

Data Policy for the SAHRA Geo-Database (SGD)

Introduction

The SAHRA Geo-Database (SGD) is designed to facilitate storage of, and access to data via the Internet. Data access refers to viewing and downloading of data from the SGD. The data will be two kinds – spatial and tabular data (including time-series data) – and will be maintained in the form of a relational database. The intention is to provide support to SAHRA researchers, non-SAHRA researchers and the general public, with different (appropriate) levels of data access. Terms used in this document are defined in the Appendices. The SGD can be accessed from:

http://sahra.arizona.edu/research_data/SAHRAGeoDB/

Purpose of the Database & Database Management

The purpose of the database is to:

- (a) Store data and metadata developed by SAHRA research groups (including but not limited to field data, processed data, model outputs, etc., along with appropriate metadata descriptions)
- (b) Store certain relevant data and metadata obtained from public sources
- (c) Support SAHRA research groups with existing data and information about data sources

The goal of SAHRA database management is to:

- (a) Ensure data integrity and appropriate data access to the maximum extent possible
- (b) Maintain and improve the database and the mechanisms for access to data
- (c) Update the database whenever new data becomes available

Purpose of the Data Policy

The purpose of this SAHRA data policy is to establish guidelines and agreements that govern data submission, access to and usage of data stored in the SGD. By doing so, we hope to facilitate the widest possible accessibility to SAHRA relevant data while protecting the data from unauthorized access and use, and maintaining a desirable level of data quality. The data policy includes agreements about confidentiality, compliance with standards (storage, metadata etc.), and methods for data submission and access.

Data Custody (Classification)

Data stored in the database are classified into three categories based on permitted level of data access:

- (a) Public Data – accessible without needing permission from the Data Custodian¹

¹ see definition of Data Custodian in the Appendices to this document

- (b) Semi-Private Data – limited to access by SAHRA Researchers only (SAHRA and Data Custodian must be acknowledged in any publications or documentation of use²)
- (c) Private Data – limited to access by data custodian and custodian-authorized users only

“Private Data” will be stored in the database pursuant to the terms of this Data Policy. Data custodians may approach the SAHRA administration to discuss special terms and conditions regarding storage of and accessibility to their data. However, in most cases, the following data policy will generally be applied:

- 1) All data collected under SAHRA funding/support must be posted on the SGD within one year of collection, along with appropriate metadata defined in the SAHRA Metadata Guidelines
- 2) Data-Custodians can have exclusive access to their own data for up-to one year from date of collection. Others who may wish to have access to the data must obtain permission from the Data-Custodian, who can grant access.
- 3) After a period not to exceed one year from date of collection, access will be opened to all SAHRA researchers by the database administrator. A data access log will be implemented (to track data usage) which can be made available to the Data-Custodian upon request. Users will be required to acknowledge³ SAHRA, NSF and the Data-Custodian in any documentation and/or publications, and are encouraged to inform the Data-Custodian of any data use and publication plans.
- 4) After a period not to exceed two years from date of collection, the data will be made publicly accessible by the database administrator. Users will be required to acknowledge⁴ SAHRA, NSF and the Data-Custodian in any documentation and/or publications, and are encouraged to inform the Data-Custodian of *any data use and publication plans*.

User Registration & Data Access

Data access is free, but all users are required to fill out a registration form the first time they enter the database, to gain access. The database can be accessed in two different ways:

- 1) Public Interface - this browser-based mode enables any user to access public data and to submit request for private data. It will also enable authorized users to access semi-private data.
- 2) Direct Access – this network-based mode (administrator-controlled) enables SAHRA users to submit and gain access to private and semi-private data, and also access the public data.

Registration to obtain data access (via username and password) is by means of the Public Interface. The following information is collected at time of registration:

- (a) Name
- (b) Email address
- (c) Phone number
- (d) Affiliation
- (e) Acceptance of the data usage agreement

² See the Appendices to this document for discussion of acknowledgements

³ See acknowledgement formats @ <https://intranet.sahra.arizona.edu:8443/administration/disclaim.html>

⁴ See acknowledgement formats @ <https://intranet.sahra.arizona.edu:8443/administration/disclaim.html>

- (f) Agreement that the requested data will not be resold
- (g) Agreement that the private data will not be redistributed without data custodian permission.

All metadata stored in the database is accessible to and searchable by all users regardless of any restrictions on access to the data.

All data stored in the database is cataloged. The catalog contains dataset name, classification, custodianship name(s), date of creation, dates for reclassification, date(s) of modification, and short description. The catalog is accessible to all users.

Direct access to data can be granted only by the database administrator. All users have direct access to data owned by them, and to all publicly available data stored in the database. Access to, and use of, Private-Data requires permission by the Data-Custodian.

Personal information about data users (and their access to data) is stored in the database and is made available only to the SAHRA Administration, except that the data access log for a Private Data set will be made available to the Data-Custodian upon request. Such information will not be made public or otherwise released to any third party. Statistics regarding data usage (which do not contain personal information about users) may be published or otherwise used by SAHRA administration (or their designees). SAHRA, as the database custodian and distributor, reserves the right to monitor data access. Monitoring information will be used to evaluate the efficiency and efficacy of database access and to prevent unauthorized database access.

Data Usage

Public-Data may be freely used for research and educational purposes. Data users are expected to acknowledge data custodian(s) in any and all documentation and publications. The form of acknowledgement is defined in metadata.

Private-Data users are required to acknowledge SAHRA, NSF and the Data-Custodian in any documentation and/or publications, and are encouraged to inform the Data-Custodian of any data use and publication plans. Users having access to Private-Data are not permitted to provide these data to any other party without explicit documentation in the form of written consent from the Data-Custodian.

The form of data acknowledgement is defined in the Appendices to this document as well as in the metadata. Any commercial use of public data requires consultation with the data custodian. Access to data by means of the SGD shall not be interpreted as authorization by SAHRA, the University of Arizona, or the National Science Foundation to make commercial usage of the data. Any opinions, findings, and conclusions or recommendations expressed by users of SGD data are not to be interpreted as reflecting the views of SAHRA, the University of Arizona, or of the National Science Foundation.

Data Submission

Users with Direct-Access to the database can load data to pre-structured tables, or can use alternative table structures in consultation with the database administrator. New data submissions

should be reported to the database administrator. Alternatively, users may submit data through the database administrator using standard data formats or follow the SAHRA Data Submission Guidelines on the database website. The upgrades of the Public-Data may only be changed by the database administrator.

Data Standards

All data submitted to the database must be accompanied by metadata describing the nature of the data. Metadata should be created based on FGDC and ISO standards. Spatial data sets must include information that enables proper spatial association of the data (e.g. in the form of geographic coordinates and projection). Preferred allowable data formats include:

- Spreadsheets (MS Excel, Quattro Pro)
- Delimited text files
- Database files (MS Access, dBase)
- ArcGIS shapefiles, coverages and interchangeable files
- Triangulated Irregular Network (TIN)
- NetCDF
- Raster data files (TIFF, JPEG, ASCII, BIL, ESRIGrid and other binary raster files).

Appendices:

Appendix A. Definition of Terms

Data – refers to digital data and its metadata derived from any research activity such as field observations, collections, laboratory analysis, experiments, or the post-processing of existing data. It includes the spatial representation of data.

Data User – refers to an individual to whom access to the data has been granted

Data Custodian – are individuals (principal investigators) or institutions directly responsible for creating and/or distributing data. The Data Custodian is identified in the metadata.

Data Distributor – refers to SAHRA as the database custodian and data distributor.

SAHRA Administration – refers to the Director, Deputy Director and Associate Directors of SAHRA.

Database Administrator – refers to the person appointed by SAHRA Administration to be responsible for overseeing all database related activities.

SAHRA Researchers – refers to scientists, students and staff formally affiliated with SAHRA and defined as such in the SAHRA Online Management System (OMS).

Appendix B. Acknowledgements

All materials produced using data accessed via the SAHRA Geo-Database, must include a clear indication of sources such as:

"This material is based upon data provided by SAHRA (Center for Sustainability of semi-Arid Hydrology and Riparian Areas) under the STC Program of the National Science Foundation, Agreement No. EAR-9876800."

In the case of Private-Data or Semi-Private-Data, the acknowledgement must also include a statement such as (please also see metadata for additional conditions of use):

"Permission to use these data was granted by SAHRA investigator <name, title and affiliation>"