

EDIT's Taxonomy and Biodiversity Information Services for Biogeographers

In recent meetings, the opinion has grown that the burgeoning field of biodiversity informatics was being led, perhaps dominated, by descriptive taxonomists. This came as something of a surprise to many of the leading players in the field, so here we want to explore how a taxonomic initiative may have real benefits for other fields, like biogeography.

The European Distributed Institute of Taxonomy (EDIT; <http://www.e-taxonomy.eu/>) is a consortium of 29 natural history institutions in Europe and beyond. It works to provide better, faster and more accountable tools to taxonomists, to significantly accelerate the global production of taxonomic knowledge. We are currently just starting the final year of a 5-year project.

We organise year-round All-Taxa Biodiversity Inventory and Monitoring programmes, we develop an Internet Platform for Cybertaxonomy with software for the management of taxonomy and biodiversity data, we create collaborative Scratchpad websites to empower research communities across the globe and we improve the training of the next generation's taxonomists. All this is in addition to helping the most powerful taxonomic institutions in the world work better together.

In this article we try to describe what the biogeographical community can gain by partnering with the recent developments in computer-aided taxonomy. It seems obvious to us that both sides can only grow from such a partnership: the increasing applied importance of taxonomy makes it rely more and more on geographical data, and the teeming mass of such data would be very hard indeed to navigate for biogeographers without an adequate taxonomic backbone.

Cybertaxonomy Platform tools

<http://cybertaxonomy.org/>

Part of EDIT's work is to develop a series of tools to accelerate and simplify scientific work. These are originally focused on taxonomy, but many of

them can be of tremendous use to biogeographers. The Cybertaxonomy Platform has now reached an open-release stage, and can be accessed at <http://cybertaxonomy.org/>.

All of the tools are free and developed using an open source process. They can be downloaded on any platform and used in the field, with or without an internet connection. We aim at providing a comprehensive, robust, intuitive platform to support scientific analysis, respecting international standards, but without burdening the user with technical details. The system works with minimal components to install. The user interface should not change dramatically if new technologies or standards are adopted. This allows researchers to use these tools on a daily basis. Here we offer a selection of tools with a biogeographical interest.

1) *EDIT MapViewer* (<http://edit.csic.es/geo/mapviewer/edit.html>).

We are aware of the benefits a taxonomist can get simply by visualizing his georeferenced records on a screen along with other geographic data. It not only enables you to get a good overview of your data, but also to discover distribution patterns; moreover, with the appropriate analysis tools, you can visualize and catch the probable location of well and insufficiently surveyed regions.

EDIT mapViewer allows you to do it without installing any software: simply access the website, upload your point data, create your maps and print them in high-quality (maximum 600 dpi) (see Figure 1). You can also get the legend of your symbolized points as well as the main map in different formats (JPEG, TIFF, PNG...) and colour or black and white.

Other utilities include: browse GBIF data, perform some spatial analysis, query your data, draw your own polygons, choose from different projections, etc.

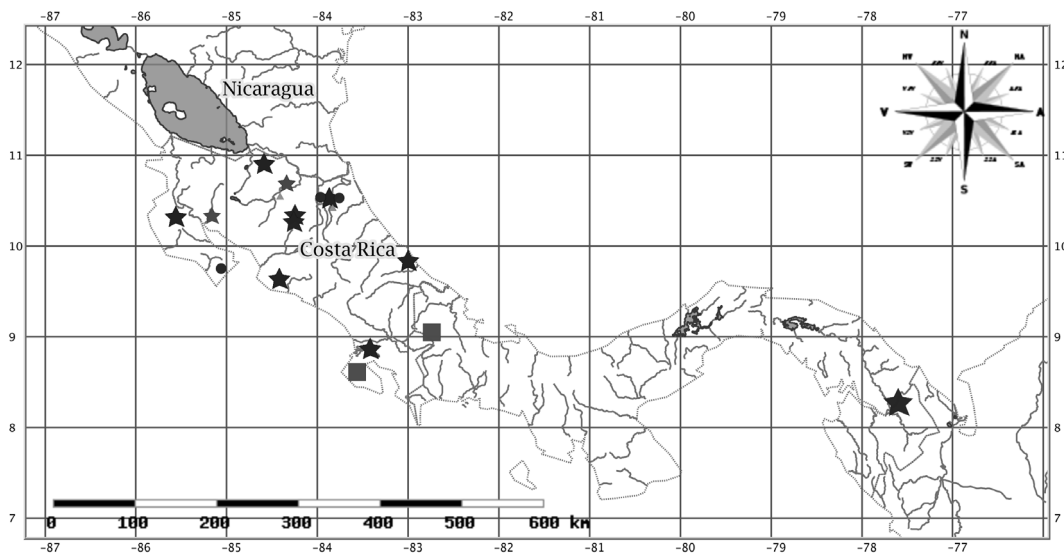
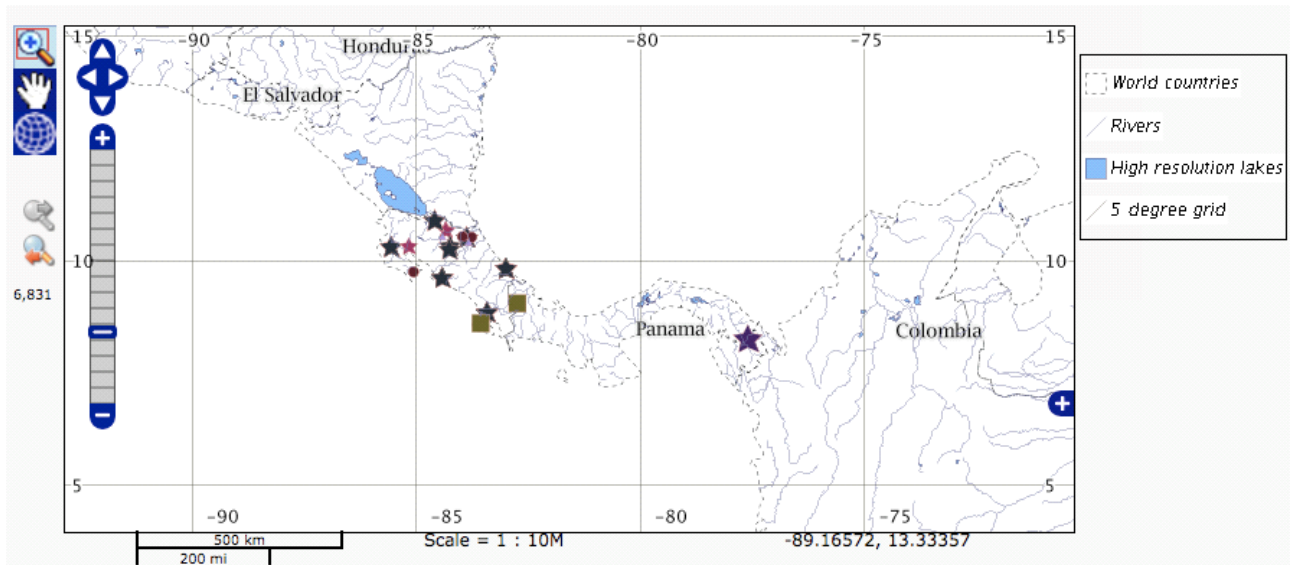


Figure 1. Partial screenshot of the EDIT mapViewer interface (up) and high-resolution (480 dpi) map generated with it (down).

2) EDIT mapRest Services (<http://dev.e-taxonomy.eu/trac/wiki/MapRestServiceApi>).

These services provide distribution maps from a URL (Uniform Resource Locator). Users may choose among a large range of symbolization parameters. To use these services, the user simply needs to construct the URLs and the mapRest Services generate a map with the selected features, symbolized as the user chooses.

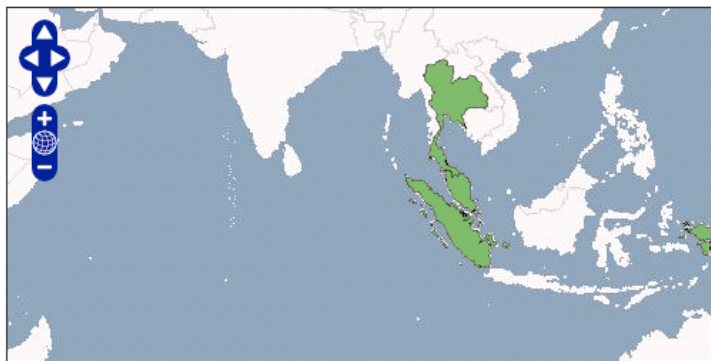
MapRest Services can return an image or a file that can be used by a client-side webmapping application (like Openlayers). Scratchpads as well as some EDIT dataportals (PalmWeb, <http://dev.e-taxonomy.eu/dataportal/palmae/> and Cichorieae, <http://dev.e-taxonomy.eu/dataportal/cichorieae/>) use mapRest services to map distributions of taxa (see Figure 2).

Point data can also be plotted from coordinates, using mapRest Services. Some All-Taxa Biodiversity Inventories (ATBI) sites like Mercantour/Alpi Marittime and Gerner are currently using it. More examples are available on <http://dev.e-taxonomy.eu/trac/wiki/MapRestServiceExamples>.

These two tools differ mainly in the degree to which human intervention is involved in handling the data. While EDIT mapViewer is specially designed for taxonomists (who usually don't know about Geographic Information Systems), EDIT mapRest Services are suitable for web-developers working for taxonomic institutes, interested in automatically creating maps that are based on user request.

Distribution

A single species, *Actinorhysis calapparia*, native to New Guinea and the Solomon Islands, now widespread as an ornamental or ceremonial plant in Southeast Asia., Malaya, New Guinea, Solomon Is., Sumatera, Thailand



Biology And Ecology

In the wild, it grows in lowland tropical rain forest at altitudes up to about 1000 m above sea level.

Common Name

Pinang penawar, pinang mawar.

Figure 2. PalmWeb screenshot using mapRest services.

3) Species Distribution Modeling.

Species distribution modeling using sampling locations and associated environmental information is commonly used in the fields of biogeography and ecology. There is a growing interest in using museum and botanical garden specimens for this type of modeling. One of the issues identified is that specimens from Natural History collections have often been collected in a qualitative rather than quantitative way, contrary to ecological studies. Therefore, typically, museum data provide information on the presence of species, but not on their absence, which is needed in most modeling algorithms. By assessing ongoing modeling activities using collection specimens, tools like MAXENT and GARP have been tested for the Geospatial components of the EDIT cyber taxonomy platform. The tests showed that they were inappropriate when there was insufficient availability of environmental data to feed into the systems. Typically for museum specimens contemporary environmental records with the collecting date are rare or need to be inferred from secondary sources or third parties modeling results. For an atlas of the birds of the Comoro Islands (Louette et al. 2008), the system of ecological envelopes around the grid where the bird was observed was used, with success, to delimit the potential ranges of distribution. Parameters used were altitude, forested versus non forested areas, rainfall, presence of rivers, lakes, villages and roads.

In conclusion, for any distribution modeling activity, is the proper collection of the specimens and recording of the site information that determines the quality of the end results. For legacy data, proper post-georeferencing, data cleaning and controls have to be performed following rigorous procedures adapted to the nature of the available data. See the publications about these topics at GBIF training resources (<http://www.gbif.org/participation/training/resources/gbif-training-manuals/>) and the georeferencing pages of the VertNet network (<http://vertnet.org/georeferencing/georeferencing.php>). The lessons learned by these networks and their recommendations were taken into account at every step of the implementation of the EDIT Geospatial Services.

The development of the system of Cybertaxonomy Platform tools within EDIT follows the Open GIS Consortium (OGC) and EU Directive INSPIRE standards to guarantee interoperability, long-term maintenance and usability of the services beyond the EDIT funding period. Information about these tools, including possibilities for reviewing, can be found at <http://www.bdtracker.net/>. BDTracker is an EDIT collection of openly available software with particular interest for researchers in biology and taxonomy.

Scratchpads (<http://scratchpads.eu/>)

Integrating the producers, users and sources of digital information is perhaps the greatest challenge facing biodiversity informatics. Scratchpads address this challenge by blending social, technical and policy developments into a platform supporting collaborative biodiversity research.

This data-publishing framework allows groups of people to create their own social networks supporting their research communities. Scratchpads are flexible and scalable enough to support multiple networks, each with its own choice of features, visual design, and data. Diverse and distributed data can be organised, curated and cited by users around multiple taxonomies. The Scratchpad framework currently serves more than 1,100 registered users across 100 sites, spanning academic, amateur and citizen-science audiences. These contributors have generated almost 160,000 pages of content in the first two years of use. Web services on standardised data elements are available for use by related initiatives such as GBIF, Encyclopedia of Life and the Biodiversity Heritage Library.

Technically, and you really don't need to know this, Scratchpads are an implementation of the Drupal content management system, which

Scratchpads in a nutshell

- * We give you a web site.
- * You fill it with stuff.
- * The site and the data remain yours, not ours.

means that the actual information is held in a MySQL database and served onto web pages as required. What you do need to know is that to create one, the best place to start is to upload a taxonomy, which is either your own encoded onto a spreadsheet, or can be selected from Catalogue of Life or other on-line sources. Either way, it should be a quick and easy step. Once the taxonomy is in place it becomes the central organiser for everything else. You can upload content from word processor files by cut-and-paste or import from spreadsheets or other databases. You can create data tables to contain any kind of data that you like. You can upload images of course. The key to this is that all information elements are tagged with keywords that allow them to be associated on a page. The default tagging is the taxon name, which is automatic where possible, but other tags, such as locality or habitat can be added as needed. The objective of the Scratchpad project is to make the workflow straightforward and easy (Figure 3).

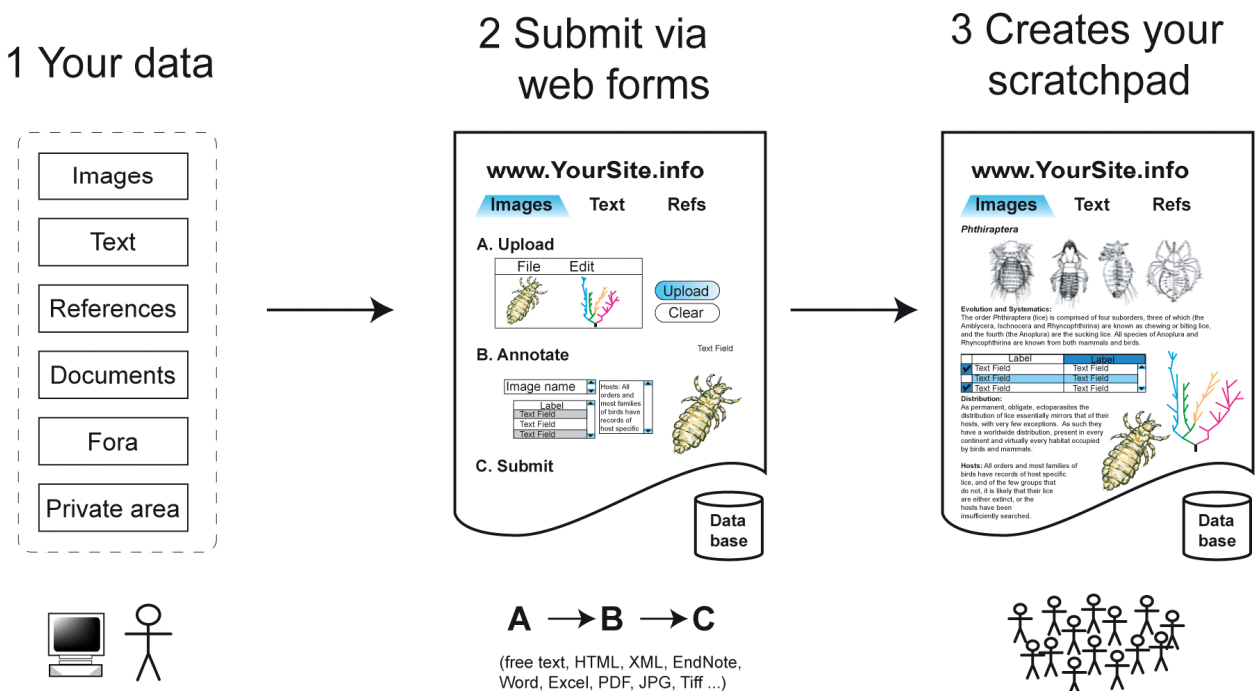


Figure 3. PalmWeb screenshot using mapRest services.

To encourage community-building, the sites are created empty and unbranded, so that each community can develop its own image. Design and layout are entirely up to you, although there are customisable templates available so that the process can be straightforward and little more than choosing a logo or an iconic image. You provide the content, which remain yours, published under a Creative Commons Licence.

The essence of collaboration is data sharing, hence the Creative Commons licence. The maintainer can control who can see what, can establish private groups within the community and can expose as much of this to public view as desired by the community. Public pages are most commonly a composite of various elements related by some tag, say a species name. These elements are arranged into a display and the resulting page can be cited by clicking a button that creates a Chicago-style bibliographic entry containing a url. The authorship is a list of those who have contributed data to the page being cited and the url points to a static archive of the page at the time the button was pressed. This means that if you cite the page in an article, the reader will see exactly the same page that the author saw, not a current view con-

taining the latest data. Clearly, if your page is cited, you collect merit points in the citation statistics but we are also working with GBIF to provide usage statistics that will measure the impact of your on-line data. Finally we have a publication interface (TAPIR) that makes your data visible to other machines on the web in a form that will give you credit for your work. This interface will be expanded to offer more data types during the coming year.

To apply for a site (<http://scratchpads.eu/>, Figure 4) you will need only to specify the scope, either taxonomic, geographical, habitat or a combination, such as the Lichens of Bermuda. We do not mind what subject you choose, provided it is broadly to do with natural history.

You provide the technical expertise to run the site. This is usually one IT-savvy person within a community who takes responsibility, the maintainer, but controls access and editing rights to all other users, including the ability to give others within the community editorial or even maintainers rights. There is a suite of tutorial videos on the main Scratchpad site, under the Help tab, that describe most of the common tasks involved in maintaining your site.

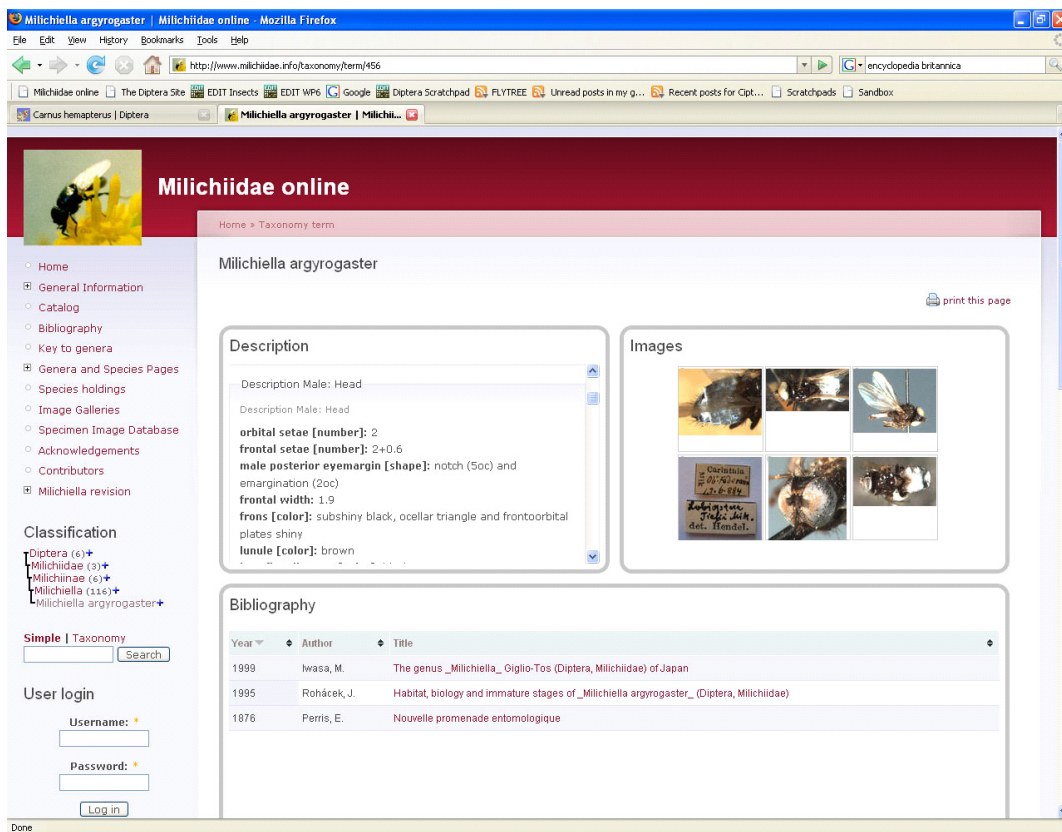


Figure 4. A sample Scratchpad.

We provide the infrastructure that keeps the site visible and we back it up. We are a very small team, so we do not offer much direct support, but we do run a forum and we do respond to bug reports and feature requests.

It's free and guaranteed to remain so for at least the next 6 years. You can download and withdraw your entire site content at any time as a MySQL dump, should you wish to migrate your content to another site, for instance, or a meteorite strikes the Natural History Museum in London. That is as far into the future as we can see right now. If your interest has been piqued by this, then visit <http://scratchpads.eu/> and try the sandbox. This is a demonstrator site that is re-built every hour, so it is impossible to break. For more than 59 minutes, anyway.

If you would like to read a longer article describing the Scratchpads from a taxonomic perspective, see Roberts et al. (2008).

References

- Louette, M., Abdérémane, H., Yahaya, I., & Meirte, D., eds. (2008) Atlas des oiseaux nicheurs de la Grande Comore, de Mohéli et d'Anjouan. Studies in Afrotropical Zoology, n° 294. Musée royal de l'Afrique centrale, Tervuren, Belgium.
- Roberts, D., Rycroft, S.D., Brake, I., Harman, K., Scott, B., & Smith, V.S. (2008) Getting Taxonomy onto the Web. *The Systematist*, 30, 3-10.

Patricia Mergen

Royal Museum for Central Africa, Belgium
e-mail: patricia.mergen@africamuseum.be
<http://www.africamuseum.be/research/cooperation>

Dave Roberts

Natural History Museum, London, UK
e-mail: dmr@nomencurator.org
<http://scratchpads.eu/>

Pere Roca

Museo Nacional de Ciencias Naturales (CSIC),
Madrid, Spain
e-mail: peroc79@gmail.com
<http://edit.csic.es/>

Gaël Lancelot

Musée National d'Histoire Naturelle, Paris,
France
e-mail: lancelot@mnhn.fr
<http://www.e-taxonomy.eu/>

Remember that being a member of IBS means you can get free online access to four biogeography journals: **Journal of Biogeography**, **Ecography**, **Global Ecology and Biogeography** and **Diversity and Distributions**. You can also obtain a 20% discount on the journals **Oikos** and **Journal of Avian Biology**.

Additional information is available at <http://www.biogeography.org/>.