

MBL MOMENT

...with

David "Paddy" Patterson
Senior Scientist, JBPC



*David Patterson's research interests in the diversity, evolution, and classification of microbes led him to the MBL in 2000. Formerly, he was head of the School of Biological Sciences at the University of Sydney, Australia. Patterson's team at the MBL is developing the information management system on biodiversity that underpins the Encyclopedia of Life. This system draws from two prior MBL efforts: uBio, a repository of taxonomic names and names-based services and tools developed in 2000 by David Remsen and Patrick Leary of the MBLWHOI Library and Patterson; and micro*scope, a names-based index of information on microorganisms developed by Patterson.*

The Encyclopedia of Life project is funded initially by the John D. and Catherine T. MacArthur Foundation and the Alfred P. Sloan Foundation.



As Grand as Life Itself

Once in a while, a scientific effort of truly historic dimensions captures the imagination of all involved. Such is the case with the Encyclopedia of Life (EOL), which has been called a "moon shot" for biology by its leading advocate, Harvard biologist E.O. Wilson. Over the next ten years, the Encyclopedia of Life will create a Web page for each of the 1.8 million named species on Earth. And, ideally, it will catalyze the discovery of new biodiversity, particularly in the largely undiscovered domains of microbes and insects. The MBL is a cornerstone institution in the giant EOL effort, along with Harvard University, the Smithsonian Institution, the Field Museum of Natural History, the Missouri Botanical Garden, and the Biodiversity Heritage Library. MBL director and CEO Gary Borisy is on the EOL Steering Committee and chairs its Distinguished Advisory Board. Cathy Norton, director of the MBLWHOI Library, is deputy director of the Biodiversity Heritage Library consortium, which is creating an open-access, digitized collection of natural history literature that will be available to the EOL. David Patterson, interviewed below, heads the EOL's Biodiversity Informatics Group.

MBL By early next year, the EOL plans to have 50,000 draft pages online. This sounds ambitious! Please explain the process of creating an EOL Web page for, say, the great white shark.

DP First, the EOL is not primarily an environment where people will write pages about species. Rather, we presume that within ten years, most information about most species will be accessible through the Internet. And our task is to find all the relevant information, say, on the great white shark, draw it across the Internet, reorganize it and present it as an EOL page. For Version 1 of the EOL, we are identifying a number of data providers whose Web sites already contain everything we would want for an EOL page, and are establishing agreements

with those providers. For example, upon agreement, our team should be able to draw out information on 30,000 fish species from FishBase within a matter of days. If there is no existing Web site that carries everything we want, then we have to bring information together from multiple Web sites and build up a new page. This process is referred to as aggregation or a mash-up.

MBL How do you create a species page that is appropriate for all potential users?

DP An important component of the EOL concept is it is a communally owned environment that will allow anybody to build screens and filters that will suit their needs. Say we have an EOL page on the great white shark that came out

"Kids, these days, are increasingly distant from the natural world and there is not much to involve them with it. With the EOL, we are developing the tools so they can go out, take pictures with their cell phones of organisms in their backyards, submit them to the EOL, and experts will help identify them."

—David Patterson



of FishBase, which is created for use by specialists. A teacher in Chicago looks at this and realizes it's pretty close to incomprehensible for a student audience. So the teacher may choose to rewrite it. What the teacher does is create a new page, which will be identified as being appropriate to a student audience. The students can then be directed to use that page. Like other visitors, they will have a "My EOL" function where they can set preferences for what kind of information they wish to see. A 14-year-old student will have a different view of EOL content than a molecular biologist will.

MBL So several versions of a species page are created in the EOL, rather than everybody having their hands on one version.

DP Yes. And that makes us very different from the standard wiki, such as Wikipedia, where there is the sense that there is one truth and the masses collectively will find that truth and articulate it. The EOL will have an enormous number of user groups, and we will create the tools to allow them to customize the environment to suit their needs.

MBL What value will the EOL bring to scientists?

DP Most biology at the moment is very parochial. That is, an individual or a small group collects a subset of information about one species or a small number of species. They publish the views that have emerged from their analyses. The raw data is then often hidden away and all one sees is a synthesis. And this creates "islands" of knowledge. There is one major exception to this: GenBank, a large, online, communal database of raw data where, for example, a scientist interested in the evolutionary history of a species can pick up molecular information on hundreds of other species and analyze it for similarities. Currently, there is no mechanism for doing that for any other aspect of the biology of an organism,

such as information on aging or life-history strategies or ecology. Our vision for the EOL is that there will be a large communal pool of data and individuals can reach in and select the information that is relevant to their agenda, such as, "I would like all the information about the life-history strategies of butterflies." This development would have a truly transformational impact on biology. It would change biology from a parochial endeavor to something that is very grand in its power. Biologists would be able to move to considerably more global questions. They would work less as individuals and more as members of large teams with much grander agendas than they currently have.

MBL The EOL seems to hold enormous potential.

DP This project is a joy to be involved in. It's inspirational, it attracts people. One kid wrote me a letter that said, "I wrote my first book on snakes when I was 12 years old, and now I'm 16. And you need me because I will bring you 'street cred.'" And he is now our advocate for the EOL on the social networking sites. Kids, these days, are increasingly distant from the natural world and there is not much to involve them with it. With the EOL, we are developing the tools so they can go out, take pictures with their cell phones of organisms in their backyards, submit them to the EOL, and experts will help identify them. The data points will go into maps, which are then used to plot the distribution of species. So kids, and all contributors, can add to the scientific analysis of global climate change—probably the matter of greatest concern for this world. •

