SAFARI 2000 DATA POLICY
Approved February 5, 2001

SAFARI 2000 OPEN DATA POLICY

The SAFARI 2000 data policy is based on open data sharing, cooperation, and synergism. The data policy is implemented to ensure:

- Participants have access to data in a timely manner;

- Intellectual property rights are protected; and

- Co-authorship, acknowledgement, or credit is given to data originators and principal investigators.

Two sources formed the basis for the SAFARI 2000 data policy:
- The International Council of Scientific Union's policy on open data and data sharing; and

- NASA's EOS Validation Program data policy which stipulates that principal investigators submit preliminary results to the project archive within 6 to 12 months from the date of measurement.

The data policy will persist beyond the life of the SAFARI 2000 Science Steering Committee.

ARCHIVE

SAFARI 2000 will establish a Southern African Regional Data Center to provide tools for documenting, storing, searching, and distributing data and images. In addition to a repository and retrieval centre, the regional data centre will be a focal point for training in remote sensing. Data generated by SAFARI 2000 will be permanently archived in southern Africa at the Southern African Regional Data Center and in the U.S. at the Oak Ridge National Laboratory Earth Observing System Distributed Active Archive Center (ORNL EOS DAAC).

SHARING OF DATA

All SAFARI 2000 data should be made available to all SAFARI 2000 participants through open online storage or submission to the Regional Data Center. Data, once quality assured, must be submitted promptly to the Center. Access to some of the data at the Regional Data Center will be limited to registered SAFARI 2000 participants for a period of 18 months after completion of the project. (see SAFARI 2000 Participant Registration section).

Eighteen months after official completion of the project (estimated to be June 30, 2003), SAFARI 2000 data will be placed in the public domain, except for commercially restricted or copyrighted data. SAFARI 2000 acknowledges that some investigators may

be required by their funding agencies to follow established agency guidelines for public distribution of project data on shorter time scales.

STUDENT PROTECTION

Data sets that are the subject of student theses or dissertations will be given special consideration. Data reserved for student use may not be published, included in a review in a manner that will compromise the originality of the student's thesis, or cited by other investigators (including SAFARI 2000 participants) prior to completion of the student's thesis, or without written permission from his/her advisor. This special consideration will be limited to 2 years after data collection for a masters student and 3 years after collection for a doctoral candidate.

To reserve a data set for student use, the student's supervisor will be asked to register data sets with the SAFARI 2000 Steering Committee, who will maintain a registry of such data and associated projects. The request must be supported with an abstract of the student's research project, indicating the scope of the use of the data and details of the data requiring special protection. The registry will be maintained as part of the Regional Data Center and will contain the abstracts and details of protected data sets. This registry will be accessible to SAFARI 2000 participants.

Registration of protected data sets for student use will not imply exclusive use of data by the student. After consultation of the student's advisor, use of data in ways that will not infringe on a student's thesis topic will be permitted. In case of uncertainty or dispute, the SAFARI 2000 Steering Committee will decide on fair use.

Students are required to register data and submit data sets to the SAFARI 2000 Data Centers within 6 to 12 months of collection to ensure that data are available after completion of dissertation. Use of SAFARI 2000 data in a student's thesis requires acknowledgement as outlined in the SAFARI 2000 Acknowledgements section below.

SAFARI 2000 PARTICIPANT REGISTRATION

A registered SAFARI 2000 participant is defined as a scientist who has registered with the Regional Data Center and been approved as a participant by the Science Steering Committee. Registration requires an explicit declaration by the registrant to abide by the SAFARI 2000 Data Policy, and completion of an abstract describing the registrant's project(s) and links to the SAFARI 2000 science framework.

Registration as a SAFARI 2000 participant provides a scientist with privileged access to SAFARI 2000 data, notification of and invitations to SAFARI 2000 data meetings, and inclusion on SAFARI 2000 mailing lists.

METADATA REGISTRATION AND DATA AVAILABILITY

Guidelines

SAFARI 2000 participants are expected to actively participate in data sharing. First, they should inform SAFARI 2000 of their primary and derived data holdings through the registration of *metadata* as soon as possible after collecting or procuring a data set. Metadata registration occurs in the Mercury system maintained by the ORNL DAAC (http://www.daac.ornl.gov/cgi-bin/MDE/S2K/access.pl). This entails completing a brief, standardized form with metadata information. Metadata are used to describe the characteristics of datasets such as content, quality, condition, and geographic extent. Metadata include the names of the investigators, the title of the data set, keywords, and descriptive information related to the project for which the data are collected.

Second, when primary data reduction and quality checking procedures are complete, or secondary data products are stabilized, participants should make such data sets available to SAFARI 2000 participants. This may entail sending complete data sets to the SAFARI 2000 Data Center. It can also be done by updating your metadata file in Mercury and providing WWW links to data sets stored on online computers. Before data sets can be considered final and complete, the investigator must also provide documentation of the data. An outline for this documentation will be provided.

The metadata database facilitates the rapid sharing of information between SAFARI 2000 participants by assembling descriptive summaries of all research activities in a searchable index. Such information allows for the identification of potential gaps and or overlaps in the SAFARI data archive and may aid in the avoidance of redundant field efforts. The metadata entries can be made publicly searchable, even though the data sets themselves may be restricted to SAFARI 2000 participants or copyright restricted.

Time Frames

Time frames for metadata registration and data availability may vary by project within SAFARI 2000. Metadata registration should be carried out at the earliest possible time after planning, collecting or procuring the data sets. The SAFARI 2000 Steering Committee strongly encourages that data be released to SAFARI 2000 investigators as soon as possible but no later than one year after collection. Early availability of data will be especially important for validation of remote sensing products and for the integration and synthesis of SAFARI 2000 results.

Access Restrictions

Originators may specify appropriate restrictions to data sets deposited at the Regional Data Center. These restrictions should be included in the metadata registration. Such restrictions may include an initial (up to 18 month) limit to registered SAFARI 2000 scientists. [Note that some data sets, such as those associated with NASA funded projects, will have totally unrestricted access and are not limited by this policy.]

Data Deposits by Country

This data policy does not replace or supercede the requirements of International Agreements between individual countries to deposit data collected from individual countries at specified repositories.

AUTHORSHIP

The analysis and interpretation of data by all participants who generate the data are supported. Co-authorships must include all who have contributed substantially to the work. In case of doubt, co-authorship should be offered. Given that collaboration signifies the spirit of SAFARI 2000, joint publications between in-region and out-of-region scientists are especially encouraged.

Participation in SAFARI 2000 is not restricted to scientists or students who participated in field studies. After registration as a SAFARI 2000 participant, investigators using data generated by other investigators should request permission from the data originator and offer co-authorship. In cases where data from other investigators are a minor contribution to a paper, the data should be cited or acknowledged in a mutually agreeable fashion.

While open data sharing is at the essence of the SAFARI 2000 Data Policy, the intellectual property rights of the data originator are not subsumed by this Policy. Sharing of data through this Policy or placing data with the Regional Data Center does not constitute publication of the data or ideas, and does not implicitly or explicitly allow use or citation of the data or ideas without proper permission.

SAFARI 2000 ACKNOWLEDGEMENTS

SAFARI 2000 investigators, as a condition of registration, agree to include an acknowledgement in each publication, report or public presentation arising from participation in SAFARI 2000. Such participation shall include field campaigns, use of facilities and equipment funded in part or whole by SAFARI 2000, logistical and administrative support, access to countries or research locations under the SAFARI 2000 name, salaries or bursaries from SAFARI 2000 associated funding or otherwise benefiting from SAFARI 2000. The wording shall be similar to the following:

> ***This study was part of the SAFARI 2000 Southern African Regional Science Initiative.***

Data providers and funding agencies may request additional acknowledgements.

Upon publication of results, SAFARI 2000 requests that at a minimum a full reference of the work be provided to the Regional Data Center for archival purposes and preferably an electronic and/or paper copy of reprint.

USE OF DATA IN SYNTHESIS AND INTEGRATION STUDIES

When data are required for modeling or integrating studies, the originator of the data should be consulted before data or derived products are incorporated or published in a review or integrated study. The scientist collecting such data shall be credited appropriately by either co-authorship or citation.

PRIOR DATA

The provisions of the data policy apply to data collected as part of the SAFARI 2000 Initiative. Pertinent data collected by participants prior to SAFARI 2000 are not included in the open access. Placement of such data in the shared data pool shall be at the sole discretion of the originator or institutional owner of such prior data. Access to such data shall be negotiated directly with the data owners, and the SAFARI 2000 name shall not be used to gain privileged access to prior data.

ACCESS TO RESTRICTED AND PROPRIETARY DATA SETS

The SAFARI 2000 Science Steering Committee will seek to enable access to restricted data sets that are pertinent to the SAFARI 2000 science aims. This will include making contact, through the Country Representatives on the Steering Committee, with relevant officials and institutions, to negotiate such access on behalf of SAFARI 2000 participants.

Further, the SAFARI 2000 Science Steering Committee will attempt to acquire and distribute relevant commercial data products on behalf of all SAFARI participants. In some cases, however, commercial or governmental restrictions beyond the control of the Steering Committee may limit distribution.

CONFLICTS

Conflicts over the interpretation of this Data Policy, or its implementation, will be submitted to the SAFARI 2000 Science Steering Committee. Participants agree to comply with the decisions of the SAFARI 2000 Science Steering Committee in such conflicts.

_____