

NGEE Arctic Policy for Data Collection, Processing, Archiving, and Sharing

The open sharing of Next-Generation Ecosystem Experiments (NGEE Arctic) data among project researchers, the broader scientific community, and the public is critical to meeting the scientific goals and objectives of the NGEE Arctic project and critical to advancing the mission of the Department of Energy (DOE), Office of Science, Biological and Environmental (BER) Terrestrial Ecosystem Science (TES) program where the strategic intent is to deliver improved scientific data and models regarding the potential effects of increasing greenhouse gas concentrations on the Earth's terrestrial biosphere and the role that terrestrial ecosystems play in the global carbon cycle.

The NGEE Arctic project is committed to upholding a rigorous and high-quality data management strategy and the implementation of that strategy in an innovative, cost-effective data collection, management, distribution, and archival framework. The goal of this effort will be to implement guidelines and procedures for collecting, tracking, storing, and providing data both within the project and with the larger scientific community. The NGEE Arctic data management policy and plans are consistent with the data policies of the sponsoring TES program.

The NGEE Arctic project seeks to address this challenge by quantifying the physical, chemical, and biological behavior of terrestrial ecosystems in Alaska. Initial research will focus on the highly dynamic landscapes of the North Slope where thaw lakes, drained thaw lake basins, and ice-rich polygonal ground offer distinct land units for investigation and modeling. The project will focus on interactions that drive critical climate feedbacks within these environments through greenhouse gas fluxes and changes in surface energy balance associated with permafrost degradation, and the many processes that arise as a result of these landscape dynamics. The overarching goal of the NGEE Arctic project is to reduce uncertainty in climate prediction through improved representation of Arctic tundra processes. A focus on scaling based on process understanding and geomorphological units will allow us to deliver a process-rich ecosystem model, extending from bedrock to the top of the vegetative canopy, in which the evolution of Arctic ecosystems in a changing climate can be modeled at the scale of a high resolution Earth System Model grid cell (i.e., 30x30 km grid size). This vision includes mechanistic studies in the field and in the laboratory; modeling of critical and interrelated water, nitrogen, carbon, and energy dynamics; and characterization of important interactions from molecular to landscape scales that drive feedbacks to the climate system. A suite of climate-, intermediate- and fine-scale models will be used to guide observations and interpret data, while process studies will serve to initialize state variables in models, provide new algorithms and process parameterizations, and evaluate model performance. The NGEE Arctic project will also develop innovative communication and data management strategies as we work both within a multi-disciplinary team environment and with the larger scientific community to chart a course for an improved process-rich, high-resolution Arctic terrestrial simulation capability.

As stated above, the NGEE Arctic project is expected to generate diverse data sets from observations, experiments, and models across field plot, watershed, regional, and global scales. These data will include automated data collected from weather stations and trace-gas systems, observations from remote-sensing platforms, manual data collection efforts during large campaign-based field work, and discrete data sets generated from chemical, biochemical, and molecular characterizations of soil, water, microbial, and plant samples. Large output files from a suite of fine- to climate-scale models will also be generated within the NGEE Arctic

project. Finally, the project will draw on a wealth of existing data products collected and generated by other national and international monitoring networks and research organizations across the Arctic.

We intend this policy to be a clear statement of the importance of implementing the data life-cycle approach for collection, quality assurance, analysis and modeling results, documentation, archiving, and use of products, and supporting the continuous flow of data and information before, during, and after the data generation activities.

The data collection activities, processing and quality assurance analyses, synthesis of information, model inputs, outputs, and code developed to meet the objectives of NGEA Arctic will be supported by the NGEA Arctic Distributed Data Management System.

The Distributed Data Management System supports a defined workflow for data collection, data and model processing, quality assurance, secure storage, archiving, and project and public access to data products.

The Policy has been expanded to a full Data Management Plan describing the implementation of the Distributed Data Management System as it supports the complete data life cycle

Applicability

This Policy is applicable to all NGEA Arctic participants, including the scientists at Oak Ridge National Laboratory, Los Alamos National Laboratory, Brookhaven National Laboratory, Lawrence Berkeley National Laboratory, University of Alaska Fairbanks, partners at leading universities and other state and federal agencies, cooperating independent researchers, and to the users of NGEA Arctic data products (see the Data Use Recommendations later in this section).

While the focus of this Policy is on sharing, archiving, and use, it is obvious that the other aspects of the data life cycle must be accomplished for archiving and use to be successful; all aspects together are the framework for optimizing data integration, interoperability, and usability. The Policy will be expanded to support the complete data life cycle in a full Data Management Plan.

Sharing, Archiving, and Discovery

The NGEA Arctic Distributed Data Management System includes provisions for long term archiving of data and model products. Depending upon the data type the existing capabilities of the CDIAC, ARM, ESG, and ORNL NASA DAAC will be utilized for long-term data archival.

The discovery (identifying and finding) of NGEA Arctic data sets, derived products, synthesis results, and models (inputs, outputs, and codes) by the scientific community and public will be facilitated through the compilation of descriptive companion metadata records and their inclusion in searchable metadata databases and clearinghouses. Applicable International Organization for Standardization (ISO), Federal Geographic Data Committee (FGDC), Network Common Data Form (NetCDF), and other standards for data and metadata format

and content will be implemented to ensure interoperability with existing data access portals and web services to optimize NGEE Arctic data integration and usability.

This Policy does assign certain responsibilities to Science Area Task Leaders; it is a shared responsibility of all participants to implement the overall data policy. In the Data Management Plan, Task Leaders, task-specific principal investigators, project personnel, and Data Center Staff will be identified and assigned clearly defined roles and responsibilities.

Data Sharing

Timeliness of Data Availability:

NGEE Arctic researchers will actively collect, process, quality assure, and document, environmental measurements, experimental data, observations, and modeling results, and submit them to the Data Archive in a timely manner. The diverse set of planned tasks varies greatly as to data type, spatial and temporal resolution, modeling or measurement, and accordingly, the amount of processing and analysis effort needed to create a given product varies. Initially these data products will be shared among NGEE Arctic participants and subsequently with the public.

To identify a reasonable time for processing and quality assurance of data products while maintaining accountability for submitting data, each Science Area Task Leader will define a schedule for submitting data to the Archive for their given products. Tentative guidance, however, is as follows:

- Generally, three months from collection of field measurements or completion of laboratory analyses is a good target for release of data for sharing within the NGEE Arctic project.
- Data from *in-situ* instruments will be made available to the NGEE Arctic science team within a week.
- For public access, timing will vary with the data type. Generally, public access will be concurrent with completion of initial quality assurance and, in some cases, open literature or web site publication of results and will be rendered consistent with BER data sharing policies. Consider:
 - Data from automated environmental measurements (e.g., weather station) will be available to the NGEE Arctic science team within a week following initial quality assurance.
 - Annual surveys and seasonal measurements (e.g. NPP) will be made available to the public within 6 months of data collection and initial quality assurance.
 - Laboratory analyses of vegetation nutrient concentrations will be posted to a public web site within 6 months of analysis and initial quality assurance.

Quality Assurance of Data:

Data products will be submitted to the Archive, initially for sharing among NGEE Arctic participants, and subsequently for access by the scientific community and public.

The level of quality assurance needed for sharing newly collected, derived, and processed data among NGEE Arctic participants for information and confirmation purposes is typically not as great as that for publication and public access. By defining the quality level of each data type

and product, it will be clear when data products are ready to be shared, and what quality checks have been performed.

Each Task Leader will define the quality assurance checks to be performed before data sharing among NGEA Arctic participants (Quality Level 1) and then before public access (Quality Level 2). When data products have been updated as a result of additional quality checks or the discovery of errors, the data should be resubmitted to the Archive with the quality level documentation changed (e.g., to Level 2).

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 (e.g., raw data, photos, hardcopy data sheets, scanned data sheets, notebooks, etc.). This may, for example, be data from an instrument logger expressed in engineering units or using nominal calibrations, or high resolution data before aggregating to a selected interval. These products should be submitted to the Data Archive for long-term storage but will not be shared routinely nor widely advertised.

Protection of Intellectual Property Rights

NGEA Arctic Science Area Task Leaders will work to ensure that the intellectual property rights of data originators and principal investigators are protected and acknowledged. Task Leaders, in deciding the timing of public data release, will balance the needs of investigators to first publish their results and desirability of expediently releasing data to a wider audience.

Data Use Recommendations

The NGEA Arctic data provided through the NGEA Arctic Data Portal will be freely available and will be furnished by project participants and cooperating independent researchers who encourage their use.

- Users are expected to inform (e-mail is appropriate) the NGEA Arctic scientist(s) of their planned use of the archived data and of any publication plans before use of the data. Contact information will be provided on the Project web site.
- The Data Archive will advise users to check the NGEA Arctic Data Archive frequently to ensure that they are using the latest version of the data.
- Users are expected to acknowledge (1) data products with a citation as provided in the data archive documentation, (2) web site information downloads with a bibliographic web citation, or (3) general NGEA Arctic information as an acknowledgment or personal communication, if no other citation form is applicable.
- When publishing original analyses and results based on NGEA Arctic data, users are expected to acknowledge the agency or organization that supported the collection of the original data (i.e., Department of Energy, Office of Science, Biological and Environmental Research).
- Users are encouraged to include these terms as publication keywords as applicable: NGEA Arctic, Alaska, and Next-Generation Ecosystem Experiments.
- Users are asked to provide an electronic reprint of their independent work to the NGEA

project so that all publications resulting from these data may be tracked, recorded, and referenced by the Carbon Dioxide Information Analysis Center (CDIAC).

Disclaimer of Liability

Data and documents available from the NGEA Arctic web site (<http://ngea.ornl.gov/>) 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 <http://www.ornl.gov/ornlhome/disclaimers.shtml>.