

Utilizing Interdisciplinary Strategies for Next Generation Ecosystem Experiments Tropics Data Organization

Emily Robles¹, Deb Agarwal¹, Danielle Christianson¹, Boris Faybishenko¹, Robinson Negron-Juarez², Gilberto Pastorello¹, and Charuleka Varadharajan³

¹Lawrence Berkeley National Laboratory

²Lawrence Berkely National Laboratory

³Lawrence Berkeley National Lab

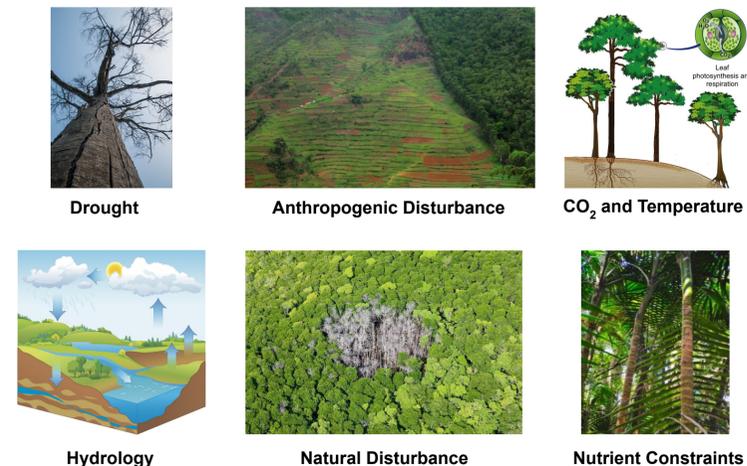
November 22, 2022

Abstract

Quality metadata and data are critical to advancing science and preserving data for long-term use. The Next Generation Ecosystem Experiments (NGEE) Tropics project funded by the U.S. Department of Energy generates and utilizes ecological, hydrological, and meteorological data from tropical forests for scientific analysis and model parameterization. The project's data team manages an archive for users to internally curate and publish data with a digital object identifier (DOI). A key focus of our project is to ensure NGEE Tropics data can be interpreted and utilized by current and future research teams. However, the education and participation of project members to prioritize and be involved in data curation is necessary to reach this goal. We have taken an interdisciplinary approach involving domain and data scientists to create a process that makes it easy for scientists to curate high-quality data packages for archival. First, the NGEE Tropics Archive and metadata reporting templates (FRAMES) were designed using user-experience research methods to incorporate user feedback through interviews and surveys. Upon submission of data packages, thorough checks are performed to ensure quality expectations are met. Each dataset is curated individually, and feedback is provided directly to scientists to identify the optimal data organization for their packages. The data team also provides training to project members using presentations, tutorials, and 1:1 training. As a result of our efforts, package and file-level metadata reporting to the NGEE Tropics archive fits within the existing workflow of scientists, establishing data curation as a core aspect of research. By educating the NGEE Tropics team through integration and communication, we have enabled the production of quality data packages that are findable, accessible and usable by any member of the public. This work will enhance the legacy of NGEE Tropics, and provide a lasting resource for the tropical research community.

OUR PROJECT

The goal of NGENE-Tropics is to develop a predictive understanding of how tropical forest carbon balance and climate system feedbacks will react to changing environmental drivers over the 21st Century. Data collected by NGENE-Tropics researchers offers insight into how tropical forests in Central and South America respond to..



A variety of data types are collected including sapflow, ecohydrological, and meteorological measurements. The data team collaborates with researchers to curate their data packages before approval and publication on the NGENE-Tropics archive.

Dataset ID	Dataset Title	Access Level	Submission Date
NGE001	FRAMES Metadata Reporting Templates for Ecophysiological Observations, version 1.1	Public	2016
NGE002	Hurricane Maria Puerto Rico Landsat Analysis	Public	2018
NGE004	El Verde Ridge, El Verde Valley, and Rio Ticoos root phosphorus and bacterial community composition (December 2015)	Public	2016
NGE004	Leaf gas exchange survey by leaf age, Feb2017, PA-SLZ: Panama	NGEE Tropics	2018
NGE005	Seven years (2008-2014) of meteorological observations plus a synthetic El Niño drought for BCI Panama	NGEE Tropics	2018
NGE004	Leaf mass area, Feb2016-May2016, PA-SLZ, PA-PNM, PA-BCI: Panama	NGEE Tropics	2017
NGE008	Diurnal leaf gas exchange survey, Feb2016-May2016, PA-SLZ, PA-PNM: Panama	NGEE Tropics	2017
NGE008	Leaf sample detail, Feb2016-May2016, PA-SLZ, PA-PNM, PA-BCI: Panama	NGEE Tropics	2017
NGE008	Leaf water potential, Feb2016-May2016, PA-SLZ, PA-PNM, PA-BCI: Panama	NGEE Tropics	2017
NGE004	CO2 response (AC) gas exchange, calculated Vmax & Jmax parameters, Feb2016-May2016, PA-SLZ, PA-PNM: Panama	NGEE Tropics	2017
NGE004	Xylem vulnerability curves of canopy branches of mature trees from Cusuco and Tapir National Forests, Para, Brazil	Public	2017
NGE002	Leaf Pressure Volume Data in Cusuco and Tapir National Forest, Para, Brazil (2011)	Public	2017

GOALS

1. Design an easy to use data archive with a streamlined submission process
2. Define standards for reporting file and package level metadata
3. Promote researcher engagement in data curation
4. Improve the quality, longevity, and reproducibility of NGENE-Tropics data

THE SUBMISSION AND APPROVAL PROCESS

Through the data archive, users can internally curate and publish data with a digital object identifier (DOI). Package level metadata is easily collected and organized using our data submission form, which is visible to archive users before downloading any data.



1:1 meetings and project wide presentations are used to train scientists on the elements of quality data packages and include topics such as..

- Package organization
- Submission steps and demonstrations
- Data package and file examples

Example

Format as you would a publication title Include type of sample, date, and location if related to publication, ideally format "Title of paper, Journal: Data"

Leaf sample details, leaf traits by age, Feb2017, PA-SLZ: Panama

Dataset ID: NGE0045
Dataset Version: 1.0
Dataset Name: Leaf sample details, leaf traits by age, Feb2017, PA-SLZ: Panama
Dataset Description: Details of leaves sampled for leaf water potential and gas exchange measurements from PA-SLZ, February 2017. Sunlit canopy leaves of APEIME, GUATDU, MICOBO, TERMMAM, VIROSP and VOCHFE species were collected before dawn and around midday. Data for each sample includes species, age, sample number and a photograph. This data was collected as part of the 2017 BNL-STRI leaf traits by age campaign. See related datasets (existing and future) for leaf water potential, leaf spectra, sap flow, LMA, gas exchange and leaf chemistry.
PA-SLZ: Bosque Protector San Lorenzo (Fort Sherman)
Dataset Field Site(s): PA-SLZ: Bosque Protector San Lorenzo (Fort Sherman)
Dataset Author(s): Kim Ely <kely@lbl.gov>, Alistair Rogers <arogers@lbl.gov>, Shawn Serbin <sserbin@lbl.gov>, Jin Wu <jinwu@lbl.gov>, Brett Wolfe <bwolfe@gmail.com>
Dataset Originating Institution(s): Brookhaven National Laboratory; STRI

Sites are chosen from provided list*

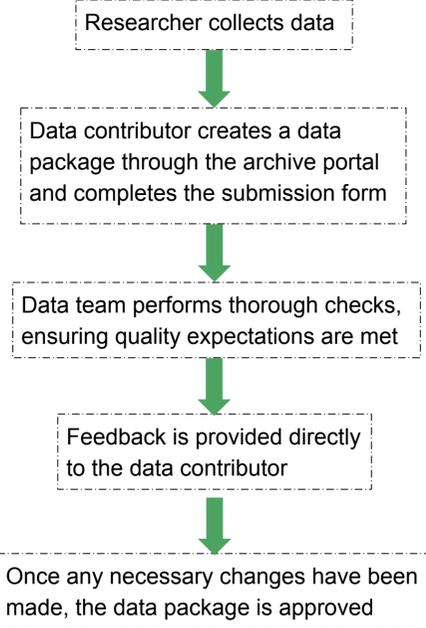
Author order will be the same as citation order

Spell out originating institution(s)

Description should formatted similarly to an abstract: Concise description of purpose and content of data package

Adequate information to decide if data is useful to a user Complete sentences, correct grammar

PUBLICATION WORKFLOW

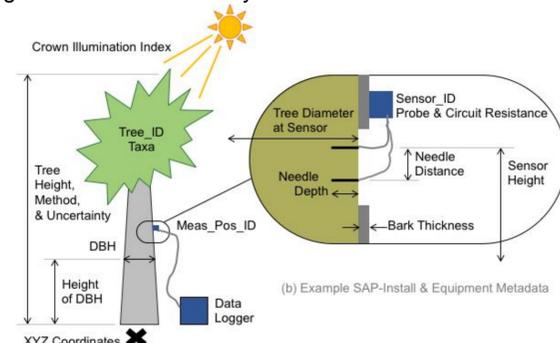


A COMMUNITY CENTERED APPROACH

Interdisciplinary group work and community outreach were utilized to meet our main objectives.

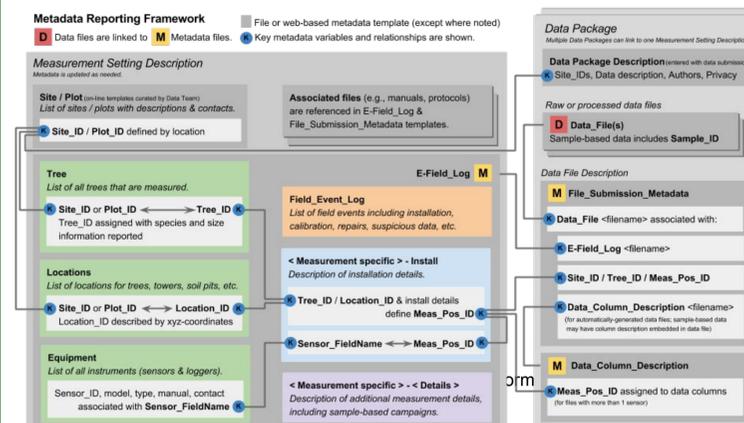
FILE LEVEL METADATA

To record file level metadata, the NGENE Tropics Archive and metadata reporting templates (FRAMES) were designed using user-experience research methods to incorporate user feedback through interviews and surveys.



Danielle Christianson, et al. (2017). "A metadata reporting framework (FRAMES) for synthesis of ecohydrological observations." Ecological Informatics Volume 42, November 2017, Pages 148-158.

The resulting three Excel and online templates describe the measurement setting, data collection, and data file organization. This standardization enables cross-site comparison for different sensor types in various formats, QA/QC, and processing levels.



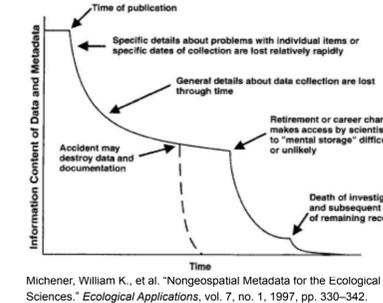
PACKAGE LEVEL METADATA

Package level metadata for each dataset is reviewed using a series of quality checks. These expectations align with community agreed standards, including those implemented by data repositories and scientific journals, and are tailored specifically for NGENE-Tropics.

A focus on community input ensured that these standards fit within the existing workflows of researchers.

ONGOING OBJECTIVES

1. Create data packages with sufficient metadata for reuse by researchers to answer multiple scientific questions
2. Maximize the longevity of NGENE-Tropics data to increase its impact
3. Increase awareness and prioritization of data package quality through educational opportunities for research teams



The scientific community benefits from your data, and **"No data set is perfect and self explanatory"** without complete metadata to accompany it.

Don't let your data die with you!

IMPACT

The NGENE-Tropics archive currently has..

107 total data packages, 45 of which are publicly available, and 172 unique users. Data packages on the archive have been downloaded 1316 times.

A focus on generating high quality metadata as part of creating the NGENE-Tropics data legacy will benefit the tropical research community for decades into the future.

ACKNOWLEDGMENTS

Danielle Christianson, et al. (2017). "A metadata reporting framework (FRAMES) for synthesis of ecohydrological observations." Ecological Informatics Volume 42, November 2017, Pages 148-158. Accessed at <http://dx.doi.org/10.15486/ngt/1419956>.

Kim Ely, Alistair Rogers, Shawn Serbin, Jin Wu, Brett Wolfe(2019). Leaf sample details, leaf traits by age, Feb2017, PA-SLZ: Panama. NGENE Tropics Data Collection. Accessed at <http://dx.doi.org/10.15486/ngt/1508122>.

Michener, William K., et al. "Nongeospatial Metadata for the Ecological Sciences." Ecological Applications, vol. 7, no. 1, 1997, pp. 330-342. JSTOR, www.jstor.org/stable/2269427.

This work was supported by the U.S. Department of Energy, Office of Biological and Environmental Research, as a part of the NGENE-Tropics