

## *The Encyclopaedia of Life*

### *E O Wilson and TED*

When renowned biologist E O Wilson collected his TED award on 8th March 2007, he wished for an 'Encyclopaedia of Life' to inspire the next generation "to search for life, to understand it and to preserve it."

The impetus to create an online central, one-stop-shop for the biodiversity data which currently resides - both on and off-line - in libraries, museums, dictionaries, almanacs and universities across the globe has been around for a while. In the 1990s, individuals were already thinking about the potential of the web for hosting pages about every living species. E O Wilson has been one of the leading proponents of the project and articulated the idea in a 2002 essay, but it is only now that technology has developed sufficiently to make the concept a reality.

In late 2005, E O Wilson, Professor Emeritus at Harvard University, wrote to the MacArthur Foundation - one of America's largest private philanthropic foundations - suggesting that the time was right to consider the feasibility of the Encyclopaedia of Life (EOL). Work on the EOL commenced in January 2006, but has recently accelerated thanks to support by the Foundation and the Harvard professor. In 2007, he received the TED Prize and used the occasion to wish for the creation of the project. The TED (Technology, Entertainment, Design) conference awards a prize of \$100,000 to anyone with the potential to inspire others to change the world. Award winners are also asked to make a wish.

Following this recent publicity, on May 9th 2007, the Encyclopaedia of Life launched a global effort to document all 1.8 million species of plants, animals and other forms of life on earth. The project has a \$20 million grant from the MacArthur Foundation and \$2.5 million from another US grant-making organisation, the Alfred P Sloan Foundation. Combined with funding from its six founding partners, the project has a total start-up pot of \$50 million.

### *Partners*

Central to the success of the EOL is the participation of the international scientific community, which holds the information the project seeks to collate and which will also be one of its primary benefactors. Speaking at the launch, MacArthur Foundation president Jonathan Fanton spelt out the project's commitment to freedom and inclusion: "for the EOL to reach its full dimensions, it must be a global endeavor, a resource of knowledge that is created by all, maintained by all, and with benefit to all."

The project is being led by the Field Museum of Natural History, Harvard University, Marine Biological Laboratory, Missouri Botanical Garden, Smithsonian Institution, and the Biodiversity Heritage Library (BHL). Representatives from a range of universities and institutions sit on the EOL's governing structures. It is hoped that new partners will join from other countries.

The BHL is a consortium of 10 of the world's largest natural history libraries - including the London Natural History Museum and Kew Botanical Gardens -

which will scan and digitize tens of millions of pages of the scientific literature. The BHL has scanning centres currently operating in London, Boston and Washington DC and has produced 1.25 million pages so far. It is both acquiring documents published prior to 1923 - and therefore not subject to copyright restrictions - and pursuing more recent publications from scientific bodies and university presses. It has also contacted larger, for-profit media groups, some of which publish thousands of journals. This resource will provide free and open access to detailed knowledge online and be a core feature of the EOL.

### *Purpose and implications*

The openness of the EOL will mean that anyone – both within the scientific community and the general public – will have access to an unprecedented volume of information. Besides raising public awareness of biodiversity and Earth's ecology, this will have a profound impact on the academic community; there will be implications both for the process of conducting research and the type of knowledge research will produce.

E O Wilson hopes that the EOL will revive the practice of taxonomy; the scientific study and practice of classification. Species taxonomy was developed by Swedish botanist Carolus Linnaeus nearly 250 years ago and is one of the most poorly funded and under-developed biological disciplines, with only 6000 biologists globally working in the field. Yet the majority of life on the planet remains unknown or ill-understood. It is not even known for certain how many formally identified organisms exist; estimates stand between 1.5 and 1.8 million. This is still only a fraction of the true number of species inhabiting Earth; estimates for this range from 3.6 million to as many as 100 million. The EOL's task of collating all the current data and literature on species may already be considered immense without taking into consideration the volume of knowledge we, as yet, do not have. But Wilson hopes that by collecting all existing information in one place, the gaps in our knowledge will stand out and point researchers towards new discoveries. For the eminent biologist, the EOL is about completing "the great Linnaean enterprise."

The species which remain undiscovered will become increasingly important as environmental damage progresses. Greater knowledge will aid estimates of species extinction and help assess the human impact on the environment. Wilson told the TED audience that in 2007 the project has become even more pressing because, as environmental threats worsen, we are at risk of losing information even before it has even been discovered.

The EOL will also accelerate the pace of research by providing open, instant access to information which researchers would once have had to travel to different museums and libraries for. Wilson describes biology as operating using Nineteenth Century methods where species identification is extremely time consuming. Researchers who are searching for a new species must examine many different specimens, typically scattered across research institutions – and across the globe. They need access to books and journals which are often old and rare and gaining access to all these resources takes time and money, which may be prohibiting factors in undertaking the research at all. The EOL will change all of this. Wilson told the *New Scientist* that he thought the online resource would enable him to complete a book in

six months which had previously taken him 25 years of part-time work.

Not only will the encyclopaedia change the how research is carried out, it is hoped that it will also fundamentally alter the nature of biological science. Wilson thinks the ELO "should be a Big Science project, equivalent to the Human Genome Project. It should be thought of as a biological moonshot with a timetable." It will have the same effect on comparative biology, ecology and other related fields as the HGP had on the biotechnology industry and allow researchers to explore patterns too large or complex to have been studied with old tools. Computers program will allow researchers to trawl massive amounts of data very quickly, thus seeing trends which were previously imperceptible. MacArthur president, Jonathan Fanton explained the EOL had the potential to serve as a "macroscope". Whereas a microscope allows us to discern the very small and particular, the EOL could find "patterns previously unseen, illuminating relationships, identifying gaps in our knowledge, and suggesting opportunities for new avenues of inquiry."

#### *How the project will work*

Collating data on 1.8 million known species whilst integrating newly discovered information along the way is a mammoth task. The project is still in the development stage and there are a number of unknowns at present, for example, there may be potential problems with intellectual property. The venture will involve utilising existing technologies as well as developing new software to make the EOL's vision for the website a reality. It will draw both on expert knowledge and incorporate user-generated information.

Initial work will emphasize species of animals, plants and fungi, but the project can also be extended to microbial life. Pages will provide written information and, where available, photographs, video, sound, location maps, and other multimedia information on each species. The EOL hope to have some material made publicly available by 2008, but it is anticipated that it will take a decade to create web pages for each species known to science. The project will, however, be a constantly evolving entity as our knowledge of known species increases and new species are discovered. This is why the international scientific community will be integral to the running of the project. They will both use it in their own research endeavours and help to provide the regular updates which will be required to keep the project current.

Unlike conventional encyclopaedias, the EOL will be developed by bringing together content from a wide variety of sources, in a similar way to how a mash-up might pull data from Google Earth and Twitter to create a new application which maps where your friends are geographically. The EOL will be closer to a Wikipedia than a Britannica. In fact, the project cites the user-generated encyclopaedia as its inspiration, saying it "gave us confidence that our tasks are manageable with current technology and social behaviour." The Wikimedia Foundation is also a member of the project's Institutional Council.

The first task is to create a dynamic catalogue of available data which will form the basis of the EOL. Providers will need to identify themselves, log the nature of their data and conditions of use. The EOL will need to work with these data providers to find common standards of information to keep data which reaches the EOL consistent and useable. This will create an index of all the known data elements regarding particular aspects of the biology of each

species. Information will be aggregated from collaborating authoritative websites which are already dealing with aspects of the world's biodiversity – for example, FishBase, a 'global information system' on fish, Tree of Life, which also collates biodiversity information from expert and amateur contributors or information from numerous online collections from museums and universities worldwide. The system will use taxonomic and semantic intelligence to group and catalogue this data. When these websites are updated, the changes will automatically and instantly be communicated to the EOL. The project will begin with the BHL content and other sources of trusted, high-quality material.

Alongside this expert-generated content, the EOL will incorporate user-generated material in a variety of ways. The data aggregated as described above underpins the EOL 'WorkBench'. The WorkBench will combine open source software tools with services from larger agencies such as Connotea (essentially Del.icio.us for scientists), Flickr or Google which allows individual users to annotate, manipulate, analyze or visualize the information themselves. The EOL will be customizable in other ways as well. MyEOL will provide a personalized experience for users, allowing them to adjust preferences such as language or level of understanding, as well as logging their favourite pages, contacts and other information.

The project envisages a three-tiered quality-control system for user-generated content. Data entering at the lowest level will not be visible automatically on the site. If it meets certain administrative standards, it is elevated to 'silver' standard. Finally, it is checked by expert reviewers and becomes 'gold' standard material which is visible to everyone on the web.

There will be a number of ways to find what you are looking for on the site. Classification and evolution trees will be accompanied by visualisations to allow users to browse as well as search through taxonomic data. Meta-data will be attached to species which describe the organisms by geographic location or physical characteristics and thus enable users to search for species by these descriptions.

But this is not only a resource for the academic world. The EOL will also prioritize developments which empower 'citizen science'. For example, they hope to add tools in collaboration with Google Earth which allow anyone to record the location and time of observations of individual specimens. This puts millions of pairs of eyes at the disposal of the EOL to create maps which show the changing distributions of species or seasonal differences and could engage the world at large in tracking the impact of climate change.

Development of web pages begins later this year, but the project will evolve over time as technologies such as handheld devices and GPS develop. There is ultimately the potential for integrating the EOL within games, popular social websites, news and current affairs programmes as well as web-based publications.

