History of the CPR database

D. Stevens and P.C. Reid

Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, The Hoe, Plymouth, PL1 2PB, United Kingdom

E-mail: dpst@mail.pml.ac.uk

The Sir Alister Hardy Foundation for Ocean Science (SAHFOS) is an international charity responsible for managing the Continuous Plankton Recorder survey in the North Atlantic and North Sea, providing a unique multi-decadal data set of plankton abundance since 1931. Subsequent to 1931 technology has advanced and the system for storing the CPR data has developed considerably. The CPR database has been affected by performance related factors such as processor speed and disk capacity as well as economic factors such as the cost of software. These issues have been overcome and the system for storing and retrieving the data has become more user friendly at every development stage.

Prior to 1969 calculations for sample positions, the time of sample and the quality control of CPR data were carried out manually. Processing to calculate year to year changes in abundance, seasonal cycles and geographical distributions of the CPR data was prepared by hand, and in 1969 the first steps were taken to automate these tasks. The CPR data were entered into a computer using punch cards and processing routines developed (Colebrook, 1975). When punch cards became obsolete, data were stored in text files and data processing was done using programs developed by CPR staff. The difficulty with this system was the reliance on "in-house" experience and skills, and there were problems interrogating the raw data files. To address this problem the data were transferred to a relational database and then to the Access database which SAHFOS uses today.

The fundamental principles behind the original data entry system and processing have not changed. For example at each development stage the same algorithms have been used to calculate the sample position and the local time. This ensures that the CPR database remains consistent, and historical information can be confidently compared with present day data. Though some automatic quality control processes have been developed over the years, CPR data is still scrutinised personally by senior analysts to maintain the standard of the data, as a computer cannot replace their specialist knowledge.

The computerised database currently contains data from 1946 to 2000 consisting of 181,262 CPR samples with 2,135,072 taxonomic abundance entries. In 2001 there were 25 separate data requests from 11 countries, highlighting the usefulness and interest there is in the CPR database.

The emphasis now is on the need to develop a system allowing users easy and immediate access to available data, and this is achieved via the World Wide Web. Phytoplankton colour and *Calanus finmarchicus* data are already available on the web for CPR standard areas. This will be developed allowing researchers to input their own geographical areas. This paper/poster is to show the developments and changes in the CPR database, highlight SAHFOS plans for the future, and emphasise the need to keep up to date with current developments.