A new infrastructure for marine biology in Europe

Marine biodiversity informatics
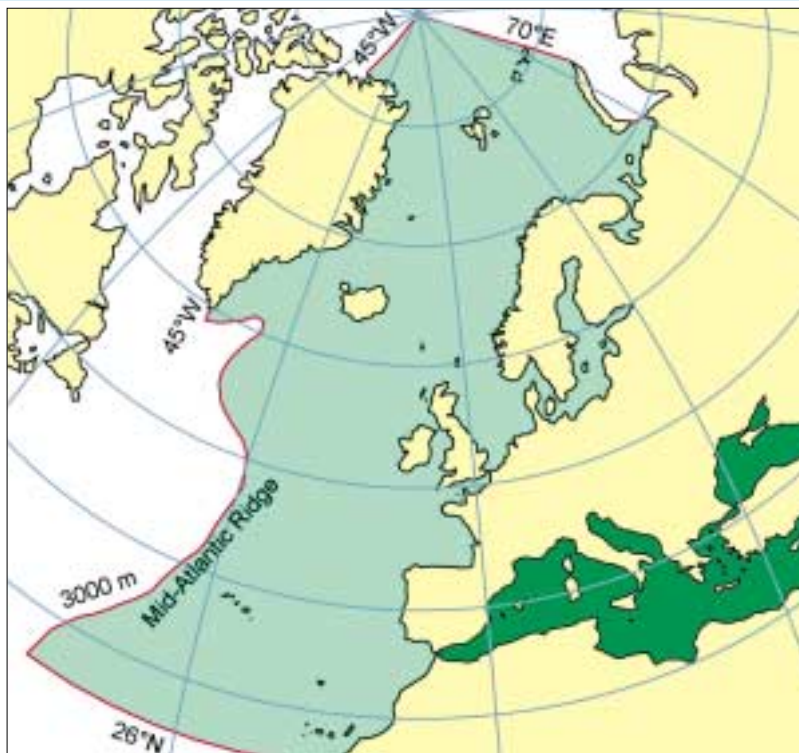

Fig 1. The sea area included in the European Register of Marine Species.

By Mark Costello

THE ADVENT OF COMPUTERS AND the internet has provided new opportunities to manage marine biological data, including data storage, analysis, archival, publication and exchange. This "marine biodiversity informatics" has parallels in non-marine areas of biodiversity, and gains added-value through interaction with non-biological marine data. This article outlines how the European Register of Marine Species (ERMS) is being updated with MARBEF, and its relationships to (a) other European initiatives, namely the Society for the Management of European Biodiversity Data (SMEBD), Fauna Europaea, Euro+Med PlantBase, Species 2000 Europa (EuroCAT), BioCASE, and (b) global initiatives, including the Ocean Biogeographic Information System, Global Biodiversity Information Facility, International Ocean Data Information Exchange (IODE) of the Intergovernmental Oceanographic Commission.

European Register of Marine Species

ERMS was created by a project funded by the European Commission entitled "A register of marine species in Europe to facilitate marine biodiversity research and management" (Contract No. MAS3-CT97-0146) and co-ordinated by Ecological Consultancy Services Ltd (EcoServe), Dublin, Ireland (www.ecoserve.ie). It involved 22 partner organisations, 170 participating scientists, communication with 42 organisations and had a budget of €385,000 euro over two years (1998-1999) (Costello 2000). It had a clear data management plan and developed a novel Intellectual Property Rights Agreement with each contributor that kept ownership of the results within the control of the scientific community that created the project results. This gave copyright of ERMS to a new society in which all contributors are life members without subscription fees. This, the Society for the Management of European Biodiversity Data, is incorporated as a limited company in Ireland, and managed by a council elected by the society members (www.ecoserve.ie/smebd/).

The outcomes of ERMS included: a book listing almost 30,000 species (Costello *et al*, 2001) in European seas (Figure 1); a website providing results; a register of 600 experts (in 37 countries) in European marine species identification; a bibliography of 840 identification guides; an analysis of gaps in identification expertise and guides, knowledge of species groups, and marine species collections; a new scientific society for the

long-term management of biodiversity data (intellectual property); and provided a model and foundation for future projects (e.g. Fauna Europaea, BIOMARE, MARBEF). The original website (erms.biol.soton.ac.uk) (ERMS 1.0) has now been replaced by the species register served from a relational database (ERMS 1.1, same taxonomy, different structure) hosted by the Flanders Marine Institute (VLIZ) (www.vliz.be/vmdcdata/erms) and project results and subsequent activities at www.ecoserve.ie/smebd/. Following a request from the European Environment Agency's EU Nature Information system (EUNIS) the register of experts has been expanded to all marine biologists active in Europe as part of the BIOMARE (www.biomareweb.org) and MARBEF (www.marbef.org) projects.

ERMS 2.0

Within the recently-funded MARBEF project, ERMS will be transformed from a static online searchable publication into a dynamic system. Because a standard species authority file is a critical infrastructure for MARBEF, this forms part of the data management component of MARBEF.

SMEBD has authorised VLIZ to manage ERMS and its updating within MARBEF. The updating will be led by Mark Costello and include Chris Emblow, Geoff Boxshall (who leads the MARBEF taxonomy activities) and Edward Vanden Berghe (responsible for MARBEF data management). The taxonomy experts who produced ERMS 1.0 have been contacted and are ready to begin updating ERMS, including correcting errors and omissions, especially trying to fill recognised

Society for the Management of European Biodiversity Data

SMEBD was founded as a product of the European Register of Marine Species project in 2000. However, its title and aims are not limited to marine biodiversity data. It was recognised that issues of biodiversity data management, both cultural and technical, were similar in terrestrial and freshwater systems. Some species migrate between marine and freshwater systems, and many taxonomists study a taxon across environments. The ERMS project and SMEBD Council include scientists involved in non-marine biodiversity data management, and in the future SMEBD may act as a trust to oversee additional data sets.

SMEBD has no funds or staff, but has a constitution and is annually audited to comply with Irish company laws. EcoServe provides its postal address and manages its website (www.ecoserve.ie/smebd/). A report on its recent activities, including involvement in three new EU projects, is available on the SMEBD website. Any revisions and expansions of ERMS, and its use and loan to other organisations, is authorised by the SMEBD Council.

It is time to elect a new Council (existing members can be re-elected) and nominations are invited.

Approved editors of species lists will be able to access ERMS 2.0 online, such that the register will be changing according to need and editorial diligence. ERMS 1.1 will remain online as a static archive and, at intervals, future versions will be similarly archived so researchers are able to go back and compare previous editions. Copies could also be made to CD for depositing in libraries if required.

What will be new in ERMS 2.0?

ERMS 2.0 will be more than an updated nomenclature of ERMS 1.0. Already, with ERMS 1.1 in a relational database, the calculation of statistics from the register can be automated (www.vliz.be/vmdcdata/erms/stats.php). The present phylogenetic hierarchy for ERMS will be reviewed in relation to hierarchies in use by Fauna Europaea, ITIS, Species 2000 and others to derive a system that is as correct, practical and stable as possible. This hierarchy will allow users to select groups of species at different taxonomic levels, and provide a standard for the calculation of phylogenetic indices of biodiversity that are being researched by ecologists. Because the ocean has so many more phyla than terrestrial systems, it is of special interest to explore these concepts and relationships in marine biodiversity.

Fig 2. A diagram produced by Species 2000 in 1998 to illustrate the plan for a complete list of all species in Europe.

gaps in ERMS 1.0. In addition, additional experts are being approached to increase direct involvement of regional experts.

Management and quality control

This new dynamic ERMS will require a different and more permanent management structure than previously. What is proposed, and is open to discussion, is that a key group of coordinators act as Editors, and an Editorial Board of 'Associate Editors' of taxonomists take responsibility for their areas of expertise. This terminology is used because ERMS 2.0 will be dynamic, more like a journal than a book. The Associate Editors would thus be responsible for quality control and advising the editors of technical and other issues that should be addressed. They would ask external reviewers and specialists in small numbers of species (e.g. a family or genus) to review and/or provide correct synonymies as necessary. A system of rating lists according to their taxonomic certainty and geographic coverage will also be developed. Costello *et al* (2001) included such a rating in the version of ERMS published as a book.

Species 2000 Europa (EuroCAT)

Species 2000 is an initiative that aims to create an online Catalogue of Life (CoL) on earth. Its philosophy has been a distributed system whereby data are searched from different databases around the world through a single portal (web page with a search interface). In partnership with a more centralised US initiative, the Integrated Taxonomic Information System (www.itis.org), Species 2000 is providing the CoL to the Global Biodiversity Information Facility (www.gbif.org). During the ERMS project, a group of scientists prepared plans for the creation of a register

of species' names for all animals and plants in Europe (Figure 2). All these initiatives have been funded, and new projects are addressing additional technical and scientific challenges.

The global Species 2000 (www.sp2000.org) initiative has obtained EU funding for a project, "EuroCat," that will create an on-line "EuroHub" that will connect to the three main species databases for Europe, namely ERMS, Fauna Europaea (www.faunaeur.org/), and Euro+MedPlantBase (www.euromed.org.uk/). Species 2000 Europa is coordinated by Frank Bisby and Fauna Europaea by Wouter Los, both members of the original ERMS project Steering committee and the SMEBD Council. SMEBD has signed an agreement with the project to collaborate, with Edward Vanden Berghe (VLIZ) as the SMEBD representative. While the technical details of how this will be done remain to be finalised, it would seem that interoperability between online databases is now relatively easy to implement, using data exchange protocols such as DiGIR (Distributed Generic Data Retrieval, digir.sourceforge.net/) (used by OBIS and GBIF) and ABCD developed by BIOCASE (www.biocase.org), and may also be adopted by GBIF.

Connecting regional and global species lists

How to interconnect regional registers of species names, such as ERMS, to global species databases (GSD), and how to merge local databases within a region, are two critical challenges for biodiversity informatics. Ideally, there would be a single authoritative file of species names, to which alternative but incorrect names of species (misspellings, synonyms) are linked. Then a search would always direct the user to the correct species name, and show the alternatives. In Europe, ERMS is this species

European examples of species information sources include:

- Marine Life Information Network for Britain and Ireland:- www.MarLIN.ac.uk;
- Seaweeds and algae:- www.algaebase.org/
- Check List of European Marine Mollusca:- www.mnhn.fr/mnhn/bimm/clemam/
- Phoronids:- emig.free.fr/Phoronida/PhoEng.htm#
- Brachiopoda:- paleopolis.rediris.es/BrachNet/
- British brachiopods:- www.earthsci.gla.ac.uk/staff/gbc/databases/brachs/brachpage.html
- Encyclopedia of Marine Life of Britain and Ireland:- www.habitas.org.uk/marinelife/
- Antarctic amphipods:- www.natural sciences.be/amphi/
- North East Atlantic amphipod identification guides:- www.amphipoda.com/index.html.

The addition of additional fields to the species in ERMS, beyond the present synonyms (which will be expanded upon) and their occurrence in the Atlantic and Mediterranean, will also be considered. Matching feedback from users as to what their needs are with available resources and technical tools, may allow matching of species to national Exclusive Economic Zones, MARBEF ATBI (all taxon biodiversity inventory) sites, conservation importance, and species information sources.

authority file. However, there are now GSD online that could connect to ERMS, such that there is only one (global) nomenclature for each taxon. Such GSD include FishBase (www.fishbase.org), AlgaeBase (www.algaebase.org), CephBase (www.cephbase.org), Hexacorallia (www.kgs.ukans.edu/Hexacoral/). In addition, there is the online European mollusc database CLENAM (www.mnhn.fr/mnhn/bimm/clemam/). How this can be achieved is a key technical objective of the recently funded EuroCAT (=Species 2000 Europa) project (see above).

European Ocean Biogeographic Information System

Another data management component of MARBEF is to connect existing databases around Europe that have species distribution

Littorina littorea: single scientific names are still not used even for common species.

data so as to establish a European Ocean Biogeographic Information System. This will require software to compare species names, translate synonyms to that in ERMS, and identify names that need expert attention to determine whether they are incorrect names or additional to ERMS. Tackling this task will benefit from the technical solutions found to integrating the global-regional species lists comparisons by EuroCAT. This European OBIS will be one of a global network of regional nodes of OBIS (www.iobis.org). OBIS is a global programme mapping marine life over the internet, and the data management component of the Census of Marine Life (www.coml.org). OBIS is the specialist provider of marine data to GBIF, and one of its largest data providers. Mark Costello is Chair, and Geoff Boxshall and Edward Vanden Berghe members, of the OBIS International committee. Carlo Heip is a member of the European CoML committee, and coordinator of MARBEF.

North Atlantic Register of Marine Species (NARMS)

The research museum at the Huntsman Marine Science Centre in Canada (called the Atlantic Reference Centre, ARC) has been developing an authoritative register of species occurring from the Arctic to Cape Hattaras in the USA, under the direction of Gerhard Pohle, Lou Van Guelpen and Mark Costello (at HMSC, 2000-2004). This register has developed from the ARC museum database and related projects. With recent funding from the Gulf of Maine Census of Marine Life programme and the CORONA project, it will be completed and placed online at the website of the Centre of Marine Biodiversity (CMB) (www.marinebiodiversity.ca) by ARC staff with the support of DFO (Fisheries and Oceans), notably Bob Branton of the Bedford Institute of Oceanography. CORONA is a US National Science Foundation project led by Cliff Cunningham of Duke University on the historical ecology of trans-North Atlantic marine species (www.biology.duke.edu/corona/). The NARMS website would search

both the ARC species register and ERMS simultaneously and return to people whether a species was recorded in either database, and access to associated information (e.g. synonyms). Species data and information from the ARC, DFO and other organisations can be searched in space and time on the CMB website, and it includes many European species (Atlantic Canada is a largely a subset of the NE Atlantic biota). This initiative may be a model for how different regions can bring added-value to their data by comparisons with each other, and reduce duplication of effort.

What does marine biodiversity informatics do for and mean for taxonomy?

More rapid and accessible authoritative information on species nomenclature will reduce incorrect and variable use of species names in research, education, recreation, industry and management. This should reduce the amount of time taxonomists spend in correcting past errors in the literature and responding to individual questions. This should free up time to produce identification guides and describe species new to science. Taxonomists are the most fundamental source of "bio-information" because putting the correct name to a biological entity is essential for quality research and management. Recognition of this is clear in the European electronic conferences (e.g. www.vliz.be/marbena/) and reports of the European Network in Biodiversity Information (www.enbi.info/). The arrival of biodiversity informatics thus increases the need for taxonomists and creates greater awareness of their key role in managing biodiversity information. ●

Dr Mark J. Costello,
Leigh Marine Laboratory,
University of Auckland, New Zealand
m.costello@auckland.ac.nz

References

Costello, M.J. 2000. Developing species information systems: the European Register of Marine Species. *Oceanography* 13 (3), 48-55.

Costello, M. J., Emblow, C. and White R. (eds) 2001. European Register of Marine Species. A check-list of marine species in Europe and a bibliography of guides to their identification. *Patrimoines Naturels* 50, 1-463. ISBN 2-85653-538-0, ISSN 1281-6213.