

ORNL TES-SFA Data Policy: Archiving, Sharing, and Fair-Use

Revision Note: This 2019 TES SFA Data Policy combines, updates, and replaces the earlier and separate SPRUCE and TES SFA Data Policies.

Overview:

The open sharing of Terrestrial Ecosystem Science Scientific Focus Area (TES-SFA) data, modeling products, and documentation among researchers, the broader scientific community, and the public is critical to advancing the mission of DOE's Program of Terrestrial Ecosystem Science.

The policy is applicable to all TES-SFA participants including ORNL Staff, cooperating independent researchers, and to the users of TES SFA data products. Data collected by TES-SFA researchers, results of analyses and syntheses of information, and model algorithms and codes will be quality assured, documented, and archived and will be made available to the public through the SPRUCE (<https://mnspruce.ornl.gov>) and TES-SFA (<https://tes-sfa.ornl.gov>) websites and in due course DOE's permanent archive Environmental Systems Science Data Infrastructure for a Virtual Ecosystem ([ESS-DIVE](#)).

Archived data products are freely available to the public. Users should acknowledge the contribution of the data provider with the citation (including DOI) as provided in the documentation and acknowledge the U.S. DOE Program for Terrestrial Ecosystem Science in any resulting presentations and publications.

TES-SFA data policies are consistent with the most recent DOE policies for "Public Access to the Results of DOE-Funded Scientific Research" https://mnspruce.ornl.gov/sites/default/files/DOE_Public_Access%20Plan_FINAL.pdf and the "Statement on Digital Data Management" <http://science.energy.gov/funding-opportunities/digital-data-management/> and with the [Memorandum of Understanding](#) between the U.S. Forest Service and UT-Battelle for management of the SPRUCE Experiment.

Scope:

TES SFA data generation tasks include large ecosystem manipulations, Carbon-Cycle observations, database compilation, and fundamental process studies integrated and iterated with modeling activities. The goal of this policy is to ensure that all the data products are preserved at a high-enough level of quality, discoverable, and readily accessible so that they are useable well into the future.

- The centerpiece is the SPRUCE experiment, a large-scale experimental manipulation testing multiple levels of warming at ambient and elevated CO₂ (eCO₂) on the C feedbacks from a *Picea-Sphagnum* ecosystem. SPRUCE has implemented a highly instrumented experimental platform for the long-term observation of the mechanisms controlling the vulnerability of organisms, ecosystems, and ecosystem functions to increases in temperature and exposure to eCO₂ treatments within a northern peatland high-C ecosystem.
- Other TES SFA efforts aim to improve mechanistic representation of processes within terrestrial biosphere models by furthering our understanding of fundamental ecosystem functions, and their response to environmental change. These tasks include smaller-scale, process-level manipulations to quantify root trait and function research, mechanistic studies of soil C-cycling modeling (i.e., MEND), and the extensive literature compilation of plant root traits and corresponding environmental data (i.e., FRED, V2). The TES SFA supports the long-term monitoring of landscape flux measurements at the Missouri Ozark Flux (MOFLUX) site and complementary measurements to better interpret responses. Also, The TES SFA

preserves the legacy data from Walker Branch Watershed, EBIS, and Precipitation Throughfall Displacement studies.

All data collected at the SPRUCE facility, all results of laboratory experiments and sample analyses, synthesis of information, genomics analyses, and model products (inputs, codes, outputs) developed in support of TES SFA tasks will initially be submitted to the respective SPRUCE or TES SFA data archive in a timely manner such that data will be available for use by project scientists and collaborators and, following publication, the public, through the SPRUCE (<https://mnspruce.ornl.gov>) and TES SFA (<https://tes-sfa.ornl.gov>) websites.

SPRUCE experiment and TES-SFA task research data that are made available to the public, will also be provided to DOE's permanent data archive, the Environmental Systems Science Data Infrastructure for a Virtual Ecosystem ([ESS-DIVE](#)). ESS-DIVE will archive and publicly share digital research data obtained from observational, experimental, and modeling research that is funded by the DOE's Office of Science under its Terrestrial Ecosystem Science (TES) programs within the Environmental Systems Science (ESS) activity.

Timeline for Data Archiving and Sharing:

The diverse set of TES SFA measurements vary greatly in their temporal measurement frequency, ranging from, for example, 30-minute averages of 1-minute air temperature measurements, lengthy soil incubations, to annual aboveground vegetation measurements. The complexity of measurement methods varies widely, from an instantaneous reading to an extensive extraction process and genetic sequencing. The amount of processing, analysis, quality assurance effort and time needed to create a given product varies accordingly.

For sharing data among SPRUCE participants: automated environmental measurements are now available within hours of collection through the data visualization and download tool (VDV); annual survey and seasonal measurement data are usually available within 120 days from the completion of the measurements; results of laboratory analyses of vegetation tissues, soils, isotopic composition, etc., are generally available within 60 days from completion of analyses.

For sharing SPRUCE data with the public: environmental measurements are provided as annual updates; annual surveys and seasonal measurement results are available with publication of analysis papers. Similarly, modeling products and results of laboratory analyses are made available concurrently with publication of papers.

SPRUCE project collaborators must agree to follow this data policy and to provide research data in a timely manner to be archived and shared with the public.

Data collected by TES SFA researchers, results of analyses and syntheses of information, and model algorithms and codes developed in support of the TES SFA tasks will be quality assured and documented during an initial period of exclusive use to publish papers based on the data. Subsequently all materials will be archived and made available to the public.

Quality Checks:

Related to the timeline for data sharing are the general quality checks to be performed prior to data sharing among participants and prior to public access. Experience has shown that most TES SFA research data are of sufficient quality for project and public use following Level 1 checks performed by the data originator. Level 2 checks are most often performed when integrating and analyzing data for use in a peer reviewed publication or developing input or validation products for modeling tasks.

Guidelines for defining data Quality Levels:

Level 1 Quality indicates an internally consistent data product that has been subjected to quality checks and data management procedures including, for example: site documentation has been reviewed for completeness; procedures and protocols were reviewed for compliance; calibrations and quality control samples have been evaluated and necessary corrections made; the data have been adjusted for "zero drift" (continuous measurements), or for "blank bias" (lab analyses) as appropriate; consistency checks have been performed with other measurements within the same data file. These internal consistency checks might include diurnal analyses to look for expected patterns, or time series analyses to detect outliers, extreme values, or time periods with too little or too much variation.

Level 2 Quality indicates a complete, externally consistent data product that has undergone interpretative and diagnostic analysis by the project participants. For example, in addition to Level 1 procedures the data have been closely examined by the data manager and/or data users for external consistency when compared to other related data. External checks might include correlation by scattergram, comparison of data with other similar data for the same time period, and comparison of a measurement made by two different methods. If comparisons were not within the precision of the measurements, then measurement records and other information have been reviewed. When data products have been updated as a result of additional quality checks or discovery of errors, the data should be resubmitted to the archive and the quality level documentation changed (e.g., to Level 2).

To be designated Level 2 quality, data may also have for example, received intense scrutiny through additional consistency checks and use in modeling tasks. As a result of these additional checks, certain fields may have been adjusted or gap-filled for completeness to meet model input needs. Data may also have been aggregated to different time intervals. These value-added products should also be archived as Level 2 products.

For completeness, *Quality Level 0* data are products of unspecified quality that have been subjected to minimal processing in the field and/or in the laboratory -- raw data, data sheets, scanned data sheets, notebooks, etc. These products may be submitted to the archive for long-term storage but will not be shared.

Code Sharing:

Public release of SPRUCE-specific E3SM code will be managed by the E3SM project as part of a collaboration agreement between the ORNL TES SFA and E3SM and subject to E3SM policies and licensing (<https://e3sm.org/resources/policies/>). Development branches of the E3SM code for research purposes will also be available through <https://github.com/E3SM-Project/E3SM/>. Code developments will be discussed and agreed upon by the TES SFA modeling team, with the understanding that our goal as a group is to make the developments here available to the larger community as soon as possible. For reproducibility, publications using model output will include information about the specific release or development branch used in the simulations. Public release removes the 'rights' of code developers to be automatically considered for co-authorship. However, we encourage users of the released model to consider informing or including those developers to the extent it would benefit the users' analyses.

For archiving of baseline model versions, TES SFA investigators may take advantage of the new Department of Energy (DOE) Office of Scientific and Technical Information (OSTI) software services platform and search tool for DOE-funded code – DOE CODE. DOE CODE provides functionality for collaboration, archiving, and discovery of scientific and business software. The DOE CODE platform (<https://www.osti.gov/doecode/>) is an open source project available on GitHub.

TES SFA Data and Modeling Products Used in Publications:

Research data and modeling research products presented in publications resulting from TES SFA research will be made available to the public concurrent with publication of the paper. This includes data that are displayed in charts, figures, images, etc. In addition, the underlying digital research data used to generate the displayed data will be made as accessible as possible. In this context, research data is defined as the data required to validate the published results and modeling codes, inputs, and outputs for transparency and potential reproducibility of results. For example, SPRUCE archives data products specific to a publication as “Title”: Supporting Data and “Title”: Modeling Archive data products. These datasets are accessible through the SPRUCE archive and are cited (with DOI) in the respective publication.

Data Fair Use Policy:

The data provided for public access are freely available and were furnished by ORNL TES Investigators, the SPRUCE Team at ORNL, the U.S. Forest Service, and cooperating independent researchers who encourage their use.

Users of these data products and project information should do the following:

- Inform (via email) the scientist(s) of your use of the archived data and of any publications that result from your use of the data. Contact information is provided on the project website.
- Frequently check the publicly accessible data archive to ensure that you are using the latest version of the data.
- Acknowledge (1) data products, including model simulations, as a citation with corresponding data DOIs, as provided in the data archive documentation, (2) project information downloaded from a website as a bibliographic web citation, and (3) general project information as an acknowledgement or personal communication. No other form of citation is applicable.
- Acknowledge the agency or organization that supported the collection of the original data when publishing original analyses and synthesis results using these data.
- Include these terms as publication keywords as applicable: SPRUCE Experiment, ORNL, ORNL TES SFA, U.S. DOE Office of Science, Marcell Experimental Forest, Northern Research Station, U.S. Forest Service.
- Provide an electronic reprint of your independent work to the TES SFA so that all publications resulting from these data may be tracked, recorded, and referenced. Contact information is provided on project websites.

Disclaimer of Liability:

Data and documents available from the SPRUCE (<https://mnspruce.ornl.gov>) and TES SFA (<https://tes-sfa.ornl.gov>) websites were prepared as an account of work sponsored by an agency of the U.S. Government. Neither the U.S. Government nor any agency thereof, or any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Further, Oak Ridge National Laboratory is not responsible for the contents of any off-site pages referenced. The complete ORNL disclaimer can be viewed at <https://www.ornl.gov/ornl/contact-us/Security--Privacy-Notice>.