

Interactive comment on “The AmeriFlux data activity and data system: an evolving collection of data management techniques, tools, products and services” by T. A. Boden et al.

T. A. Boden et al.

bodenta@ornl.gov

Received and published: 24 April 2013

First of all, thank you for your thorough review of our paper. We very much appreciate your comments and suggestions. The following addresses our response to your comments following the numbering sequence in your review.

1) Before submission of the final version of the paper we will look for opportunities to shorten text without compromising content and necessary detail. We are bound to the publishing guidelines and formats imposed by GI. We agree wholeheartedly that Figure 7 (i.e., the data flow diagram) would be more powerful if shown earlier in the manuscript rather than at the end.

C20

2) Again, we are bound to the publishing guidelines and formats imposed by GI. Whether we can include a "Personal Experiences" section in the final manuscript will be determined largely by GI's willingness to deviate from the prescribed format. We have sprinkled in a variety of personal experience statements (e.g., off-the-shelf commercial software packages typically do not satisfy the requirements of diverse environmental data systems without customization) throughout the manuscript.

With respect to your question regarding effort, we presently devote ~ 0.5 person-years/yr to embellishing and maintaining the AmeriFlux database and user interface. Your fundamental point is an excellent one. Long-term dbs and interfaces require constant care and improvement, especially as usage increases and expands. This happened in the case of AmeriFlux. The original network was a handful of sites and the original database system architecture reflected this smaller size. In our case, we simply outgrew the original architecture and design requiring a migration to a newer architecture. Of course, the migration takes time and resources too.

Our preferred mode of receiving data is for sites to submit files using a secure, dedicated file transfer system at our laboratory. Files are quarantined and cyber-checked properly before transfer to our own servers and thus from a cybersecurity standpoint is preferable. Mirroring scripts and utilities ease automation but the bottom line is we simply want to get data. Larger irritants for us are file inconsistencies and poor data quality resulting from a fundamental lack of attention on the part of measurement teams.

Metadata review, conveying QA results to PIs for resolution, and reviewing QA plots require manual attention. Initial file transfers and creation of processing reports are easily automated. If we could start over we would advocate standardized reporting requirements (format, content, units) and promote steps to improve data quality (e.g., dedicated data technicians at the sites, replicate measurements where cost effective, remote monitoring of instruments to reduce down time, etc.).

C21

3) AmeriFlux has helped document and understand interannual variability in ecosystem productivity under managed and natural scenarios. AmeriFlux has also advanced fundamental process-level understanding (e.g., factors driving and influencing ecosystem respiration) in a variety of ecosystem types.

4) Site teams are responsible for maintaining their instruments, including calibration. AmeriFlux deploys a roving system to assist sites identify potential problems but the primary responsibility rests with the sites. The measurement teams are also responsible for maintaining records of their calibration efforts but this varies greatly across the network. AmeriFlux would be wise to adopt rigid calibration and reporting requirements like the DOE Atmospheric Radiation Measurement Program where this information is integrated with the data quality evaluations and long-term data archive. Some AmeriFlux measurements are traceable to NIST-like standards but not all.

5) Multiple flags are used in the FLUXNET level 3 data products including flags to identify spikes, low variability, and marginal points (e.g., records immediately preceding a long gap).

6) Yes, level 4 products also include flags indicating whether a record is original or gap-filled and, if gap-filled, an indicator of the reliability of the gap-filled record. Error estimates are not currently provided in higher-level AmeriFlux/FLUXNET data products but we plan to include error estimates in future CDIAC-produced AmeriFlux data products.

7) Our web-based data interface allows access to our MySQL database, however, users are limited to the suite of queries we offer through this interface. Users are not provided open access to the database. Edits and changes to the MySQL database are done internally. Downloading and visualization capabilities exist in our present AmeriFlux data interface but other remote/synced operations are not - interesting suggestion.

8) We use multiple missing value representations (e.g., -9999 represents a missing value while -6999 indicates the measurement was never made) and all are numeric. For many of our processing efforts (e.g., SAS), alphanumeric or character representa-

C22

tions are an annoyance in generating summary statistics easily.

9) We determine the "reasonable values" based on our expertise and experience and on existing datasets, such as earlier years from the same site or other sites within the same climate zone and ecosystem type.

10) In designing the gap-filling methods for the AmeriFlux meteorological records, we considered the applicability of each method to gap sizes. For example, the mean diurnal cycle method is generally limited to a gap of 15 days or less. Weather station data are used for gaps upwards of 30 days and NARR data are used to fill gaps exceeding one month. If a procedure fails because of a larger gap, this gap is unfilled at the current step and will be tried again in the next step. Therefore, with the current design, a gap larger than a year will be left open. In other situations, gaps spanning several months could also be unfilled if the measurement records are very short (1-2 month) in that year because the linear relationship between measurements and NARR data can be very poor and unreliable in this case.

11) Good suggestion, will make this addition in the final version of the manuscript.

12) Good catch, we will correct this typo in the final version of the manuscript.

Interactive comment on Geosci. Instrum. Method. Data Syst. Discuss., 3, 59, 2013.

C23