Please see the following manuscript for detailed methods used to harmonize the datasets in LAGOS-NE; in particular, many of the specific methods and procedures are found in the online ‘Additional Files’ available at: https://gigascience.biomedcentral.com/articles/10.1186/s13742-015-0067-4

Soranno, P.A., E.G. Bissell, K.S. Cheruvelil, S.T. Christel, S.M. Collins, C.E. Fergus, C.T. Filstrup, J.F. Lapierre, N.R. Lottig, S.K. Oliver, C.E. Scott, N.J. Smith, S. Stopyak, S. Yuan, M.T. Bremigan, J.A. Downing, C. Gries, E.N. Henry, N.K. Skaff, E.H. Stanley, C.A. Stow, P.-N. Tan, T. Wagner, K.E. Webster. 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science and data reuse. GigaScience 4:28 doi:10.1186/s13742-015-0067-4

Detailed information on data sources are found in “Additional File 5” in Soranno et al. (2015). Briefly, the data source for lakes and streams in the 17 state area was the NHD (http://nhd.usgs.gov). The hydrologic boundaries (i.e., for three of the spatial classifications, HUC12, HUC8, HUC4) came from the Watershed Boundary Dataset (WBD; http://nhd.usgs.gov/wbd.html). In addition, we used the digital raster dataset of elevation for watershed delineation from the National Elevation Dataset (http://ned.usgs.gov/). All download dates for these data sources are provided in “Additional File 5” in the above citation.

All methods to create this module are described in Soranno et al. (2015). The most challenging and time-consuming part of building this module was connecting the sampling locations from the lake water quality datasets (which each contained different types of unique identifiers, and sometimes only lake names) to a georeferenced location in the NHD. When data providers included the lake latitude and longitude, we were able to mostly automate the procedure. Nevertheless, even when coordinates were available, there were many cases where the latitude and longitude did not intersect the NHD lake polygon boundary, requiring manual interpretation. We also had to devote significant resources to developing automated procedures for creating lake watersheds, which are all described in Soranno et al. (2015).