NEON Data Product Concept and Production Plan

NEON data from observations and experiments will be processed in sophisticated ways that make it accessible to the widest range of users. Data processing proceeds through a number of defined levels. Initial processing includes conversion from instrument output (Level 0 data) into calibrated biological units (Level 1). Subsequent processing steps range from spatial registration of airborne data, averaging or smoothing to meaningful time steps up to processing through complex models for inference of parameters or unobserved quantities (Level 4 data). Without advanced data processing—the transformation of data to information—NEON data would be usable only by specialists and the network's goals could not be met. Development of advanced data products will increase the accessibility of NEON data for interdisciplinary research and education. NEON data will be distributed as distinct data packages with Digital Object Identifiers to allow their citation. The use of bibliographic identifiers for data sets increases the discoverability of NEON data and allows tracking of use via citations.

An example may help illustrate the planned NEON data flow. The Fundamental Instrument Unit measures a wide variety of water, carbon, and energy budget components at high temporal frequency. However, observations may be missing when instruments fail or conditions do not meet required assumptions, causing data gaps. In addition, many quantities desired by ecologists must be computed or inferred from the measurements made. Data processing steps for biophysical data typically involve:

1. Calculation of basic physical quantities from measurements (for example, carbon flux is computed from CO_2 concentration and wind data).
2. Calculation of key quantities during data gaps using statistical models to allow calculation of temporal averages of integrals (e.g, monthly or annual totals).
3. Calculation of desired quantities from measured quantities. For example, evaporation and transpiration can from be inferred from daily variation in the measurement (their sum, the latent heat flux), and photosynthesis and respiration may similarly be inferred from the measurement (their sum or Net Ecosystem Exchange).
4. Inference of parametric information from calculated quantities. For example, once photosynthesis and transpiration have been estimated, Water Use Efficiency may be computed. Using respiration and temperature, the temperature dependence (Q_{10}) of respiration may be computed.
5. Estimation of process controls from observed, calculated, and parametric quantities. For example, the effect of drought on water use efficiency between statistically normal and dry years can be estimated from interannual variations in precipitation, soil moisture, transpiration, and photosynthesis.

Analogous processing flows will be followed for other data types. For example, population estimates will be computed from small mammal data using mark-recapture models. Population parameters (fecundity, mortality, etc) will be calculated from population estimates. Different sorts of NEON users will be interested in different levels of data. A research biophysicist might be interested in data from the earlier steps in the example given above, whereas an agricultural economist might need the sensitivity estimated in step five for inclusion in a forecast of climate change impacts on the farm sector.

All NEON information will be associated with metadata, describing the characteristics of the measurement (location, time stamp, instrument ID, calibration information, etc), tracking uncertainty, and documenting provenance for derived data products (primary measurements used, algorithm ID and version, etc). NEON metadata will be based on the EML metadata language, with extensions developed to accommodate NEON requirements (for example, currently EML has limited capability for tracking provenance of complex derived products). The Use of EML will encourage interoperability of NEON with LTER and US federal NGDC-compliant data sets.

NEON Data Policy

NEON data will be accessible to the community under an open data policy that emphasizes the free and open exchange of scientific information. All data, whether from core sites, relocatable or mobile deployment systems, or experimental programs will be securely archived within the NEON cyberinfrastructure. The basic NEON data streams from the FIU, FSU, LUAP, AOP, and experiments, and all data taken as part of instrument calibration, engineering, or software testing will be open-access and available as soon as possible once basic QA/QC checks have been applied. Data archives shall include easily accessible information about data holdings, high-level data products, including quality assessments, supporting relevant information, and guidance for locating and obtaining the data. Metadata associated with NEON data will meet or exceed national requirements and be closely associated with the archived data. Data will be made publicly available to users through an open-access web interface. NEON will enforce a principle of non-discriminatory access so that all users will be treated equally. Open-access means that data will be freely-distributed without charge; there may be charges for the cost of reproduction and delivery when access is not web-based. Data distribution will be unrestricted (unless for reasons of national security, protection of endangered species and other legal restrictions). NEON encourages investigators working in experimental set-aside areas to make their research data freely and openly available as soon as possible and will require data sets so collected to be registered with the NEON data archive. NEON will permit restricted access for a reasonable period not to ever exceed two years for data collected by investigators using PI instruments connected to NEON cyberinfrastructure or through PI-requested deployments of NEON facilities. At the end of a restricted access period, all data will be freely and openly available.