



Two huge database projects are amassing everything fisheries' managers and anyone else interested could want to know about marine organisms

Basis for change

By Patricia Sorongon, Marianne Pan & Steve White

People have been observing life in the seas for centuries. Well before we were able to scuba dive, we were cataloguing the things we netted or dredged up from the bottom. The results of those explorations are scattered across the globe in libraries, universities and research institutes, and crucially, that is where most of it stays.

Back in 1987, Daniel Pauly, a French fisheries biologist now based in Canada, was working for ICLARM (International Center for Living Aquatic Resources Management, now known as the WorldFish Centre) in Manila. Doing research in the tropics proved a frustrating experience at times: Pauly was where much of the diversity of life was, but where much of the literature wasn't.

This led him to conceive of a database that would stitch together all the information we have ever accumulated on the fish of the oceans, gathered from every possible source around the world.

In 1990, together with colleagues Rainer Froese and Deng Palomares, Pauly launched his project. Called FishBase, it was initially planned to serve the needs of scientists and policy specialists by concentrating on 2,500 commercial fish species: giving data on identification, growth, mortality and biology to aid in fisheries management issues.

The database makes extensive use of the established (but often esoteric and/or out-of-print) literature, as well as 'grey' sources such as unpublished papers, reports by non-governmental agencies and the like. All of this is reviewed by scientific experts before it is encoded to join countless other snippets of data.

FishBase aimed first to build as complete a taxonomical backbone – using the Catalogue of Fishes, assembled by William Eschmeyer of the California Academy of Science – to the database as possible, adding biogeographic information where it was available. Now

though, the scope of the database is expanding, with deeper information being added on the biology of the species.

The usage of FishBase has also expanded considerably – exponentially so since the Internet became a part of everyday life for so much of the developed world. Today, the scientists and resource managers using the site have been joined by educators, recreational divers, aquarium enthusiasts, customs officers (identifying illegally traded species) and many,



many others. The biggest are school children doing their homework and altogether the site receives somewhere around 25 million hits per month.

Buoyed by their roaring success, Pauly and his team are now embarking on something still bigger.

"In one of the annual meetings of the FishBase Consortium – at the natural history museum in Paris in 2003 – the members discussed the request of colleagues in marine biology for a similar information system for non-fish organisms", says Palomares.

Today she is project coordinator of what

is known as SeaLifeBase, which aims to catalogue every living thing in the seas bar the fish. Funded by the Oak Foundation, it very much builds on the example set by FishBase.

Pauly is not the only one with the idea for such repositories of information though. There are rival databases out there, many of them well funded, but according to him, they typically make several crucial mistakes.

One major error in his eyes is to use the scientific name as the main identifier when most fish have one or more common names. Doing this results in what Pauly calls, "a cupboard full of clothes hangers with nothing hanging on them" – there is less chance of others finding the entry by chance or by design, and adding information to what is already there. Often databases even raise the barriers to entry still higher by requiring you to register, giving a username and password.

Pauly says it is crucial too that numeric data of all types is "hard-coded and standardised" This means making it searchable so it can be analysed and readily cross-referenced. This has led to many insights: "We see patterns emerging that were not visible before", he says. "Many specialists think that their stuff is so different that they never compare it with the work of others."

As an example he mentions a French scientist who had weighed the brains of fish – not just one species but 100s of species. FishBase constructed a table for this data and this has led in part to the discovery that reef fish have much bigger brains than other fish.

Having worked so hard to secure the necessary funding for his projects, Pauly is particularly unhappy with what he sees as wastage elsewhere: "Many people think it is how it looks that is the thing", he says.

A squat lobster, *Munida zebra*, recorded in deep-sea environments off New Caledonia and Palawan. **Facing page:** When research is fun – diving for data.

“They blow all the money on the software.” Typically this software-intensive approach means quarrying the net and connected resources for data, using state-of-the-art artificial intelligence techniques to mine for valuable nuggets buried in all the ore. This costs big money and Pauly is not convinced of its effectiveness.

He believes the mash-up databases are all cannibalising each other’s data and are not encoding anything new. “All the money is spent on an intelligent agent to read the books – it is much better to train somebody who has a Masters degree in marine biology.” The encoding team of SeaLifeBase in the Philippines is presently adding around 5,000 species a month.

“We know from the experience with FishBase that the small investment we make on training research assistants in the Philippines gives a large return in terms of the work done and in terms of the benefits these people make from being in the project, such as getting a good graduate education”, says Palomares.

“When I set up the SeaLifeBase project, one of my objectives was to hire fresh graduates from the University of the Philippines to form the team,” she says. “I had a very good mentor in Daniel Pauly and I thought I should pay it forward. So, when Daniel asked me to be project coordinator for SeaLifeBase, I saw right away that I would be able to plant a few seeds. I am now doing my best to nurture



Above: A showy species of mantis shrimp, *Odontodactylus* sp. found off Palawan. **Bottom left:** Daniel Pauly, the architect of FishBase and SeaLifeBase.

the SeaLifeBase’s small team of young fresh graduates, encourage them to learn and get the necessary training they need to compete in the western scientific world, and most of all, assure that the projects we built so far will have the new generation to see them to completion.”

This Asian-based workforce also helps to address another of Pauly’s concerns: that of geographical bias. “When US or European developers work on a database, they are mostly concerned with their own waters. Few people have a global mandate.” For that you need lots of collaborators – in FishBase’s case, around 1,000 all over the world.

Pauly believes in giving all these people the credit they deserve, each getting their name and picture on the database. But collaboration doesn’t extend to being able to structure the design of the database, for that would lead to paralysis as rival factions fight to have their point of view prevail. Instead, Pauly’s databases go back to the classics on the bookshelf. They use the standard reference work on a given family of species or aspect of their biology as their blueprint, refining it where necessary with selected experts from the field.

This is a challenge for SeaLifeBase in particular which began by looking at groups of marine animals for which there were no

databases available or no authorities that had taken the initiative to collect basic information. They tend to be the smallest groups of organisms: groups like the microscopic phyla *Cycliophora*, *Micrognathozoa* and *Placozoa*, each of which has only two species. Now the project is moving on to larger groups.

Marine mammals (<200 species) and seabirds (<500) have already been documented and Palomares is looking forward to seeing some of her own earlier work on sea cucumbers (>1,700 species) find a place in SeaLifeBase. Beyond that, much larger groups beckon such as the molluscs with around 90,000 species.

Meanwhile FishBase isn’t standing still. “Data encoding continues at a regular pace, even after 15 years of existence,” says Palomares. “Our work is never complete as new data keeps getting published. Database maintenance, though is more straightforward, as we have acquired the expertise over the years.”

To contribute your own data or just view the 30,000 fish species already logged, see www.fishbase.org. If it is other sea critters you are after, see www.sealifebase.org ΔΔ

