Continuous Plankton Recorder database: evolution, current uses and future directions

Darren Stevens1,*, Anthony J. Richardson1,2,3, Philip C. Reid1

1Sir Alister Hardy Foundation for Ocean Science (SAHFOS), The Laboratory, Citadel Hill, The Hoe, Plymouth PL1 2PB, UK
2CSIRO Marine Research, Cleveland, Queensland 4163, Australia
3Department of Applied Mathematics, University of Queensland, St Lucia, Queensland 4072, Australia

ABSTRACT: The Continuous Plankton Recorder (CPR) survey, operated by the Sir Alister Hardy Foundation for Ocean Science (SAHFOS), is the largest plankton monitoring programme in the world and has spanned >70 yr. The dataset contains information from ~200 000 samples, with over 2.3 million records of individual taxa. Here we outline the evolution of the CPR database through changes in technology, and how this has increased data access. Recent high-impact publications and the expanded role of CPR data in marine management demonstrate the usefulness of the dataset. We argue that solely supplying data to the research community is not sufficient in the current research climate; to promote wider use, additional tools need to be developed to provide visual representation and summary statistics. We outline 2 software visualisation tools, SAHFOS WinCPR and the digital CPR Atlas, which provide access to CPR data for both researchers and non-plankton specialists. We also describe future directions of the database, data policy and the development of visualisation tools. We believe that the approach at SAHFOS to increase data accessibility and provide new visualisation tools has enhanced awareness of the data and led to the financial security of the organisation; it also provides a good model of how long-term monitoring programmes can evolve to help secure their future.

KEY WORDS: Data accessibility · Visualisation tools · CPR data

INTRODUCTION

Large data repositories are common within the physical, chemical and biological oceanographic community. It is generally accepted that such data should be shared openly and freely, providing a wealth of valuable information for researchers, ecosystem modellers, policy makers and the general public. However, data are not always easily accessible, and the volume of diverse information can be difficult to integrate and synthesise (Vanden Berghe et al. 2004). This requires organisations to continually make data more accessible and provide simple yet evocative visual representations of data in easily digestible forms. Here we describe these challenges as they pertain to the largest plankton monitoring programme in the world, the Continuous Plankton Recorder (CPR) survey.

The CPR survey provides a long-term baseline of the near-surface distribution, abundance and diversity of phyto- and zooplankton. This has been used to assess biodiversity (Beaugrand et al. 2000, Beaugrand & Ibañez 2002), especially in terms of impacts of climate change (Beaugrand et al. 2002, Richardson & Schoeman 2004), over-fishing (Reid et al. 2000), pollution (Batten et al. 1998), eutrophication (Edwards et al. 2001a) and the spread of invasive species (Edwards et al. 2001b). The survey was initiated in 1931 and has been operated since 1991 by an international charity, the Sir Alister Hardy Foundation for Ocean Science (SAHFOS).

The CPR itself is a high-speed plankton sampler towed monthly behind commercially operated ships of opportunity. Since its inauguration the survey has operated in the North Sea and, since 1939, in the North
Atlantic, with a break during World War II. Operations started in the North Pacific in 1997. Water enters the CPR and flows down a tunnel through a silk filtering mesh. Upon return to the laboratory the silk is cut into 10 nautical mile sections that equate to samples (for more information see Warner & Hays 1994, Richardson et al. 2006).

In this contribution we describe the CPR database and the approach taken at SAHFOS over the years to increase data accessibility and provide new visualisation tools. We first describe the evolution of the CPR database and access issues. We argue that as the data have become more accessible, both to researchers within and outside SAHFOS, not only is more high-quality research being conducted, but the CPR data are being used more directly in marine ecosystem management. We then describe 2 recent visualisation tools, the WinCPR North Sea plankton browser and the digital CPR Atlas, which make CPR data available to both researchers and non-plankton specialists. Enhanced awareness and use of CPR data have contributed to the financial security of the organisation, and provide a model of how other long-term monitoring programmes can evolve to help secure their future.

CPR DATA

Since September 1931, the CPR has been towed >5 million nautical miles, with 196 120 samples counted (see Fig. 3). This equates to >2.3 million records of individual taxa being present or ~90 million data points, including zero records. Samples are analysed for plankton abundance under a microscope in 3 stages: (1) phytoplankton, (2) small zooplankton <2 mm counted in a traverse across the silk and (3) zooplankton >2 mm that are removed from the silk and counted under low magnification. More than 450 different taxa have been identified, over half to species level (Reid et al. 2003). A description of the methods of the CPR survey, along with a list of all taxa and relevant information for each taxon, is given in Richardson et al. (2006).

In addition to plankton data, concurrent environmental data, such as temperature, chlorophyll and salinity, are measured. Environmental data are collected on approximately half of the CPR routes; this will hopefully be expanded to all CPR deployments in the future, as funding allows. There is also extensive auxiliary information on the attributes of each tow and sample. In terms of the tow, the name and average speed of the ship, the latitude and longitude of deployment, retrieval and course changes during the passage of the ship, the identification numbers of both the CPR and the interchangeable internal mechanism and the propeller angle of the CPR are recorded. For each sample, the location and local time at the midpoint of the course is calculated, and the name of the person who counted the sample is documented.

STORAGE AND ACCESS

Over the last 70 yr, the storage methods for CPR data have changed as technology has advanced (Stevens & Reid 2004), allowing improved data access (Fig. 1). Prior to 1969, CPR data were stored on cards and large maps, with data in >1 format (e.g. by tow and by species). As all calculations were carried out by hand, analysis of the data was time consuming and limited to basic operations. In 1969, the first computerised database containing CPR data was developed on a KDF9 computer. Initially this database only stored processed data (monthly means for standard areas), but no raw data, and access by researchers was limited because the database was housed offsite. By the early 1970s, a database to store raw data from newly analysed samples was developed, and the CPR team took steps to ensure that historical data were entered retrospectively. Data were now accessible faster than before, but the employment of a programming specialist was necessary to extract data from the file-based database. The database has evolved considerably over the years; initially it was accessed using ALGOL (ALGORithmic Language), then by IMP (Implementation Language), after that concurrently by PASCAL and FORTRAN (IV) G, and then by FORTRAN 77 running on an IBM OS/2 platform. In 1993 the first relational database for CPR data was developed in ORACLE, but this was never fully adopted because of financial constraints.

In 1995, the CPR database was transposed into the ACCESS relational database still in use today. This gave researchers within SAHFOS easier access to raw data. However, calculations for sample position and time were still processed by the old FORTRAN system. A further difficulty in gaining access to the data was the limited number of staff members with the skills required. Almost all research at this time was carried out by those directly involved with the survey, as data were not easily available to external researchers.

By the end of the 20th century a significant change in the philosophy at SAHFOS concerning data accessibility had evolved. In May 1999, SAHFOS amended its data policy to comply with the emerging Global Ocean Observing System (GOOS) programme, making the data freely available for non-profit research (Reid et al. 2003). Since then, the CPR survey has formed part of the Initial Observing Programme of GOOS. As part of our commitment to GOOS, data on important indicators of primary (phytoplankton colour) and secondary...
(Calanus finmarchicus) productivity are freely available directly from the SAHFOS website (www.sahfos.org). Data are available as pre-processed monthly means for CPR standard areas (pre-defined areas used historically within the survey). At present, no information is collected on who is accessing these data or the number of times they are being downloaded; this should be redressed in the future.

Currently, researchers requiring any data other than phytoplankton colour and Calanus finmarchicus abundance in CPR standard areas need to complete, sign and return a data licensing agreement available from the SAHFOS website. A dedicated full-time database manager has processed these data requests since August 1999, enabling us to monitor the increase in data requests (Fig. 2). We believe this is linked to the increased scientific profile of SAHFOS, enhancing awareness of CPR data, and the more open data policy, as well as the investment in computer hardware and software improving accessibility. Since this time, there have been >100 requests for CPR data from 14 different countries across 3 continents (Canada, France, Germany, Iceland, Italy, Netherlands, Norway, Portugal, Republic of Ireland, Spain, South Africa, UK and USA), emphasising the research and management value of the archive.

In 2001 the FORTRAN programs were rewritten in Visual BASIC for ACCESS, because of concerns about future compatibility. In 2002, temperature data along selected CPR routes since 1996 were made available via the SAHFOS website.

During 2004, in collaboration with the Ocean Biogeographic Information System (OBIS) project of the Census of Marine Life (CoML), CPR plankton presence data have been made available via the World Wide Web (www.iobis.org). Consequently, via OBIS, CPR data are available through the Global Biodiversity Information Facility (GBIF). The CPR dataset supplies more records to OBIS than any other provider and is thus also a significant provider to GBIF. OBIS allows users to draw simple distribution maps for nearly 40 000 marine species, from sponges to whales. Making CPR data available to the global community...

---

**Fig. 1.** Changes in storage and accessibility of Continuous Plankton Recorder (CPR) data through time. GOOS: Global Ocean Observing System; PCI: phytoplankton colour index; SAHFOS: Sir Alister Hardy Foundation for Ocean Science
Plankton data accessible via such large international data portals will further enhance the awareness, accessibility and use of the dataset.

The policy currently adopted by SAHFOS provides data as monthly and annual means. Researchers requiring access to raw sample data need to visit SAHFOS in Plymouth (UK) to obtain the data. This allows the researcher to witness the process of counting samples and become familiar with the idiosyncrasies of the CPR methodology and data.

**PRODUCTS**

Recent efforts have focused on making CPR data more available by developing software to allow users easier access and increased flexibility of data interrogation. Two such products have been developed, i.e. SAHFOS WinCPR and the digital CPR Atlas. These products provide oceanic researchers with graphical outputs that aid data interpretation, increasing the accessibility of CPR data to a wider audience. Both WinCPR and the CPR Atlas were originally programmed in MATLAB, to overcome the problem of spatial and temporal biases in the sampling. The aim was to provide simple access to CPR data in a summary form, but this was hindered by the sophistication and cost of the software. Therefore, user-friendly, WINDOWS-compatible front ends to the browsers were built.

WinCPR is a gridded database browser of North Sea plankton, containing data from a 50 yr period (1948 to 1997). It targets not only the marine science community, but a wider audience, including the general public and students from schools and universities (Vezzulli & Reid 2003, Vezzulli et al. 2004). The user-friendly information and buttons on the opening page are visually appealing, clear and allow the user to perform sophisticated analysis quickly. The software allows users to produce output summarising changes in monthly and interannual abundance of plankton taxa. The grid consists of 172 pixels centred on 1° longitude x 0.5° latitude. A total of 110 plankton taxa, as well as phytoplankton colour (an estimate of phytoplankton biomass), have been averaged for each month within a 40 or 50 yr time span (50 yr for zooplankton and phytoplankton colour, 1948 to 1997; 40 yr for phytoplankton, 1958 to 1997). Products available include distribution maps averaged annually or over the entire period, graphs of abundance through time and seasonal cycles, and month-by-year contour plots for individual and groups of pixels. Underlying gridded data can also be exported; this will be particularly useful for ecosystem modellers in validation and initialisation. The software is available for download via the SAHFOS website (www.sahfos.org/winCPR.htm).

The other major product recently developed to increase data accessibility is the digital CPR Atlas. The first hardcopy CPR Atlas on spatial distribution of plankton in the North Atlantic was published in 1973, and was based on only ~40 000 CPR samples from 1958 to 1968 (see Fig. 3 for number of samples per grid square). This atlas contributed to our knowledge of the biogeography of ~260 taxa.

Work was started on updating the atlas in 2001, for the celebration of the 70th anniversary of the CPR survey, and was published as a supplement of the Marine Ecology Progress Series in 2004 (CPR Survey Team 2004). This new hardcopy atlas is based on >150 000 CPR samples, collected from 1958 to 1999. Fig. 3 shows the distribution and number of samples in the North Atlantic, with highest numbers in the NE Atlantic and North Sea. A statistically robust procedure was developed to reduce bias associated with the irregular sampling through time of the CPR survey (Beaugrand 2004). Maps were produced using Lambert conical projection rather than the Mercator projection used in the 1973 edition, removing the problem of distortion for large areas away from the Equator. This atlas will be a powerful tool for researchers, an invaluable tool for para-taxonomists needing to know whether a species is found in a certain area, but will also help to define important biological regions for international biodiversity initiatives such as the CoML.

The new digital CPR Atlas is based on the hardcopy atlas released in 2004, but includes several powerful added features. It not only allows users to view distribution maps of the North Atlantic for the entire period (1958 to 1999), but also enables maps to be drawn for each decade, and for day and night periods. Decadal maps will help document the spread of non-indigenous species and provide a critical baseline to assess climate impacts on plankton. For example, it is clear from the...
decadal maps produced that a number of subtropical species have moved further north over the last 4 decades as sea temperatures have warmed (also see Beaugrand et al. 2002). Maps showing day–night differences highlight species that undergo extensive diel vertical migration (also see Beaugrand et al. 2001). The digital CPR Atlas will become available in 2006.

Fig. 3. Number and distribution of samples throughout the North Atlantic and North Sea for the CPR Atlas in: (a) 1958 (1973 version) and (b) 1958 to 1999 (2004 version; note: scale here is logarithmic base 10)

CPR data have underpinned high-quality research for >70 yr, and the growth in resulting publications has continued to increase (Fig. 4). A comprehensive bibli-
ography of publications relating to the CPR survey is available on the SAHFOS website, managed by the National Marine Biological Library, Plymouth, UK.

In the early years, CPR data were used to describe and document general plankton ecology in terms of the timing of phyto- and zooplankton blooms (Colebrook & Robinson 1961) and regional differences in productivity (Colebrook 1964). At this time data were used almost exclusively by researchers within the CPR survey. More recently, the long time series has been used as a baseline against which global change can be assessed. Much of the research has focused on impacts of our changing climate on biodiversity (Beaugrand et al. 2000, 2002), plankton productivity (Richardson & Schoeman 2004), timing of seasonal cycles and synchrony between successive trophic levels (Edwards & Richardson 2004) and fisheries productivity (Beaugrand et al. 2003, Beaugrand & Reid 2003). Other impacts of global change assessed include over-fishing (Reid et al. 2000), pollution (Batten et al. 1998), eutrophication (Edwards et al. 2001a) and the spread of invasive species (Edwards et al. 2001b). With greater data accessibility over recent years, researchers not affiliated with SAHFOS have been increasingly using CPR data. This is highlighted by the growth in external data requests since records began in 1995 (Fig. 2) and a number of recent high-profile publications driven externally to SAHFOS (e.g. DeYoung et al. 2004, Thompson et al. 2004).

**Marine management**

Information from the CPR survey has been used extensively to support marine management in the areas of fisheries and, environmental protection, and in the study of ecosystem response to environmental change (Brander et al. 2003). In fact, the survey was initiated to better manage the herring fishery in the North Sea by reducing its variability and improving its efficiency (Reid et al. 2003). In contemporary times, the survey has focused on documenting changes in ecological indicators that are sensitive to alterations in ecosystem health. CPR data were used in the Quality Status Report of the North Sea (e.g. North Sea Task Force, NSTF 1993) and in an annual Ecological Status Report of the North Atlantic (Edwards et al. 2004; see www.sahfos.org). Specific procedures have also been designed and implemented for monitoring climate-driven changes in copepod biodiversity (Beaugrand 2004, Beaugrand & Ibañez 2004). Other indicators sensitive to environmental change that have been developed include the presence and location of harmful algal blooms, the relative dominance of mero- and holozooplankton, unusual range extensions and the dominance of warm- or cold-water taxa. Such indicators contribute to assessments of the health of European waters by the UK Department of Environment, Food and Rural Affairs (Reid et al. 2004), the UK Joint Nature Conservation Committee (e.g. Edwards & John 1998), the European Environment Agency, the International Council for the Exploration of the Sea and OSPAR (Reid 1999; Quality Status Report 2000).

**THE FUTURE**

SAHFOS is actively enhancing methods of data storage and developing new ways of making CPR data more accessible.

**WinCPR**

We have acquired funding to further develop the WinCPR software through 2 future expansions. The first of these is to extend the dataset for the North Sea up to the present, so the database will include a further 6 yr of CPR data. There will also be an improved method of gridding, similar to the one used to create the CPR Atlas. The size of each pixel will be changed in an effort to reduce artefacts of the gridding process. This development will also include other environmental variables, such as sea-surface temperature (SST), cloud cover and wind speed, so that relationships between the environment and plankton can be assessed. The second phase of development is to extend the geographical coverage of the browser to the NE Atlantic.
Upgrade of the CPR database

The size of the CPR dataset makes data extraction sluggish under the current Microsoft ACCESS system. Investigations are underway to upgrade the system, with advice from the British Oceanographic Data Centre and the University of Plymouth. The new computerised database will include further variables such as the microscope number (important to assess whether there are inter-microscope differences), the height of the tow point, the length of the tow wire, location of stored sample in the sample archive, whether a sample was collected during the day or night, a link to a taxonomic manual that is in preparation (containing information on each species) and CPR data, currently in paper form, from 1931 to 1938. The new database will also allow improved quality control by providing information on the distribution of each taxon, to highlight potential irregularities, and by comparing samples with others taken along the same route. Easy access to the CPR samples within the archive via the database will allow taxonomic (Lindley 1982), genetic (Kirby & Reid 2001) and pollution (e.g. plastics, Thompson et al. 2004) research to be conducted retrospectively.

Database developers are presently taking a fresh look at the way CPR data are managed, from the moment the samples are returned to the laboratory to the point when standard analysis is complete and samples are archived. The aim of this project is not only to make the data more secure and easier to manage, but also to decrease the time between the arrival of samples at the laboratory and the release of data for use by researchers. CPR data will be entered directly into the computerised database, rather than entering plankton abundance counts on paper first, as is the current procedure. Another development will be the release of the phytoplankton colour index data earlier than the plankton abundance data. The new system will allow for phytoplankton colour data to be released within 6 to 8 wk of collection, in contrast to the current situation in which all plankton data are released in September of the following year.

Environmental data

Environmental information should be derived for each CPR sample and stored in the database. Data on SST, cloudiness and winds are available from the International Comprehensive Ocean–Atmosphere Data Set (www.cdc.noaa.gov/coads/) on a 1° monthly scale and dating back to 1860. Temperature, salinity and fluorescence data collected via CPRs will also be incorporated.

CONCLUSIONS

The CPR dataset is one of the most valuable marine biological surveys in existence. The CPR database has evolved with changes in technology, increasing data access to researchers within and outside SAHFOS. The organisation has found that it is imperative to plan for software development and budget for further design to ensure that the product will reach the widest audience. We have seen, over the last 10 yr, that efficient access to data has resulted in its increased use, in many high-impact publications within and outside the organisation, and in expanding the role it plays in marine management, all of which have raised awareness of the dataset. Monitoring the number of people accessing data and their affiliations provides a measure of the
value of the dataset and helps justify continued financial support of the organisation into the future. SAHFOS has found that quick and easy access to CPR data will ensure the continued survival of the survey.

With the wealth of data now available on the web, it is easy for researchers to become overwhelmed. We have shown that to promote wider use, additional tools need to be developed to provide representations that are visually appealing. Two such software tools developed at SAHFOS, WinCPR and the digital CPR Atlas, allow easy access to CPR data for both researchers and non-plankton specialists. A critical feature of these tools is that they provide researchers with the ability to generate summary statistics instantly about the dataset, in order to refine their data queries. We believe that the approach at SAHFOS to increase data accessibility and provide new visualisation tools has enhanced awareness of the data and led to the financial security of the organisation; it also provides a good model of how long-term monitoring programmes can evolve to help secure their future.

The next challenge is linking the CPR database with other larger environmental and biological datasets. The role of distributed database systems, such as OBIS, will allow the integration of data from different sources providing greater spatial and temporal resolution. Data providers will be able to utilise the centralised tools provided by such systems, reducing duplication of development. Such initiatives will undoubtedly enable rapid data access, allowing more effective planning and targeted research. They will also help to ensure the continued survival of long-term monitoring programmes by developing new ways of making data accessible, informative and useful by a broad cross-section of the research community, policy makers and the general public.

Acknowledgements. The authors are grateful to all past and present members and supporters of SAHFOS and its predecessors; their sustained help has allowed the establishment and maintenance of the long-term CPR dataset. Currently, the main support for SAHFOS is from the Department of Environment, Food and Rural Affairs (UK) and the Natural Environment Research Council (UK). Consortium support for the CPR survey is also provided by agencies from the following countries: UK, USA, Canada, Faroe Islands, France, Iceland, Ireland, Netherlands, Portugal, Spain, and by the International Oceanographic Commission and the European Union. We thank the owners, masters and crews of the ships that tow CPRs on a voluntary basis. Our appreciation goes to the anonymous reviewers for their constructive comments on our paper.

LITERATURE CITED


Edwards M, John AWG, Johns DG, Reid PC (2001a) Case-history and persistence of the non-indigenous diatom

Editorial responsibility: Howard I. Browman (Associate Editor-in-Chief), Storebø, Norway

Submitted: December 1, 2004; Accepted: December 12, 2005
Proofs received from author(s): June 19, 2006