

I hope this initiative also works as a call to arms. For too long we have been collecting in some of the most diverse and threatened ecosystems on Earth, sorting out our specimens of interest and leaving all the other material sad and lonely in freezers or collection rooms. With the world becoming a global village connected by instant messaging and social media, and a rampant biodiversity crisis, now more than ever is the time to share!



Coming back “home” in speed boats (“voadeiras”) after a long day of work.

Consider budgeting for specimen sorting services in your next expeditions, and see the amount of available material for taxonomic and evolutionary research go up exponentially for the benefit of all of all of us—and the growth of scientific knowledge! ◦

TreatmentBank and Biodiversity Literature Repository: a further steps towards access to the data in scientific publications

By: Donat Agosti, Plazi, Bern, Switzerland;
agosti@amnh.org

Digitization offers great opportunities for taxonomists. Hymenopterists started early in this area, built one of the first online catalogues for a megadiverse taxon (Nature, 2002), obtained the first grant from Smithsonian’s Atherton-Seidall foundation to digitize publications (together with the mosquitoes and *Biologia Centrali Americana* projects) in the pre-Biodiversity Heritage Library era, have been activists calling for open access to taxonomic data for everybody, and the Hymenoptera Name Server is still one of the largest online catalogues for a given large taxon. Time is moving beyond the PDF into sub-article elements and facts of traits. Hymenopterists have been involved in the launch of the first semantically enhanced taxonomic, probably scientific overall, publication Pensoft launched in 2010, using Taxpub, a schema developed together with the National Institutes of Health and Pensoft. Norm Johnson *et al.* (2010) provided at this time the first manuscript produced by a script right off his database.

Norm has also been the first who made use of a follow-up project, the Biodiversity Literature Repository

at Zenodo/CERN, and uploaded the first 4,000 publications. This project started as a side project to TreatmentBank, which has its roots in 2002, in an early binational US-NSF-German DFG project to model taxonomic literature. This lead eventually to the foundation of Plazi in 2008 as the institution that took care of developing the text to data conversion tool GoldenGate, maintains Taxpub, and TreatmentBank, the repository for taxonomic treatments. The Biodiversity Literature Repository had originally the role to host all the PDF we used in the data conversion. But soon, we realized its full power. This collaboration with Zenodo/CERN offers not only almost unlimited repository space—you can add your PDFs here too—the highest degree of probability that the archive will not be sold to a commercial publisher or another private enterprise,



Hymenoptera genitalia illustrations recovered from digitized publications; see text for full description.

but it also mints Digital Object Identifiers (DOI) for free, vital today to cite properly scientific articles and data in our publications. Additionally, they also offered space to store illustrations (and provide DOIs), and encouraged the users to add as many links to related items as possible. With its generous and liberal data policy, articles up to year 2000 can be made open access, and small publishers could obtain a DOI for their new articles. For Plazi, this offered a unique opportunity to follow through not only to extract taxonomic treatments, named entities and

observation records from articles (Agosti & Egloff, 2009), but also the illustrations (Egloff *et al.* 2017). Together with the tools at hand, 16 journals are fully automatically harvested daily. Additionally, all the bibliographic references are extracted, as well as the illustrations. The latter are deposited at BLR, are given a DOI, and they are cited from within the treatments in TreatmentBank.

For example, for 2016, from 72 journals 3,223 articles have been processed resulting in 43,739 taxonomic treatments, including those of 5,051 new species, 16,416 figures extracted and deposited on BLR, and 133,447 bibliographic references.

The value of access to scientific illustrations will become obvious, once the new viewer of the illustrations at BLR will be available, which will look like this view of extracted illustrations on Hymenoptera genitalia. (<https://zenodo.org/communities/biosyslit/search?page=1&size=20&q=aedeagus&keywords=hymenoptera>) ◦

References:

1. Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. *BMC Research Notes* 2: 53 <https://doi.org/10.1186/1756-0500-2-53>
2. Biodiversity Literature Repository. <http://biolitrepo.org>
3. Clarke T (2002) Formidable catalogue puts army of ants online. *Nature* 416, 115 (14 March 2002) <http://dx.doi.org/10.1038/416115b>
4. Egloff W, Agosti D, Kishor P, Patterson D, Miller J (2017) Copyright and the use of images as biodiversity data. *Research Ideas and Outcomes* 3: e12502. <https://doi.org/10.3897/rio.3.e12502>
5. Johnson NF *et al.* (2017) Hymenoptera Online (HOL). [Online] Available from <http://hol.osu.edu> [accessed 25 September 2017]
6. Polaszek A, Masner L, Rajmohana K, Johnson N (2010) World species of the genus *Platyscelio* Kieffer (Hymenoptera: Platygasteridae). *ZooKeys* 50: 97–126. <https://doi.org/10.3897/zookeys.50.485>
7. TreatmentBank. <http://plazi.org/resources/treatmentbank/>

Australian Hymenoptera evolution: connecting digitisation and research at the Australian National Insect Collection

By: Dr. Juanita Rodriguez and Nicole Fisher, Australian National Insect Collection (ANIC)-CSIRO; Juanita.Rodriguez@csiro.au, Nicole.Fisher@csiro.au

The Australian National Insect Collection (ANIC) is recognized both nationally and internationally as a major research collection. Being the world's largest collection of Australian insects and related groups (*e.g.*, mites and nematodes) we house over 12 million specimens one fourth of which are hymenopterans.

We are currently working on a number of rapid digitization projects comprising bees (~50,000 specimens), Pompilidae (~5,000 specimens) and Mutilidae (12,000 specimens). Funding coming from Bush Blitz (Australia's largest nature discovery program) and CSIRO has allowed us to bring a team of 10 casuals and volunteers who are overhauling the curation, imaging and databasing of over 300 collection drawers. Through DigiVol, a crowdsourcing transcription center, which is used in Australia and globally, we are able to retrieve information from the labels transcribed by volunteers.



Juanita collecting specimens in Lago Yelcho, Chilean Patagonia (2017).

During the initiation of these digitisation projects, the main emphasis was on databasing and exposing the collection to research opportunities, and it seems our digitisation efforts are paying off. For example, over 20 drawers of Mutilidae specimens collected in the early 1980s had been left to sit for over 35+ years identified only to family level, and thanks to digitisation they are now finding their way to scientific discovery. We have shared image files with overseas taxonomy experts, like Kevin Williams at the California Department of Food and Agriculture (CDFA), who can use the images to identify specimens as best as possible. Researchers can also look at specimen images and decide which specimens to ask on loan.



ANIC specimens curated and digitised ready for research.

But not only have these images served for identification and curation purposes, current research in ANIC is