ICES data at our fingertips: an overview of new and upcoming technologies

Colin Millar, Arni Magnusson, Scott Large, and Carlos Pinto

Bridging the gap between data and analysis

Traditionally, the task of getting and preprocessing scientific data has been quite separate from the subsequent data analysis. This can make it challenging to trace where the data came from and how they were preprocessed. It can also be difficult to repeat the analysis with a slightly different dataset.

In this paper, we introduce new technologies and pathways to read ICES data directly into the R software platform, where the data analysis takes place. This benefits scientists, both when working with scripted workflows and when exploring data interactively.

R packages and web services

The ICES Secretariat has recently developed R packages that provide functions to read data directly from ICES databases via web services. The packages are available on CRAN, the central repository maintained by the R Core Team.

Four packages have been released so far, to access trawl survey data, stock assessment results, stock database entries, and general reference codes.

Diagram here

icesDatras

icesSAG

icesSD

icesVocab

Underlying web services

The pathway from the ICES databases to individual R sessions uses a technology called web services. When a data request is submitted as a special web address (URL) the web service returns the data in a standard machine-readable format. The data are then converted (parsed) to create an R data frame.

The R functions perform these tasks behind the scences, but the user is provided with a simple and efficient interface to specify the data to retrieve.

Reproducible research

Access to additional ICES databases is likely to follow in the future.

[Perhaps mention that some timestamp mechanism would be required to achieve fully reproducible research. When a previous analysis is rerun it should give the same result as before, even if the contents of the underlying database have changed since the original analysis.]

Blah blah (R Core Team, 1999; Alonso *et al.*, 2004; Richardson and Ruby, 2008; Marcial and Hemminger, 2010).

References

Alonso, G., Casati, F., Kuno, H., and Machiraju, V. 2004. Web services. *In* Web services, pp. 123–149. Springer.

Marcial, L. H., and Hemminger, B. M. 2010. Scientific data repositories on the web: An initial survey. Journal of the Association for Information Science and Technology, 61: 2029–2048. Wiley Online Library.

R Core Team. 1999. Writing R extensions. R Foundation for Statistical Computing. Vienna.

Richardson, L., and Ruby, S. 2008. RESTful web services. O'Reilly Media, Inc.