Data quality in fishery science

Laurent Dubroca,Norbert Billet,Joel Vigneau,Alice Vastel,Nicolas Goascoz,Anne-Sophie Cornou,
Sébastien Démanèche,Jérome Weiss,Mathieu Merzereaud,Alastair Pout,Liz Clark,Mike Armstrong,
Chun Chen,Kelig Mahe,Maria Hansson,Ana Ribeiro Santos,Eva Velasco,Christoph Stransky

In fishery science, data compilation is the cornerstone on which the accuracy and the precision of stock assessments are built. Data compilation is the step dedicated to the preparation of the information related to a stock. It covers the computation of the species composition of the landings, effort, discards and their associated length or age structures estimated from declarative statistics and sampling data. A very first step of data compilation is the verification and validation of the data, which includes data quality checks in a broad sense. During the fishPi projects (DG-MARE 2014/19 WP4) and the sister-project related to the Mediterranean fishery (DG-MARE 2014/19 WP4), we designed R tools to make this crucial step objective. The implemented procedure audits the data structure, verifies the sampling protocol implementation and carries out some advanced data checks based on outlier detection. An automated reporting system provides information on the quality status of the data for the end-user. An easy import and export procedure allows the accurate identification and correction of errors using generic spreadsheet software. These tools are versatile (ie able to tackle other format definitions with few modifications), open source and already available. The procedure is illustrated with the quality checking of a dataset related to the sole fishery in European waters.

Contact author:
Laurent Dubroca
laurent.dubroca@ifremer.fr
Datacall Response Unit (CREDO)
IFREMER, Avenue du Général de Gaulle, 14520, Port-en-Bessin-Huppain, France