

Southwest Experimental Garden Array Data Access Policy, Data Access Requirements, and General Data Use Agreement

31OCT2013 Revision

Southwest Experimental Garden Array (SEGA) Data Access Policy

The SEGA data policy includes three specific sections designed to express policies regarding the release of SEGA data products, user registration for accessing data, and the licensing agreements specifying the conditions for data use. These policies are intended to align with current National Science Foundation guidelines and the policies developed by the Long Term Ecological Research network office.

SEGA Data Release Policy

Data and information derived from publicly funded research, totally or partially from SEGA funds from NSF, Institutional Cost-Share, or Partner Agency or Institution where a formal memorandum of understanding with SEGA has been established, are made available online with as few restrictions as possible, on a nondiscriminatory basis. SEGA scientists are expected to make every effort to release data in a timely fashion and with attention to accurate and complete metadata.

Data

There are two data types:

Type I – data are to be released to the general public according to the terms of the general data use agreement (*Appendix 1*) as soon as possible after collection and no later than two years from collection. Type I data include both streaming data collected by the SEGA instrumentation (climate and engineering data) and user-collected data (data collected as part of specific experiments by researchers). Streaming data collected by SEGA will be made available in near-real time in most cases. User-collected data may include typical ecological data (i.e. plant physiological measurements), genetic data, or other periodic or one-time data not collected by the SEGA instrumentation.

Type II - data are to be released to restricted audiences according to terms specified by the owners of the data. Type II data are considered to be exceptional and should be rare in occurrence. The justification for exceptions must be well documented and approved by the lead PI (name) and the SEGA Data Manager (Paul Heinrich). Some examples of Type II data restrictions include: locations of rare or endangered species, data that are covered under prior licensing or copyright (e.g., SPOT satellite data), or confidential data on human subjects covered by the U.S. National Research Act (45 CFR 46; see also The Belmont Report of 18 April 1979: <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html>). Researchers that

make use of Type II Data may be subject to additional restrictions to protect any applicable commercial or confidentiality interests.

While the spirit of this document is to promote maximum availability for ecological data in either Type I or II status, there also are criteria by which priority for data release may be determined. Primary observations collected from core research activities directly supported by SEGA must receive the highest priority for data release. In general, all data collected by the SEGA instrumentation will be made available to the public in near-real time. Data collected by other sources to which SEGA supported research has added value is also a high priority. Other types of data including non-SEGA data that were acquired for SEGA research, student thesis data, schoolyard SEGA data, or legacy data that already suffer from inadequate documentation or format obsolescence may be ranked a lower priority with justifications provided in their data management policy. Finally, some data may be determined to be of lowest priority for archiving on the grounds that they are interim data that led to final products that carry the scientific value. These might include data files created during stages within an analytic workflow, raw or replicate data values that were subsequently aggregated or processed for release, or individual outputs from stochastic models.

Metadata

1. Metadata documenting archived/online data sets of all types listed above will be made available when, or before, the dataset itself is released according to the terms above.
2. All metadata will be publicly available regardless of any restrictions on access to the data.
3. All metadata will follow SEGA recommended standards and will minimally contain adequate information on proper citation, access, contact information, and discovery. Complete information including methods, structure, semantics, and quality control/assurance is expected for most datasets and is strongly encouraged. Whenever possible Federal Geographic Data Committee standard metadata (<http://www.fgdc.gov/metadata>) will be used for spatially explicit data and Level 5 Ecological Metadata Language (EML) metadata (<http://knb.ecoinformatics.org/software/eml/eml-2.1.1/>) will be used for non-spatially explicit data.

SEGA Network Data Access Requirements

Access to all SEGA data is subject to requirements set forth by this policy document to enable data providers to track usage, evaluate their impact in the community, and confirm users' acceptance of the terms of acceptable use. These requirements are standardized within SEGA to provide contractual exchange of data between Data Providers and Data Users that can be encoded into electronic form and exchanged between computers. This will allow direct access to data via a common portal once these requirements have been fulfilled. The following information may be required directly or by proxy prior to the transference of any data object:

Registration

1. Name
2. Affiliation
3. Email Address
4. Full Contact Information

- Acceptance of the General Public Use Agreement or Restricted Data Use Agreement, as applicable.
- A Statement of Intended Use that is compliant with the above agreements. Such statements may be made submitted explicitly or made implicitly via the data access portal interface.

Data providers wishing to impose further requirements beyond these are encouraged to include them in their Restricted Data Use Agreements accompanying the datasets.

Data Use Agreements

Datasets released by SEGA will be accompanied by a use agreement that specifies the conditions for data use. For Type I data, this shall be the General Data Use Agreement (*Appendix I*). This document specifies general roles and the obligations and rights enjoyed by each regarding the use of most dataset released for general public use. For Type II datasets, a Restricted Data Use Agreement must be provided with the dataset that identifies the specific restrictions on the use of the data and their justification. Because these are expected to be unique to the dataset, no template is provided, although in most cases the General Data Use Agreement can be modified as needed to create a Restricted Data Use Agreement. Grounds for restricting data may include (but are not limited to) the need to restrict access to species, habitats or cultural resources protected by legislation; rights of privacy granted by human subjects legislation; or protection of intellectual, financial or legal rights over the data held by a third party.

This policy becomes effective when approved by the SEGA Executive Committee. It may be revised by, or at the request of, the same body.

Appendix I

General Data Use Agreement

Definitions

“Data Set” – Digital data and associated metadata derived from any research activity, such as field observations, collections, laboratory analysis, experiments, or the post-processing of existing data, and identified by a unique digital object identifier (DOI) issued by a recognized cataloging authority such as a site, university, agency, or other organization.

“Data User” - individual to whom access has been granted to this Data Set, including his or her immediate collaboration sphere, defined here as the institutions, partners, students, and staff with whom the Data User collaborates, and with whom access must be granted, in order to fulfill the Data User's intended use of the Data Set.

“Data Set Creator” - individual or institution that produced the Data Set.

“Data Set Owner” – individual or institution that holds intellectual property rights to the Data Set. Note that this may or may not be defined as a legal copyright. If no other party is designated in the metadata as Data Set Owner, it may be presumed that these rights are held by the Data Set Creator.

“Data Set Distributor” - individual or institution providing access to the Data Sets.

“Data Set Contact” - party designated in the accompanying metadata of the Data Set as the primary contact for the Data Set.

Conditions of Use

The re-use of scientific data has the potential to greatly increase communication, collaboration and synthesis within and among disciplines, and thus is fostered, supported, and encouraged. Permission to use this dataset is granted to the Data User free of charge, subject to the following terms:

- 1) *Acceptable use.* Use of the dataset will be restricted to academic, research, educational, government, recreational, or other not-for-profit professional purposes. The Data User is permitted to produce and distribute derived works from this dataset provided that they are released under the same license terms as those accompanying this Data Set. Any other uses for the Data Set or its derived products will require explicit permission from the dataset owner.
- 2) *Redistribution.* The data are provided “as is” for use by the Data User. The metadata and this license must accompany all copies made and be available to all users of this Data Set. The Data User will not redistribute the original Data Set beyond this collaboration sphere.
- 3) *Citation.* It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the Data User will properly cite the Data Set in any publications or in the metadata of any derived data products that were produced using the Data Set.

Citation should take the following general form: *Creator, Year of Data Publication, Title of Dataset, Publisher, Dataset identifier*. For example:

McKee, W. 2001. Vascular plant list on the Andrews Experimental Forest and nearby Research Natural Areas: Long-Term Ecological Research. Corvallis, OR: Forest Science Data Bank: SA002. [Database]. <http://www.fs.fsl.orst.edu/LTER/data/abstract.cfm?dbcode=SA002>. (21 October 2004)

4) *Acknowledgment.* The Data User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this dataset in any publications where the Data Set contributed significantly to its content.

Acknowledgments should identify the supporting party, the party that received the support, and any identifying information such as grant numbers. For example:

Data sets were provided by the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the U.S. Forest Service Pacific Northwest Research Station, Corvallis, Oregon. Significant funding for collection of these data was provided by the National Science Foundation Long-Term Ecological Research program (NSF Grant numbers BSR-90-11663 and DEB-96-32921).

5) *Notification.* The Data User will notify the Data Set Contact when any derivative work or publication based on or derived from the Data Set is distributed. The Data User will provide the data contact with an (electronic) reprint of any publications resulting from use of the Data Set and will provide copies of, or on-line access to, any derived digital products. Notification will include an explanation of how the Data Set was used to produce the derived work.

6) *Collaboration.* The Data Set has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, collaboration and/or co-authorship with the Data Set Creator.

By accepting this Data Set, the Data User agrees to abide by the terms of this agreement. The Data Owner shall have the right to terminate this agreement immediately by written notice upon the Data User's breach of, or non-compliance with, any of its terms. The Data User may be held responsible for any misuse that is caused or encouraged by the Data User's failure to abide by the terms of this agreement.

Disclaimer

While substantial efforts are made to ensure the accuracy of data and documentation contained in this Data Set, complete accuracy of data and metadata cannot be guaranteed. All data and metadata are made available "as is". The Data User holds all parties involved in the production or distribution of the Data Set harmless for damages resulting from its use or interpretation.

Forward comments to:

Paul Heinrich

Merriam-Powell Center for Environmental Research

Northern Arizona University

*Applied Research and Development, Room 225
Flagstaff, Arizona, 86011*