

## NGEE ARCTIC DATA MANAGEMENT PLAN \*

JANUARY 26, 2019

**Data Types and Sources:** The Next-Generation Ecosystem Experiments (NGEE Arctic) project is composed of an array of researchers from multiple science disciplines across multiple national laboratories and universities. The NGEE Arctic project generates diverse datasets from observations, laboratory measurements, experiments, and models across field plot, watershed, regional, and global scales. These data include automated data collected from weather stations and in situ systems, observations from remote-sensing platforms, manual data collection efforts during large campaign-based field work, and discrete datasets 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 by the NGEE Arctic project. The project draws on a wealth of existing data products collected and generated by other national and international monitoring networks and research organizations across the Arctic including the NASA-funded ABoVE field campaign

**Content and Format:** The NGEE Arctic project leverages existing tools and expertise to provide data management support to the project by adopting a standards-based, open-source approach to ensure interoperability with future data systems and other projects. The ORNL Online Metadata Editor (OME) is a Web-based tool that allows users to create and maintain robust metadata stored as eXtensible Markup Language (XML) files, the preferred metadata output format, with output that conforms to and satisfies widely-adopted metadata standards – specifically Federal Geospatial Data Committee (FGDC) and ISO19115 metadata standards. The OME captures information about the specific data contributors, projects, parameters, time periods, quality assurance, and locations associated with the data. This metadata is the basis for the NGEE Arctic Search and Access Tool for data products. The OME also enables the automatic generation of a unique data citation and Digital Object Identifier (DOI) for each dataset. Users may upload data files plus additional documentation using the OME. The preferred non-proprietary file format for public sharing of tabular data products is the comma separated value (csv) format. For geospatial imagery, GeoTIFF and NetCDF are the preferred formats for raster data and shapefiles or KML for vector products.

**Sharing and Preservation:** All NGEE Arctic metadata records are available to the public in the [NGEE Arctic Data Search and Access Tool](#). Metadata may be classified as either Planned, Internal, or Public. Planned datasets include metadata only and are listed in order to promote visibility and communication within the NGEE team. Datasets marked Internal are accessible to authenticated NGEE Arctic team members only, and Public datasets are released without restriction to the public and are shared with the ESS-DIVE data center. Descriptive information captured in the OME enables advanced data search options using keywords, geospatial, temporal, contributor name, and scientific task filters. The NGEE Arctic data will be released publicly under a [Creative Commons Attribution 4.0 license \(CC-BY\)](#) license. The NGEE Arctic portal provides a data sharing policy (available in Appendix), data submission guidance, and data citation recommendations plus guidance on data submissions. Early sharing of data within the project is urged because it encourages vital scientific collaboration within the project, identifies planned research, allows access to the data as soon as possible for researchers and modelers, and ensures long-term preservation. All data and associated metadata produced by NGEE Arctic should be made available as soon as feasible but no later than 12 months after acquisition. All public data will be submitted to the DOE's ESS-DIVE data center. At the end of the project, all internal data and associated metadata will be made public and will be transitioned to ESS-DIVE.

**Protection:** NGEE Arctic will not store personally identifiable or sensitive environmental information in its data system. If any are discovered, it will be removed. Intellectual property rights of investigators are protected by access restrictions and promoted through data citation guidance and DOIs. Stored data are protected from loss by routine and test backup protocols. Cybersecurity planning and infrastructure for the NGEE Arctic data system will be provided as part of the Oak Ridge National Laboratory cyberinfrastructure.

**Rationale:** The open sharing of 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) program. The NGEE Arctic project seeks to understand how surface and subsurface processes and properties are interconnected across permafrost-dominated tundra ecosystems. Ultimately, the NGEE Arctic project is developing a process-rich ecosystem model, extending from the bedrock to the vegetative canopy-atmospheric interface, in which the evolution of Arctic ecosystems can be modeled at the scale of a high-resolution ESM grid cell.