

## Keeping it clean: Creating and maintaining high quality fish data for the U.S. EPA National Rivers and Streams Assessment

Blocksom, K.A. and Peck, D.V.

One primary biological indicator of condition used in the National Rivers and Streams Assessment (NRSA) is the fish assemblage. Data for the 2008-2009 assessment were collected on field forms from over 2100 sites. After field forms were scanned into the NRSA database, we developed a complete taxa list while simultaneously reconciling names from field forms to AFS accepted names. Vouchers were sometimes used to complete positive identification of species collected. Data validation is an iterative process, and once an initial analysis dataset was created, data analysts identified additional corrections to assure accurate and complete data. Corrections were relayed back through the data gatekeeper, a single point of contact able to update the database directly. Original and updated values are maintained in the database in order to track data changes and allow reversion back to the data on a specific date if necessary. All datasets were distributed through this data gatekeeper to users to ensure data analysis teams were using the latest version of datasets. This approach is critical to maintaining clean, reliable data for all users and providing important documentation of changes from original to final data. This is an abstract and does not necessarily reflect EPA policy.