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). Almost all data sources for this module are from national-scale datasets and thus use standardized methods throughout the study extent.

All methods to create this module are described in “Additional files 5, 7, 8, 13, and 14” in Soranno et al. (2015). Briefly, we calculated the metrics for this module by developing project-specific GIS tools in the ArcGIS environment, which are referred to as the LAGOS GIS Toolbox (and made available here: https://github.com/soranno/LAGOS\_GIS\_Toolbox). The toolbox outputs multiple individual data tables of calculated values organized by the above three data themes that are then imported into LAGOS-NE-GEO for different spatial classifications, including values calculated at the level of the individual lake, 100 m and 500 m buffers around each lake, the lake IWS, states and counties, hydrologic units, and ecological drainage units (an ecoregion spatial classification). The unique identifiers for this data module are the zone ID’s for each spatial classification for which we calculate these metrics. In other words, we calculate land use around a lake in each of the zones of the many spatial classifications in LAGOS-NE. However, the data are exported into individual tables by spatial classification. Therefore, there are different numbers of rows in each table; for example, there are 51,101 rows for the land use metrics calculated for the 100 m lake buffer because there are 51,101 lakes that have a 100 m buffer area, but only 17 rows for the land use metrics calculated for the state spatial classification.