

## **Biological Data Submission Guidelines**

Contributors: Beverly Law and members of the AmeriFlux, Fluxnet, and NACP communities

05/10/2007

Updated 3/9/2009

This document explains the biological variables and gives guidance for reporting these data. The biological data augments the flux-met data.

Biological variables may be irregularly, periodically, or regularly. This document addresses the irregular or periodic sampling. Regularly sampled data should be submitted following the regional network submissions guidelines for the flux meteorology and micrometeorology data.

For example, soil CO<sub>2</sub> efflux (Rs) may be measured periodically with a LI6400 or automatically with a LI8100, or home-made system with LI820 or LI6262 gas analyzer. The LI6400 measurements are reported with the BADM protocols with the accompanying soil temperature (Ts) measurements. The LI8100 measurements are reported with the flux meteorology protocol; that reporting is outside the scope of this document.

Raw files from specific instruments should also be submitted to the network archive. The format for these submissions is still TBD. The variables and instruments are:

- LAI. Raw files from an LAI-2000 must include timestamps.
- ACi. Raw data files from the LI6400 must include timestamps.
- PSN\_LT. Raw photosynthetic light response curve data from the LI6400 combined from multiple samples and dates.

### ***Data submission mechanisms***

The irregularly sampled biological data can be submitted in one of three ways:

- The “Ancillary Data Update” form at <http://www.fluxdata.org>. This link is protected – only Fluxnet Principal Investigators or their designated site ancillary or biological data submitters can access this site.
- A “BADM template” composed of four Excel spreadsheets. Multiple spreadsheets are used to handle the richness and variation in the data and the associated metadata describing both the data and data collection practices. Additional documents describing the actual Excel spreadsheets can be found at <http://www.fluxdata.org/>; the spreadsheet in Excel 2003 format or Excel 2007 format can also be downloaded from the website.
- A columnar spreadsheet. When submitting a large amount of data, a columnar spreadsheet may be used instead of the template. The spreadsheet must meet specific criteria listed on the next page.

Completed template or columnar spreadsheets should be submitted to the appropriate regional network; please also send a copy to [fluxdata-support@fluxdata.org](mailto:fluxdata-support@fluxdata.org). If the site is not affiliated with a regional network, the spreadsheet can be mailed directly to [fluxdata-support@fluxdata.org](mailto:fluxdata-support@fluxdata.org).

A copy of the most recent cleaned and curated data can be downloaded from the <http://www.fluxdata.org> site. These “CADM” files contain all data submitted by web form, template, and columnar spreadsheet. These files are also available to the regional networks. Additional cross-site summaries are also available on <http://www.fluxdata.org>.

Using a web form or BADM template spreadsheet can be cumbersome when reporting large amounts of data. This data can be submitted in a columnar format similar to the flux meteorology and micrometeorology data (the flux-met format). A sample of spreadsheet in Excel 2003 format can be downloaded as a starting point.

Submitted columnar spreadsheets must:

- Include the same header giving the site ID, site name, submitter email contact, and creation date.
- Use the same variable names as the BADM template. For example, LAI<n>, Rs<n>\_MEAN or Ts<n>.
- Include any associated parameters such as measurement date, measurement depth, or comment. For example, soil CO<sub>2</sub> efflux would include RS<n>\_MEAN, Rs<n>\_SIGMA, RS<n>\_HOUR, RS<n>\_DATE, and Rs<n>\_COMMENT and all LAI measurements would include LAI<n>\_SIGMA, LAI<n>\_DATE, LAI<n>\_CLUMP, LAI<n>\_TECHNIQUE, and LAI<n>\_COMMENT.
- Dates should follow the DOY/YYYY format.

### ***Biological data groupings***

#### *Chemistry of Soil, Forest Floor and Live Foliage.*

These variables include soil parameters such as percentage of silt, clay and sand, soil bulk density as well as soil, root, litter, and foliage carbon and nitrogen content.

These variables should be measured at least once per year.

#### *Phenology*

These variables capture the dates for budbreak, cotyledon formation, flowering, leaf full, leaf senescence and leaf fall. The data are to be reported by species.

#### *Plot Condition*

These variables are general characteristics of the site. Three different site descriptions give the history, fetch, and general overall characteristics of the site. Other variables

include the stand age and land ownership. These variables tend to be reported once and updated only infrequently.

Plot condition variables also include disturbance such as fire, harvest, storm, or fertilizer application. Each disturbance is characterized by a controlled vocabulary descriptor, a descriptor qualifier, and a date which may be approximate. Each disturbance is reported separately and should be reported no later than when the associated flux-met data are submitted to the regional network. If you can't find an appropriate term in that vocabulary, please contact [fluxdata-support@fluxdata.org](mailto:fluxdata-support@fluxdata.org) for help.

### *Productivity*

These variables characterize the annual changes in site foliage, wood, harvest, litter, and root productivity. The above ground productivity is partitioned into crops, shrubs (foliage and wood), non-woody plants such as grasses or forbs, and trees (overstory and wood). Coarse and fine roots are reported separately; the depth to which the roots are measured is included. Wood radial increments (tree rings) are also reported.

The net ecosystem production computed from the biological data is also reported. This is an aggregate over some number of years. For example, in 2000, NEP might be computed over 1995-2000.

### *Soil Respiration*

These variables include the soil water content, soil temperature, and soil carbon flux. These data are reported as biological data only when measured from time to time and not continuously.

### *Tree Physiology*

Only the sapflow measurement methodology is reported as part of the biological data. The sapflow data are reported in the flux-met format. The raw data files (ACi) and the raw photosynthetic light response curve (PSN\_LT) data from the LI6400 are reported directly to the regional network..

### *Vegetation Condition*

These variables characterize the changes in the site biomass, woody debris or crop residue, canopy height, leaf area index and species characteristics.

The site IGBP class is also reported.

Note that the controlled vocabulary used for the vegetative classification is a subset of the full IGBP set; the subset was developed in collaboration with CarboEurope and omits terms that tend to be imprecise such as SNO. Using the IGBP classifications provides a

uniform classification scheme aimed at facilitating synthesis and modeling efforts and allows for simple comparison with MODIS and other remote sensing products.

### *Data submission guidelines*

Biological variables are measured:

- Rarely in response to changes in the site. Examples include SITE\_DESC (site description) and IGBP (vegetative cover) and any DIST<n> disturbance such as wildfire.
- Annually. Examples include CR\_BIOMASS (coarse root biomass) and HEIGHT\_C (canopy height).
- Multiple times during the year. Examples include LIT\_MASS<n> (litter mass) and ROOT\_N<n> (root nitrogen concentration).
- Across species. Examples include SPP\_O<n>\_PERC (overstory species percent) or WOOD<n> (woody tissue carbon concentration), and LEAFFULL\_SPP (leaf full date).
- At different depths. Examples include SOIL\_N (soil nitrogen measured across a profile) or SOIL\_WATER\_CAP (soil water capacity measured at a single depth). Variables which are measured at depth also include an optional associated horizon.
- Across plots at a site. A plot variability error expressed as a standard deviation is associated with all but the rarely measured variables. If the plot variability is estimated rather than computed, the associated \_COMMENT field should indicate the estimation method.

Variables which can be reported at measured at different time in a year or across different species are indicated by an “<n>” in the variable name.

A few examples will illustrate this.

- The soil composition variables CLAY\_PERC, SAND\_PERC, and SILT\_PERC are measured annually. There is no <n> in the variable name and the actual measurement date need not be submitted.
- LAI can be measured multiple times during the growing season. A single LAI<n> measurement includes the actual measurement, the date of that measurement, and the < LAI<n>\_CLUMP, and LAI<n>\_TECHNIQUE.
- Disturbances such as fires are reported only when the event occurs. There may be no, one, or multiple events during a year. For example, a crop site may be tilled, planted, fertilized, suffer an insect infestation, and harvested. Each of those is reported as a distinct DIST<N> with the associated time and disturbance-specific qualifier.

The spreadsheet template has four different sheets:

- The SiteBioAncData and SiteBioMetadata sheets can be filled out once and then only updated and resubmitted when the site characteristics or methodology change. These sheets can be blank otherwise.
- The DisturbanceData sheet can hold multiple disturbances across multiple years or a new sheet may be used for each year. Like the SiteBioAncData and SiteBioMetadata sheets, the DisturbanceData sheet can be left blank when there are no new disturbances.
- A new BioData sheet should be used each year. For example, new data obtained in 2007 should not be added to the spreadsheet previously submitted in 2006; instead a new spreadsheet should be used.

It is important to report the measurement date of many of the biological variables. In the spreadsheet template, the data format depends on the variable.

- For some variables, no date is necessary. The year of the data submission is assumed to be the year the measurement was obtained.
- For disturbance data, the spreadsheet template date format is DD/MM/YYYY or (day/month/year). For historical site disturbance data where the exact day or even year may be unknown, the day (if possible) and year can be approximated; the DIST\_DATE\_QUAL is used to indicate the approximation
- The preferred spreadsheet template date format is DOY/YYYY; DD/MM/YYYY can also be used. For Rs and Ts, the Rs<n>\_HOUR row is used to report the hour and minutes in HHMM format just like the flux-met data submissions.

The web form accepts either DD/MM/YYYY or DOY/YYYY for any date. Note that it is important not to submit dates as MM/DD/YYYY or MM/YYYY. These will be misinterpreted as DD/MM/YYYY or DOY/YYYY.

Only actual measurements need to be reported. It is not necessary to submit data with “not available”, “not appropriate”, “NA”, “N/A”, “unknown”, “none”, “no” or similar text string. When using the spreadsheet template, cells should be left empty. If cells are filled with -9999, they will be interpreted as empty.

Each measurement can be associated with a COMMENT. The comment can be used for any additional text describing the measurement or measurement qualifiers such as species or to cite a published paper, web site, or personal communication. Please do not embed comments in the spreadsheet - any such comments will be omitted by the processing.

All numerical variables have defined units. The reported numbers will be assumed to be in those units. Numerical values should be reported with a “.” as the decimal point and no “,” separators. If there is a “,”, it will be interpreted as a “.”; entries with multiple “,” will be automatically rejected both by the web form and the template spreadsheet processing. In other words, a value of “1234.567” is valid while values of “1,234.567” or “1,234,567” are not valid and a value of “1234,567” will be interpreted as “1234.567”.

Numerical variables can be reported as ranges using the separator “-“. In other words, a value of “123-456” is valid while the values of “123 to 456” or “123\_456” or “about 123 to 456” are not. Invalid values will be automatically rejected both by the web form and the template spreadsheet processing.

Species are specified using the standardized codes in the Natural Resource Conservation Service (NRCS) PLANTS database January 2000 version (<http://plants.usda.gov>). Identification to species only is expected, but if the subspecies is known, please enter the appropriate NRCS code. If a plant cannot be identified confidently, assign a NRCS PLANTS genus or unknown code appropriate to the species. The scientific name or common name can also be provided in the associated comment field. If the measurement spans all species at the site, ‘(All)’ should be used.

Some text variables such as IGBP, SPP, or DIST have controlled vocabularies. Only text strings that belong to those vocabularies will be available in the web form. The template spreadsheet implements a pull down menu box in the appropriate cells. If you can’t find an appropriate term in that vocabulary, please contact [fluxdata-support@fluxdata.org](mailto:fluxdata-support@fluxdata.org) for help.