Archiving ABoVE: Insights into ORNL DAAC's Data Management Efforts for the Arctic-Boreal Vulnerability Experiment

Debjani Singh, Tammy Walker, Hannah Blanco, Matthew Donovan, Scott Pearson, Tom Ruggles, Carol Sanderson, Michele Thornton, Yaxing Wei, Jessica Welch, and Bruce E. Wilson,
Oak Ridge National Laboratory Distributed Active Archive Center (https://daac.ornl.gov)

Abstract

The mission of the ORNL DAAC is to assemble and maintain a comprehensive archive of observations and models relevant to research in the fields of terrestrial biogeochemistry and ecological dynamics, and to facilitate research, education, and informed decision making in support of NASA's Earth Science research objectives.

Background

The ORNL DAAC for Biogeochemical Dynamics is a NASA Earth Observing System Data and Information System (EOSDIS) data center managed by the Earth Science Data and Information System (ESDIS) Project and operated by the ORNL.

Archive for ABoVE

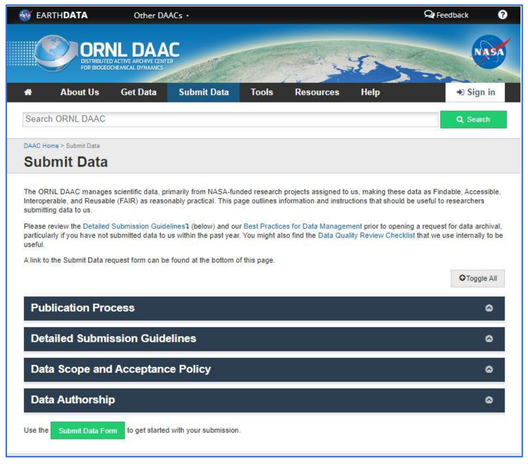
The ORNL DAAC continues to support ABoVE as the campaign archive and as a resource for data publication.

The ORNL DAAC manages scientific data, from NASA-funded ABoVE research projects, making these data as Findable, Accessible, Interoperable, and Reusable (FAIR) as reasonably practical..

190 Published ABoVE Datasets

- 27 Airborne Science
- 18 Carbon Dynamics
- 18 Fire Disturbance
- 42 Hydrology & Permafrost
- 3 Project Standards
- 74 Vegetation
- 8 Wildlife

Data Publication Resources



Submit Data https://daac.ornl.gov/submit

Learn
- Review the *Detailed Submission Guidelines* and *Best Practices for Data Management.*
- Understand ORNL DAAC's *Data Scope and Acceptance Policy* and *Data Versioning and DOI Policy.*
- Review NASA's *Data Product Development Guide for Data Producers.*
- Contact us early at ingestlead@daac.ornl.gov whenever you have question about data publication.

Request
- Double check that your data follows our standards and best practices.
- Complete the *Submit Data Form.* We will review all requests within a few working days. If we have the authority to publish your data, we open a Submission and email you instructions. Where we do not have that authority, we work with you on next steps..

Submit
- Submit your data package, following the emailed instructions. The package includes all data files, as well as relevant supplemental files, code/scripts, papers, and manuscripts.
- Answer the *Data Provider Questions* using the link in the emailed instructions.
- We reserve your DOI after we receive all necessary materials. The DOI will not yet resolve, but you will know what it will be and you can use it preparing manuscripts.

Preprint
- Where needed, we work with you to publish a *Preprint Dataset* to support NASA Open Source Science (OSS) data policies and to support the publication of scientific manuscripts..

QA
- We review your data using the *Data Quality Review Checklist*. Ensuring that your data can be understood and manipulated is an important part of data reusability.
- We will reach out if we find any problems or need more information.

Doc
- Using your documentation and data files, we prepare structured metadata, formalize the dataset citation, and write a comprehensive user guide. If we did not already make the DOI resolve in creating a preprint dataset, we take care of it in this step.

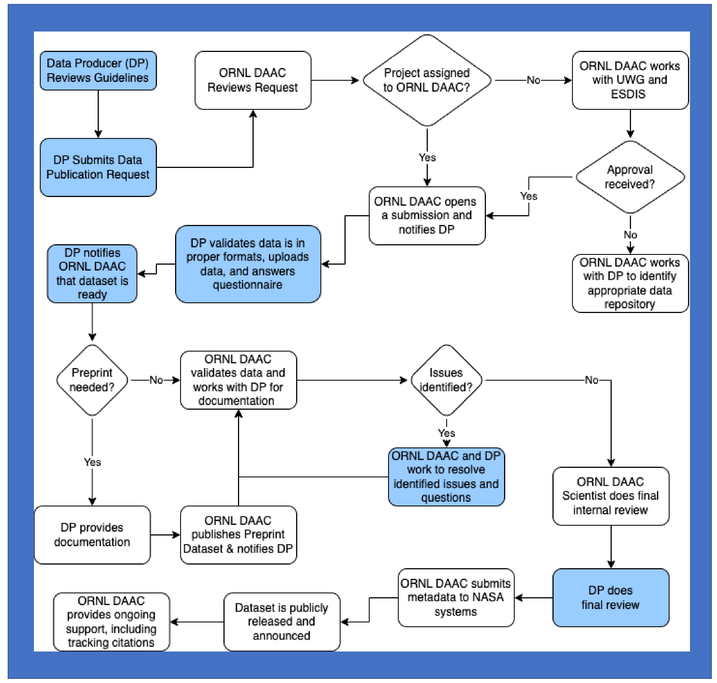
Tools
- We determine which tools for data discovery, subsetting, visualization, and transformation are appropriate for your data and begin the work to get your data into those tools. See the *Detailed Submission Guidelines* for links to learn more about the *Earthdata Level of Service Model* and *Earthdata Data Tools.*

Review
- We email you instructions and a link to review your data package.
- You perform the review and let us know of any corrections.
- We make those corrections, and check back with you if needed.

Publish
- We publish the data package, providing a landing page with links to data and services for your data and pushing metadata to the NASA EOSDIS Common Metadata Repository (CMR). CMR provides metadata to other relevant catalogues, and it is the source for Earthdata Search. We ensure the data package is advertised online through email, social media, and ORNL DAAC news.

Support
- We continue to provide and maintain tools to discover, explore, access, and extract data.
- We ensure long-term availability including multi-site backups and integrity checks.
- We address user questions, serving as a buffer between users and data producers.
- We maintain usage, download, and data citation statistics, which are provided on dataset landing pages.
- We update data formats and documentation as needed to ensure continued reusability.

Publication Process



*Blue Boxes indicate Data Producer (DP) steps.

Publication Metrics

- Published ABoVE data archived at ORNL DAAC have been accessed 90,740 times by a total of 16,322 unique users (identified by IP address) since January 2015.
- 170 publications have cited ABoVE (150 dataset-level and 20 project-level)
- Current: 190 active datasets totaling 27.9 TB
- 73 preprints processed from the ABoVE project since 2015
- 124 unique first authors, 592 total unique authors
- 30+ file formats
- Range of dataset size: 7KB through 17TB
- Range of number of granules in a dataset: 1 through 164 K

Oak Ridge National Laboratory is managed by UT-Battelle for the US Department of Energy