New species discovery, description and data sharing in less than 30 days

March 27 2014

A new spider species was discovered during a student field course in Malaysian Borneo. The species was described and submitted online to Biodiversity Data Journal through a satellite internet connection. The manuscript was peer-reviewed and published within two weeks. On the day of publication, software tools extracted the occurrence data of the species and submitted these to the Global Biodiversity Information Facility. In a similar way data was exported to Encyclopedia of Life.

Researchers and the public can now have immediate access to data underlying discovery of new species of life on Earth, under a new streamlined system linking taxonomic research with open data publication.

The partnership paves the way for unlocking and preserving a wealth of 'small data' backing up research conclusions, which often become lost within a few years of an article's publication in an academic journal.

In the first example of the new collaboration in action, the Biodiversity Data Journal carries a peer-reviewed description of a new species of spider discovered during a field course in Borneo just one month ago. At the same time, the data showing location of the spider's occurrence in nature are automatically harvested by the Global Biodiversity Information Facility (GBIF), and richer data such as images and the species description are exported to the Encyclopedia of Life (EOL).

This contrasts with an average 'shelf life' of twenty-one years between
field discovery of a new species and its formal description and naming, according to a recent study in *Current Biology*.

A group of scientists and students discovered the new species of spider during a field course in Borneo, supervised by Jeremy Miller and Menno Schilthuizen from the Naturalis Biodiversity Center, based in Leiden, the Netherlands. The species was described and submitted online from the field to the *Biodiversity Data Journal* through a satellite internet connection, along with the underlying data. The manuscript was peer-reviewed and published within two weeks of submission. On the day of publication, GBIF and EOL have harvested and included the data in their respective platforms.

The new workflow established between GBIF, EOL and Pensoft Publishers' *Biodiversity Data Journal*, with the support of the Swiss NGO Plazi, automatically exports treatment and occurrence data into a *Darwin Core Archive*, a standard format used by GBIF and other networks to share data from many different sources. This means GBIF can extract these data on the day of the article's publication, making them immediately available to science and the public through its portal and web services, further enriching the biodiversity data already freely accessible through the GBIF network. Similarly, the information and multimedia resources become accessible via EOL's species pages.

One of the main purposes of the partnership is to ensure that such data remain accessible for future use in research. A recent study published in *Current Biology* found that 80% of scientific data are lost in less than 10 years following their creation.

Donald Hobern, GBIF's Executive Secretary, commented: "A great volume of extremely important information about the world's species is effectively inaccessible, scattered across thousands of small datasets carefully curated by taxonomic researchers. I find it very exciting that
this new workflow will help preserve these 'small data' and make them immediately available for re-use through our networks."

"Re-use of data published on paper or in PDF format is a huge challenge in all branches of science", said Prof. Lyubomir Penev, managing director of Pensoft and founder of the Biodiversity Data Journal. "This problem has been tackled firstly by our partners from Plazi who created a workflow to extract data from legacy literature and submit it to GBIF. The workflow currently launched by GBIF, EOL and the Biodiversity Data Journal radically shortens the way from publication of data to their sharing and re-use and makes the whole process cost efficient", added Prof. Penev.

The elaboration of the workflow from BDJ and Plazi to GBIF through Darwin Core Archive was supported by the EU-funded project EU BON (Building the European Biodiversity Observation Network, grant No 308454). The basic concept has been initially discussed and outlined in the course of the pro-iBiosphere project (Coordination and policy development in preparation for a European Open Biodiversity Knowledge Management System, addressing Acquisition, Curation, Synthesis, Interoperability and Dissemination, grant No 312848).
