being allocated when required. The focus of the within-region data manager will be to capture data sets existing in the region and to integrate new data sets as they become available.

The physical location of the of the (interim) data management centre will be within the Environmental Engineering Department at Wits University. The department has a host of tools, including PCs and the provision of web services (etc.), to facilitate the collection and initial screening of data. The responsibility for data screening is to be that of Leon Herbert, an MSc student at the department. Leon Herbert will work in conjunction with Lance Coetzee in establishing and maintaining the data base.

If a larger regional Data Centre is formed, plans will be put in place to merge the established structure with the new one and to make it a co-operative venture. Both data and experience will be shared. The main aim of commencing with the interim measures is to collate as much of the information currently available prior to the August-September flight campaign.

5.5 UNIVERSITY OF BOTSWANA DATA HANDLING PLANS
Dozie Ezigbalike, University of Botswana

Dozie Ezigbalike titled his presentation “Progress, Problems and Plans”, and started by indicating what was required in terms of the establishment of a data centre at UB, including:

- Minimum configuration of a mirror of the ORNL data centre
- Space, computers and peripherals
- Network connectivity through university's IT infrastructure
- Staff of two: Manager/Director and Assistant
- Funding?

Possible funding for the establishment of a rational data centre at UB could be obtained from USAID. UB were interested but required to know "What's in it for them". It was decided that more than a mirror site was needed, and that the establishment and functioning of the centre would respond to RCSA's mandates which include:

- Sustainable Natural Resources Management;
- Inclusion of training for policy makers and support for environmental policy analysis; and
- The prerequisite of a regional outlook.

The REDMAT Concept

The concept of **R**egional **E**nvironmental **D**ata **M**anagement **A**nd **T**raining Center (REDMAT) has been introduced at UB. Any developments with regard to the SAFARI 2000 data centre would need to be evaluated in terms of the this concept. REDMAT incorporates:

- SAFARI data management requirements;
- RCSA's Strategic Objectives #3 (i.e. cross border management of environmental data and policy development);
- SADC environmental management programs; and
- UB's teaching, research and community engagement mandates.

SAFARI-friendly objectives of REDMAT include the following:

- Acquiring and maintaining equipment for downloading, storing, processing and disseminating environmental monitoring data, specifically remote sensing and other geo-spatial data;

- Act as a data archival and dissemination center for the SAFARI 2000 Regional Science Initiative; and

- Act as a regional resource center for servicing requests for unprocessed and semi-processed images or final data products.

In terms of the staffing requirements of the data centre it is envisage that between 8 and 11 people need to be employed on a phased-basis over a four year period. The salary bill of meeting such a staffing requirement would exceed $ 1 million by year four. The office space requirements could be

accommodated through integration with UB's proposed IT centre building. Hardware and network capabilities could similarly be shared with this proposed facility.

UB Concerns Regarding the Safari-2000 Data Centre

UB was uncertain how the proposal would impact on other funding proposals, specifically KELP/EDDI (Education for Democracy and Development Initiative) with its emphasis on educational technology. The recommendation was made that the two proposals be integrated by taking the following steps:

- Adopting an environmental management theme for KELP/EDDI;
- Establishing a data center to support the activities;
- Use modern edu-tech to deliver REDMAT instructional objectives; and
- Increase research contents of original KELP.

The position of REDMAT in UB's organizational structure was also questioned. A semi-autonomous center with Board of Directors drawn from SAFARI, SADC, UB, Government was envisaged. Further UB concerns included: questions regarding the duration of RCSA and EDDI funding, and reluctance to adopt a natural resource/environmental management theme for KELP.

Current Status and Next Steps

A final proposal has been submitted for KELP/EDDI funding and has high chance of success, but no explicit data archiving component has been included in this proposal. The RCSA is still, however, interested in the data management components of REDMAT. Further consideration will be given regarding the equipment and staff required by the two aspects.

It should be noted that the question was asked during the data management team discussion held at this workshop yesterday, whether the data centre should still be at UB. The LEAD South African Project is to create independent data bases and a metadata system at Wits University with professional support. The data centre to be established there will use compatible tools to the ORNL DAAC (i.e. it will most likely implement the Mercury System). During April and May, Dozie Ezigbalike will visit Wits University to work on a proposal for RCSA with Leon Herbert and consult (via email) with Bob Cook and David Landis. This proposal will be structured so as to address RCSA's strategic objectives and expressed desire.

The proposal will ask for essential equipment and staff, and will motivate for the establishment of more than just mirror site. A centre which will facilitate some data preparation services for to the region is envisaged, in addition to the coordination of training in the region on use of the data. The proposal will suggest to RCSA to call for EOIs from other institutions, not just UB. If UB is selected as the base, RCSA will need to ask UB for space, rather than UB proposing it. In short, African Data Handling Plans need not necessarily be focussed on UB.

5.6 SAWB DATA BASES
Andrew van der Merwe, SAWB

The South African Weather Bureau maintains the regional meteorological data on behalf of the World Meteorological Organization (WMO), incorporating data from Malawi, Zimbabwe, Zambia, Angola, Mozambique, Swaziland, Lesotho, and Botswana. Telex machines and email is used to collate the data. Further arrangements are needed to get the data timeously.

A climatological data bank (120 years of data) for southern Africa is available from the SAWB. Climate surface data is available from 1930. Meteosat satellite imagery has been stored for the past 7 years in CD format. Raw meteorological data is stored in WMO message format (one would need to know the codes and format to make use of these data). Data can be sent to whichever regional data centre was established.

Information subsequently received from Louis Botha with regard to facilities at the SAWB Databank is included in Appendix C

Questions and Answers:

Q Are there any restrictions on the production and distribution of data from the SAWB?
A There is a copyright on the Meteostat data. All other data is classified as Resolution 40 data, i.e. available to everyone.

Comment Data originators such as the SAWB must be recognized by co-authorships.

Comment Internet access for Pietersburg is currently being worked on. The option of dial-up accounts would be expensive. Ideally UNINET could be used so that it is not necessary to pay for individual calls. Leon Herbert is to investigate the possibility of this so that individual dial-up accounts are not necessary.

5.6 FUTURE MEETINGS, PROJECT PUBLICITY AND ACTION ITEMS
Bob Swap, UVA and Tim Suttles, NASA GSFC

Future Meetings

The following workshops are anticipated within the next three years:

- A Data Workshop will be held early in 2001. It is uncertain whether this workshop should deal with data handling or first results. This workshop could possibly be coupled with the wet season campaign, its aim being to get the dry season science team together to talk about the best case studies and about quality assurance. It was agreed that 6 months after a campaign is the best time to do this.

- It is likely that First Results will be presented as a scientific conference in the later half of 2001.

- A Data Synthesis Workshop will be held early on in 2002. This is likely to be at Lusaka, and will represent the next major meeting in Africa after the experiment. Mr Mukelabai is the champion for this meeting.

- The Science Safari Conference will be held late in 2002 or early 2003. The American Geophysical Union will hold the conference.

Publicity

The handling of publicity from NASA's perspective was addressed by Tim Suttles. Dr Suttles indicated that a publicity officer, namely Steve Cole, will be in charge of publicity for the campaign from NASA's side. A plan is in place to compile a video of all NASA's activities. As part of the making of this video, it was proposed that footage be taken at Pietersburg during the flight campaign. The meeting was asked whether scientists would feel it an intrusion and whether it was possible to film from an aircraft. The meeting felt it would not interfere, and the Aerocommander team indicated that if filming from one of their aircraft was done during a one hour flight outside of project activities then it would be fine.

Scientists were asked to include poster material for the public open day to be held on 12 August. All aircraft will be on the apron and posters and ground displays will be in place on this day. The E-Theatre will be out to the region during the campaign.

Item Requiring Action:

The loss of capabilities in Mozambique in terms of Rawisonde capabilities. The recommendation was made to the SAWB use the additional sondes for data validation and inter-comparison. SAWB personnel were requested to put together a proposal in this regard to submit to the committee. (Beira, Lilongwe and Harare were indicated as potential sites for the use of the additional sondes. Beira was indicated as being the location of choice. Equipment at Lusaka and Mongu are to be repaired.) The meeting indicated that only a dozen of the sondes would be needed in Pietersburg to facilitate inter-comparison studies. Others would be needed north of Harare.

5.7 MERCURY SYSTEM DEMONSTRATION
Bob Cooke, ORNL DAAC

Mercury represents a metadata search and data access system which is designed to provide access to dynamic, distributed research data. Advantages of the system include:

- Low requirements for the data provider;
- It incorporates a software tool available to help produce data documentation;
- No complicated data ordering function;
- It is easily implemented for new projects (i.e. weeks, no months);
- It is currently being used by EOS Land Validation, SAFARI 2000, DAAC's Regional and Global Data Initiative, and LBA.

Mercury Support for SAFARI 2000

The system is an internet-based, distributed system that will search metadata to identify data sets of interest and delivery them to a user. As such it facilitates up-to-date (nightly) links to SAFARI 2000 data. Metadata and data files reside on investigators' servers connected by the Internet, e.g. field measurements for Kalahari Transect at UVa, ETM+ data at EROS Data Center, AERONET data at NASA Goddard, and regional data at ORNL DAAC.

Mercury can be used to identify data of interest. Where Internet connections are fast, data can be downloaded from investigator's Web site. Where Internet is slow or data are voluminous, the Data Center can package files on tapes or CD-ROMs and distribute these data to the necessary parties.

Overview of Mercury

Documentation, data files and standards represent the foundation for Mercury. The investigator uses the software tools to document the data. The metadata and data are linked and placed in a predefined location at the investigator's computer and the investigator's computer connected to the internet. The metadata downloaded from the ORNL has links to the full metadata, documentation and data at the researcher's site.

The ORNL DAAC Mercury search screen is illustrated in Figure 54. From the pull-down topic menu it is evident that searches can be done per site, project, source, (etc.). If data were grouped for a specific site (e.g. Skukuza or Mongu) it would be possible to gain access to all data for these sites across all disciplines and including ground-based, airborne and satellite data.

In order to facilitate access to your data set through Mercury you would need to fill in the on-line form which will create the metadata and documentation required for your data set. *Metadata* descriptors

enable data to be found. The *documentation* facilitates the understanding of the data. Metadata variables include:

- Contact information of principal investigator(s);
- Data set information (e.g. data set title, project, site, spatial coverage, time period, parameter description, keywords, abstract)
- Data access information (e.g. data set status, data center contact, data centre URL, data set citation and data set link.)

Mercury makes available an extensive glossary in order to explain what is meant by certain terms in the on-line questionnaire, and offers tutorials on how to complete the information required for the creation of the metadata and documentation.

An example of the metadata summary for one data set is given in Figure 55 and the metadata report generated for this data set indicated in Figure 56. The underlined text in Figure 56 represent active areas which may be used to gain direct access to the websites at which the data are stored.

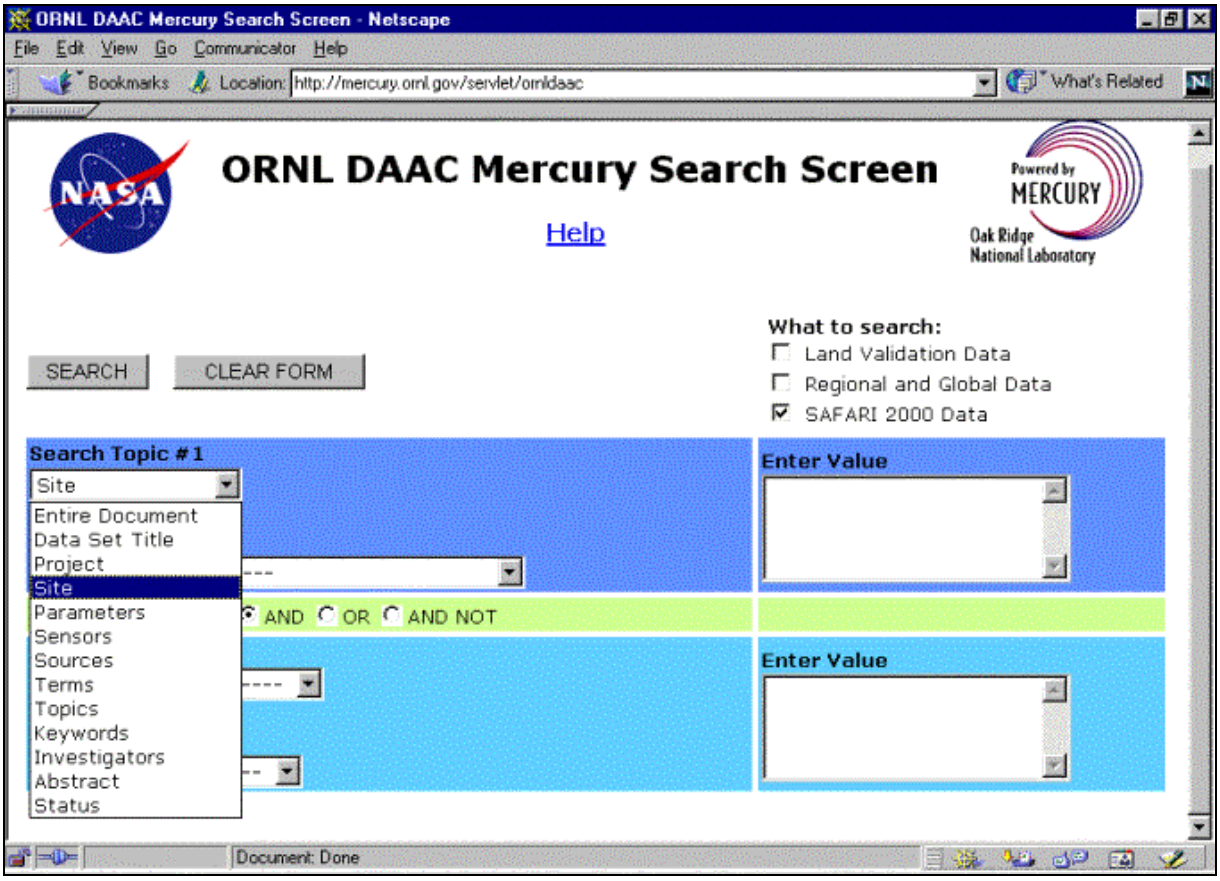

Figure 54. ORNL DAAC Mercury Search Screen.

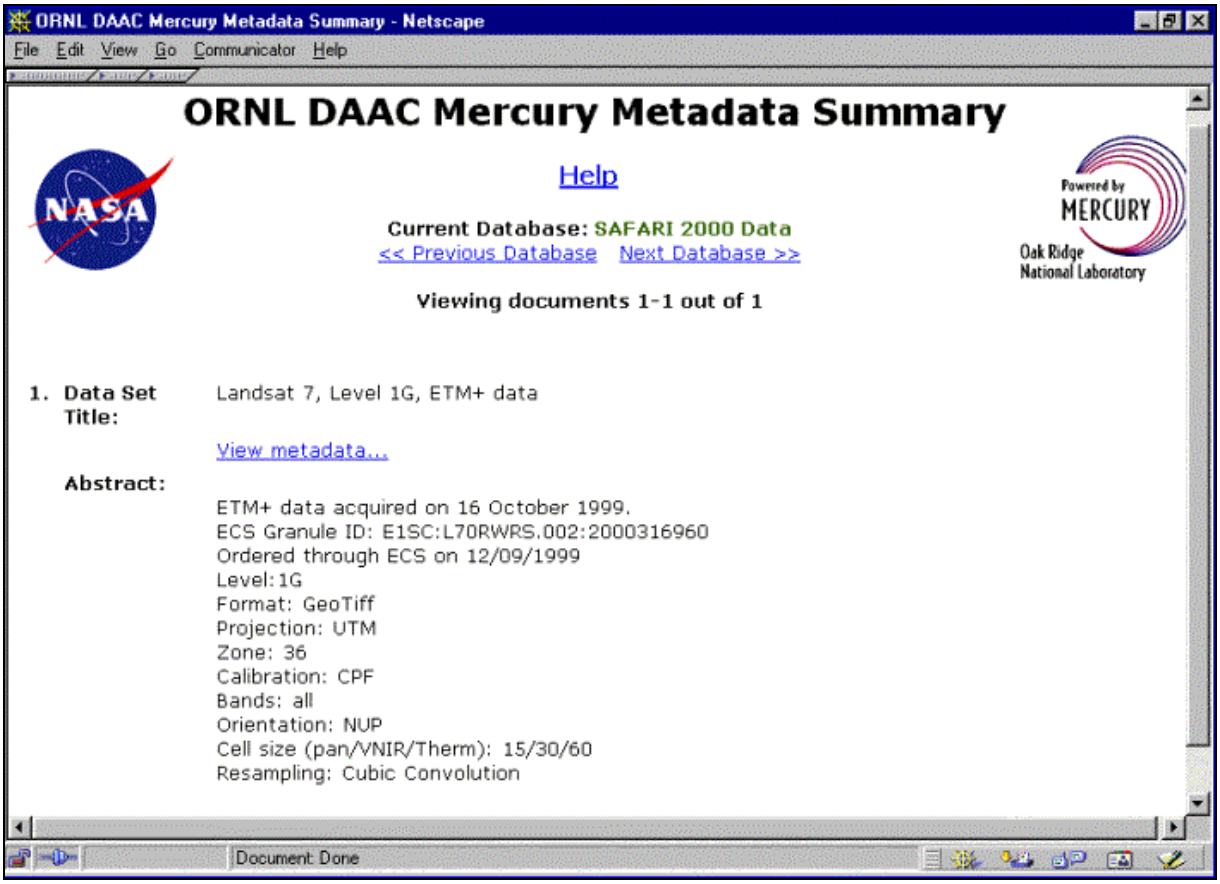

Figure 55. Example of ORNL DAAC Mercury Metadata Summary.

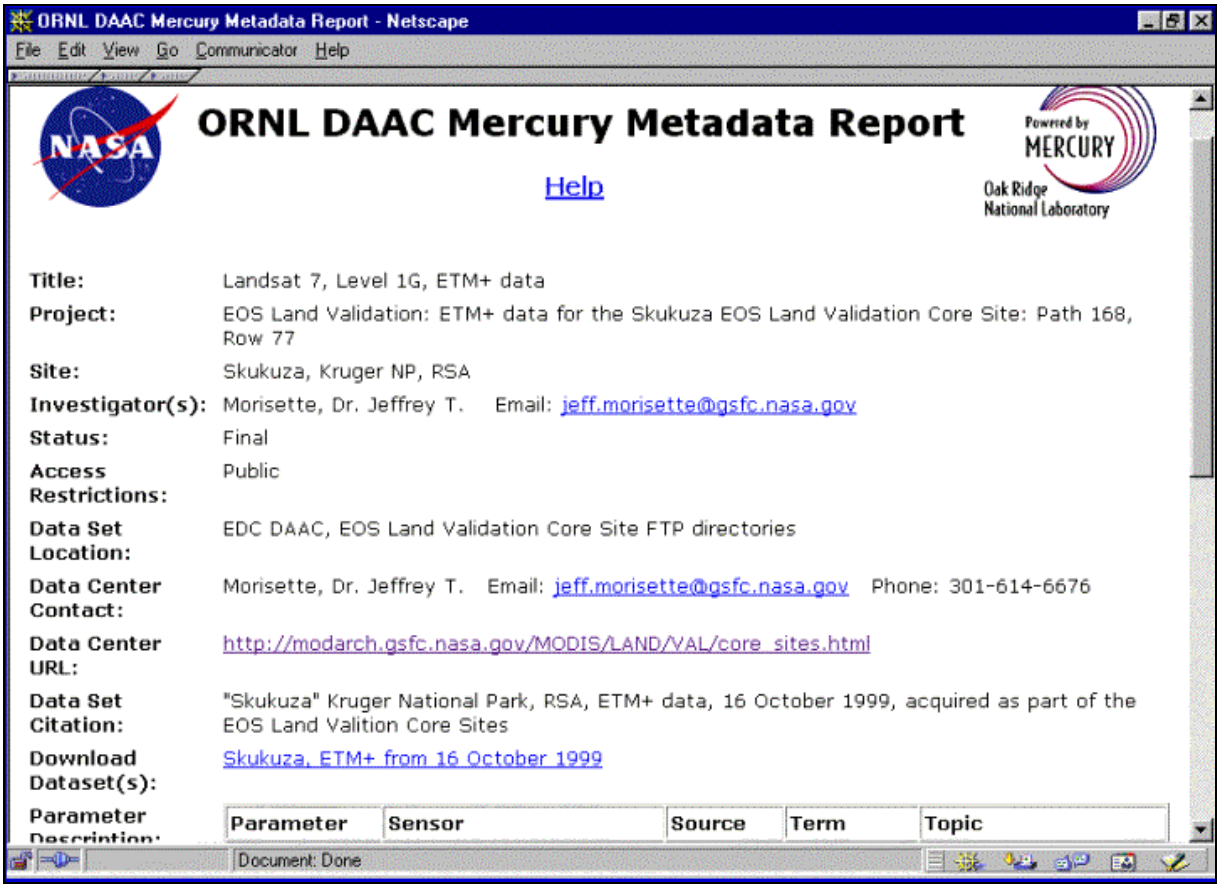

Figure 56. Example of ORNL DAAC Mercury Metadata Report.

Regional Data Sets for SAFARI 2000

Key regional SAFARI 2000 data sets are currently being compiled at the ORNL DAAC for use in SAFARI 2000. These data will be of value to the SAFARI 2000 modeling teams as well as to SAFARI 2000 researchers in general. All data will be available through the Mercury metadata search and data access system.

The approach being adopted is as follows: a series of southern Africa data sets for historical and future climate, hydrology, soils, digital elevation, and vegetation will be generated from existing global and African data sets. These data will be subsetted using the following bounding coordinates:

Westernmost long.	+5
Easternmost long.	+60
Northern most lat.	+5
Southernmost lat.	-35

Based on discussions with the SAFARI 2000 science team, data sets that will be subsetted include:

Climate data:

- East Anglia Climate Research Unit Data (monthly mean climatology for 20th Century)
- Cramer and Leeman's Potsdam Climate
- Future climate scenarios from Hadley Centre, Canadian Climate Centre, and NCAR

Soil Data:

- IGBP DIS Soils Data Set
- ISRIC -WISE Data Set
- Zobler data
- Zinkie data

Vegetation Data:

- Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation
- Wilson, Henderson-Sellers' Global Vegetation & Soils, 1-Degr
- Haxeltine and Prentice
- 1km Land Cover Data, Global Land Cover Facility, University of Maryland
- Treecover data, Global Land Cover Facility, University of Maryland
- Potential Vegetation data set, Ramankutty and Foley

Land Cover/Land Use Change

- Leeman
- 1700-1992 Historical Croplands data set, Ramankutty and Foley

Hydrology

- Cogley Global Hydrographic Data
- Freshwater Wetlands (Stillwell-Soller)
- Climate, People, and Environment Program (Global River Discharge Data)

Others

- Digital Elevation Model
- Population

At this time, three data sets have been subsetted, viz.:

1. Olson's Major World Ecosystem Complexes Ranked by Carbon in Live Vegetation
2. Wilson, Henderson-Sellers' Global Vegetation & Soils, 1-Degree
3. ISRIC-WISE Global Data Set of Derived Soil Properties on a 1/2 x 1/2 degree grid

It is envisaged that 22 data sets will eventually be available. The data files will be provided in ASCII Grid format for ArcInfo. Each file contains a single ASCII array in a geographic (lat/lon) projection.

The ASCII file consists of header information containing a set of keywords, followed by cell values in row-major order. The file format is

<NCOLS xxx>
<NROWS xxx>
<XLLCORNER xxx>
<YLLCORNER xxx>
<CELLSIZE xxx>
{NODATA_VALUE xxx}
row 1
row 2
row n

where xxx is a number, and the keyword NODATA_VALUE is optional and defaults to -9999. NCOLS and NROWS are the number of columns and row in the array, XLLCORNER and YLLCORNER are the longitude and latitude value for the lower left corner of the array. Row 1 of the data is at the top of the grid, row 2 is just under row 1 and so on. The end of each row of data from the grid is terminated with a carriage return in the file.

The files will be in UNIX compressed format (".gz") in order to provide for faster download. Each data set will include a README file with information specific to that data set as well as information on the procedure used to subset the data.

All data sets are provided in geographic projection and retain their original resolution. The data sets will be clipped to an area roughly contained within the following latitude longitude coordinates:

minimum longitude 5W	maximum longitude 60E
minimum latitude 35S	maximum latitude 5N

A short data summary taken from the metadata that will be on-line in Mercury for the three data sets currently available is given in Appendix C. For questions or comments with regard to the regional data sets, please contact Bob Cook and/or Veronica Fisher:

Bob Cook	cookrb@ornl.gov
Veronica Fisher	vfisher@facstaff.wisc.edu

Or visit Mercury on-line at:

http://mercury.ornl.gov/servlet/ornldaac

The Metadata Editor Tool URL is:

http://www-eosdis.ornl.gov/cgi-bin/MDE/S2K/access.pl

To request an ID for access it is necessary to call DAAC User Services (+1) 865-241-3952, or send an e-mail to mye@ornl.gov. All comments on this tutorial, Metadata Editor, or Mercury should be sent to editor@daacs.esd.ornl.gov.

DATA INFORMATION SHEETS

Questionnaires on the types, volumes and dates of availability were circulated at the workshop. All data sheets completed at the workshop, or prior to the finalization of these proceedings are included in Appendix E. Supplementary information were also obtained from the UK Met Office in terms of their instrumentation which will be in use during the campaign (Appendix F). Information is also available from the Kruger National Park Scientific Services on their metadata listings (please contact Harry Biggs for the more current listing at biggs@parks-sa.co.za).