

# **Introduction to CephBase [www.cephbase.utmb.edu](http://www.cephbase.utmb.edu)<sup>a</sup>**

by

**James B. Wood<sup>b</sup>, Ronald K. O'Dor<sup>c</sup> & Uwe Piatkowski<sup>d</sup>**

## **Abstract**

The combination of relational databases with the Internet allows for the exchange of information on a scale never before experienced. Database driven web sites can rapidly publish large sources of data on a platform accessible across geographic and political boundaries. This rapid and cost effective method of exchanging copious amounts of information will greatly help scientists understand the biocomplexity of natural systems.

CephBase is an example of a taxon -specific web site powered by a relational database. CephBase serves data on all 703 extant cephalopod species. It is part of the "Census of Marine Life" effort that grew out of an U.S. initiative into a wider international undertaking in support of implementing documentation requirements of the Convention on Biological Diversity.

The purpose of CephBase is to collect and provide information on life history, distribution, catch and taxonomic data on all living species of cephalopods (which includes octopuses, squids, cuttlefish and nautilus). Technical information on its current structure and access is discussed below together with an outlook on how cooperation between divergent biological databases could be facilitated.

a) Prepared for the INCO -DEV International Workshop on Information Systems for Policy and Technical Support of Fisheries and Aquaculture, Los Baños, Philippines, 5-7 June 2000;

b) National Resource Center for Cephalopods, University of Texas Medical Branch, Galveston, TX 77555 - 1163, USA. E-mail: [ceph@is.dal.ca](mailto:ceph@is.dal.ca)

c) Biology Department, Dalhousie University, Halifax, Canada B3H 4J1. E-mail: [odor@is.dal.ca](mailto:odor@is.dal.ca)

d) Institut für Meereskunde, Universität Kiel, Düsterbrooker Weg 20, D -24105 Kiel, Germany. E-mail: [upiatkowski@ifm.uni-kiel.de](mailto:upiatkowski@ifm.uni-kiel.de)

## Introduction

Cephalopods have short life spans, fast growth rates (exponential when young), and they tend towards semelparity. They are quite distinct from fish, not only in their morphology, but also in their life history. Cephalopod fisheries are a growing percentage of the world's catch. However, as is typical of invertebrates, they are understudied and much of their basic biology is still unknown. Models developed for fish have not reliably been able to estimate cephalopod life cycles (Jackson & Choat, 1992). From a fisheries management perspective, more information is needed on this resource (Piatkowski & Wood, 2000).

CephBase is a dynamic html (dhtml) relational database -driven interactive web site. The prototype version of CephBase was developed at Dalhousie University, Halifax, Canada. This project was initially sponsored by the Sloan Foundation following the Workshop on Non-Fish Nekton in Boston, Massachusetts on December, 1997. The purpose of CephBase is to collect and provide information on life history, distribution, catch and taxonomic data on all extant cephalopods (octopus, squid, cuttlefish and nautilus). This pilot project was designed to help define the goals of the Census of Marine Life and to demonstrate what was possible with emerging Internet technology (Wood *et al.*, 2000). The web is an ideal globally accessible platform for collaboration and communication of data and interest in this new tool is growing like wildfire.

CephBase provides reliable referenced data on cephalopods and a platform for collaboration both within the cephalopod community and with all marine sciences. This is accomplished with a dynamically driven web site powered by a relational database. Since CephBase went online in August, 1998, it has been referenced in NetWatch twice (*Science*, 282:587 and 285:2027). CephBase and other Census of Marine Life Projects were featured in an upcoming full-length article in *Science*.

CephBase is similar in concept to FishBase ([www.fishbase.org](http://www.fishbase.org)), but instead of distributing data on yearly CDs, it was designed from the outset to operate dynamically from the web. This has several advantages: 1) everyone with an Internet connection has full access, 2) the latest version is always available, 3) CephBase functions on many platforms and operating systems, 4) publication costs are extremely low.

CephBase recently received a grant for the National Oceanographic Partnership Program and is expanding its scope. We are working with the Ocean Biogeographical Information System (OBIS) to explain the diversity, distribution and abundance of marine life (Grassle, 2000). With this support, the main CephBase database will soon be moved to a Client-Server database at the National Resource Center for Cephalopods in Galveston, Texas. We are adding new types of data as well as continuing to populate existing tables.

## How does CephBase work?

Under the hood of CephBase you will find a Microsoft Access database. This relational database holds the information in various tables. SQL (Structured Query Language) is used to manipulate these data. An NT server linked with Cold Fusion (an API or Application Programming Interface) serves dynamic web pages to users. This allows our database to be fully accessed by anyone with a computer linked to the Internet. Despite the complexity of CephBase, the web page is designed to be user friendly.

## What are the current features of CephBase?

1. Classification of all known cephalopods. We currently have all the taxa, authorities, and the year the taxa was described online for all of the 703 living species of cephalopods listed by Sweeney and Roper (1998). This information is fully searchable by using our search engine. Results are listed in a dynamic table. Simply click on any species in the table and all the taxonomy, from class to subspecies, for that cephalopod is displayed. Furthermore, synonymies, type repositories, type localities, references and common names are listed. References are listed in brief with full references only a click away. This demonstrates what can be done and it isn't hard to imagine life history, ecological and morphological data being accessed in the same manner.
2. Distribution and mapping. We currently have over 3,150 localities for about 320 species in our database. Maps are made on the fly using the Xerox Parc Map Viewer. All latitude and longitude data used to generate maps are from published sources and are listed in tables and referenced. In most cases, the individual specimens used to populate our database can be tracked to a museum repository.
3. Ecological data. These data are needed to fit cephalopods into global models. At present, our database holds 420 predator records and 979 prey records. All are referenced and all have the complete Latin name of the cephalopod involved in the interaction. Papers that just list "squid" are not included in the database.
4. Pictures of selected species. The Internet is a graphical medium. Providing online images of as many species as possible is a very useful resource. Cephalopods change texture, color and shape and these traits are important for taxonomic, ecological and evolutionary studies (Hanlon & Messenger, 1996). At present, links to pictures of 30 selected species are available from The Cephalopod Page maintained by J. Wood. One of our current initiatives is to place thousands of images from the National Resource Center for Cephalopods (NRCC) online.
5. International Directory of Cephalopod Workers is maintained to help foster global collaboration.

CephBase is global in species coverage. We made a concerted effort to include cephalopod data from all over the world, i.e. all 703 extant species are covered. However, there is likely a North American bias in the location, predator and prey data, because the initial CephBase team was based in North America. This is an area that additional partners in other regions could help to improve. Data can be fed to the existing database from anywhere in the world.

## What species level information does CephBase provide today?

The following information on cephalopod species is available on CephBase:

1. A classification of the recent Cephalopoda to subspecies level.
2. Type locality.
3. Type repository.
4. Common names.
5. Synonymies.
6. Distribution and real-time plot of latitude/longitude data (from Xerox Parc Map Viewer).

[In addition](#), the following information is available for some species:

1. Links to pictures of approximately 30 selected species.
2. Cephalopod predators.
3. Cephalopod prey.

### **Future Directions for CephBase**

1. Improved on the fly mapping system.
2. Move to client-server architecture for increased security.
3. Addition of life history and fisheries data to the database.
4. Addition of more predator, prey and location data to the database.
5. Maintain and expand the International Directory of Cephalopod Workers.
6. Establish mirror sites in other countries to improve data sets from outside North America.
7. Create a database of taxonomically and behaviorally important images.
8. Establishment of a references database.

### **Some ideas on future management of Internet databases**

It would be a great catalyst to create and support a generic blank species level database and promote its use for higher taxa not yet covered. This would also encourage others to start new database projects on groups in their area of expertise. An important additional benefit is that new projects that start from the same point and use the same protocols will be a lot easier to integrate and could cooperate with each other in the future. This idea should be a priority and needs to be well funded and supported. However, it would pay high dividends by jumpstarting new database development and facilitating collaboration of similarly set up databases (Wood *et al.*, 2000).

It would also be useful to create a more organised "family" of \*.base and have all members provide links to all other members. It would be desirable if this were done in the same placement on each site so users would have some consistency. Some minimal standards should be met before a database is included. This would be most useful to developers of species level sites if it was created independent of the many sites that want to serve as portals to the sites that gather and publish actual species level data.

A list server for those who are actively involved with biological relational databases should be created. All of us are going through many of the same problems as we develop these projects and some mutual help and teamwork would go a long way to increase efficiency.

CephBase uses maps generated in real time. While there are occasional problems with the unsupported and free Xerox Parc server we use, this method is certainly the way to go at this time. When a user clicks on a species distribution map he or she can zoom in from a global view to a local area to get details. This option alone would take scores of static maps for each species that would take up a lot of hard drive space. Furthermore, when we add a new record to the database, it is automatically and instantly included in maps requested by users. The alternative is to plot all those points by hand on scores of static maps for each species, then upload them all. Generating maps in real time from your database is the way to map specimen locations and was first done for a marine species database by CephBase. With help from the Ocean Biogeographic Information

System (OBIS) we hope to eventually move away from Xerox Parc to more powerful software where we can overlay environmental conditions (depth, temp, etc) on our distribution maps. The cephalopods are a small group compared to fish. This makes them an ideal test group for new methods.

When setting goals, scientists should keep in mind that the database structure is excellent at handling certain kinds of information and terrible at others. Databases do extremely well with hierarchical information such as classification. Standard fields, such as weight or age at maturity, also work well. Data in odd formats, data fields of widely varying size, datasets with many missing fields, or repeated data can all cause problems and do not work well (Forta, 1998). It is critical for primary investigators to have some idea about what can and cannot be done in order to set realistic goals.

Other database projects use a variety of software and hardware. CephBase is currently using MS Access and even others using the same software will have set up their tables and relationships differently. Meta data standards need to be agreed to and put in place as soon as possible; otherwise, database projects will grow apart instead of together.

## Conclusion

Relational databases like CephBase provide a powerful tool for the storage and retrieval of many types of data commonly used in science. Examples of data suitable for databases include taxonomy, ecological relationships between species, catch data, growth data, and much more. HTML and the Internet provide a new media that has many benefits for scientists. Publication costs are extremely low and scientists as individuals or using a peer reviewed processes can directly control the content of the media. Large datasets and color images can easily be published and shared with colleagues. The combination of relational databases with the Internet allows for the exchange of information on a scale never before experienced. Database driven web sites can provide large sources of information that are made rapidly available on a platform accessible across geographic and political boundaries. This rapid and cost effective method for exchanging large amounts of information will greatly help scientists understand the biocomplexity of natural systems. This technology is in its infancy and a little collaboration and coordination between workers now will yield high dividends in the future.

## References

- CephBase: <http://www.cephbase.utmb.edu/cbplay1.html>
- Forta, B. 1998. The ColdFusion web application kit. Que, Indianapolis, Indiana, 1000 p.
- Froese, R., 2001. The FishBase Information System: Key features and approaches. Brussels, *ACP-EU Fish.Res.Rep.*, (8):36-38.
- Grassle, F.J., 2000. The Ocean Biogeographical Information System (OBIS): an on-line, worldwide atlas for accessing, modeling and mapping biological data in a multidimensional geographic context. *Oceanography*, 13:5-7.
- Hanlon, R.T. & J.T. Messenger, 1996. Cephalopod behaviour. Cambridge University Press, 232 p.
- Jackson, G.D. & J.H. Choat, 1992. Growth in tropical cephalopods: An analysis based on statolith microstructure. *Can.J.Fish.Aquat.Sci.*, 49:218-228.

- Piatkowski, U. & J.B.Wood, 2000. Cephalopods: What makes them an ideal group for an Internet database. *ACP-EU Fish.Res.Rep*, (8):39-43.
- Sweeney, M.J. & C.F.E. Roper, 1998. Classification, type localities, and type repositories of recent Cephalopoda. *Smithson.Contr.Zool.*, 586:561-599.
- Wood, J.B., C.L.Day, P.G.Lee & R.K. O'Dor, 2000. CephBase; testing ideas for cephalopod and other species-level databases. *Oceanography*, 13:14-20.
- The Cephalopod Page: <http://is.dal.ca/~ceph/TCP/>