Blumea 54, 2009: 13–15
www.ingentaconnect.com/content/nhn/blumea

RESEARCH ARTICLE

doi:10.3767/000651909X474032

Interactive web-taxonomy for the *Araceae*: www.cate-araceae.org

A. Haigh[1], S.J. Mayo[1], T. Croat[2], L. Reynolds[1], M. Mora Pinto[3], P.C. Boyce[4], L. Lay[1], J. Bogner[5], B. Clark[6], C. Kostelac[2], A. Hay[7]

Key words

Araceae
CATE
taxonomy
web revision

Abstract CATE (Creating a Taxonomic E-science) is a pilot project funded by the UK Natural Environment Research Council (NERC) to test a model of internet taxonomy which aims to construct and maintain online a full descriptive taxonomic revision as a collective enterprise carried out by the specialist taxonomic community. The software application includes the functionality to allow taxonomists to make contributions and proposals for change that are passed for peer review to an editorial and moderating body drawn from the taxonomic community. The model is being tested on the Hawkmoths (*Sphingidae*) and Aroid (*Araceae*) families. The paper describes the aims of the project and current progress on the *Araceae* e-revision.

Published on 30 October 2009

THE CATE PROJECT

CATE (Creating a Taxonomic E-science) is a 3-year project, in progress since December 2005, to model a particular approach to implementing α-taxonomy on the internet, using two families of organisms, the *Sphingidae* (Hawkmoths, http://www.cate-sphingidae.org) and the *Araceae* (Aroids, http://www.cate-araceae.org). It is funded by the UK's Natural Environment Research Council (NERC) under their E-Science Programme and is a consortium led by the Natural History Museum London, with the Royal Botanic Gardens Kew and the University of Oxford. The Missouri Botanical Garden plays a key role as a major institutional collaborator for the *Araceae* website.

The basis for the project was first clearly articulated in a paper by Godfray (2002) that described the challenges facing taxonomy in the internet age and suggested that the future for taxonomic revision lay in exploiting the internet, not only as a means of communicating taxonomic information, but also to carry out many of the operations of taxonomic research. An important idea in Godfray's vision is the notion of peer-reviewed consensus taxonomy, i.e. the expert community agree on a version of the taxonomy of a group of organisms as a dated 'edition' that holds for a specified period. The consensus is then later replaced by an updated version resulting from changes submitted and peer-reviewed before being uploaded. This concept of consensus taxonomy implies that the taxonomists cooperate to produce, maintain and update, on a permanent basis, the consensus taxonomy website. The justification for consensus taxonomy is the demand from a wider public, including other scientists, for easily accessible, authoritative taxonomic treatments of organisms. Godfray argued that by providing strict, dated versioning and the technology for continual updating of the taxonomic revision, the societal need for 'standard' taxonomies could be reconciled with the inevitable and necessary changes brought about by continual growth in scientific understanding of biodiversity and its taxonomic units. More recent discussion of developments in internet taxonomy are provided by Godfray et al. (2007) and Scoble et al. (2007).

Consensus taxonomy leads to a number of requirements for the software system, such as the need for:

1. an editing system to enable taxonomists to make specific proposals for change to a web-implemented taxonomic treatment;
2. a versioning system which enables the smooth and regular production of a new and precisely dated versions of the taxonomy;
3. a system to facilitate the peer-review of new proposals for change;
4. functionality to present alternative taxonomic views. Consensus taxonomy also creates the need to tackle some challenging human and professional issues.

The taxonomic community needs to become sufficiently organized that an editorial body can be formed, but the human organization supporting the website must also be open, transparent, and self-renewable. It must attract the next generation of taxonomists but also secure the allegiance and commitment of the established specialists. Rewards and credit must be of sufficient value to compete with those of conventional publication in career evaluation. The website must also be able to attract non-professional biologists and natural historians who have expertise and important information to contribute.

The CATE project was set up to test as much of this vision as possible as a feasibility study. The three main aims are to

1. build the websites for each family, to include about 1300 species each and upload edited content from existing literature;
2. build a software framework for on-going revision of the uploaded information by the relevant taxonomic community;
3. attract the taxonomic community to form the human framework to sustain, update and develop the consensus revision.

[1] Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AE, United Kingdom.
[2] Missouri Botanical Garden, P.O. Box 299, St Louis, MO 63166-0299, USA.
[3] Department of Biological Sciences, Box 870345, The University of Alabama, Tuscaloosa, AL 35487-0345, USA.
[4] Malesiana Tropicals, Suite 9-04, Tun Jugah Tower, No. 18, Jalan Tunku Abdul Rahman, 93100 Kuching, Sarawak, Malaysia.
[5] Botanischer Garten München-Nymphenburg, Menzinger Str. 65, 80638, München, Germany.
[6] Department of Zoology, University of Oxford, South Park Road, Oxford OX1 3PS, United Kingdom.
[7] National Herbarium of New South Wales, Royal Botanic Gardens, Mrs Macquaries Road, Sydney, New South Wales 2000, Australia.

THE SOFTWARE

The design of the software is based on existing taxonomic information models such as Taxon Concept Schema (TCS, http://www.tdwg.org/activities/tnc/tcs-schema-repository/), Structure of Descriptive Data (SDD, http://www.diversitycampus.net/Projects/TDWG-SDD/index.html), Access to Biological Collections Data (ABCD, http://www.bgbm.org/TDWG/CODATA/Schema/) and the Common Data Model (CDM, http://wp5.e-taxonomy.eu/EDIT-Architecture.html). The aim is that the CATE software application will be generic, i.e. usable for other plant and animal groups and the software architecture is database agnostic, i.e. it will work over a variety of different database formats and structures; the data is decoupled from on-screen delivery, which allows very flexible presentation. LSIDs (Life Science Identifiers) are being implemented in the software to make it possible to integrate easily with global web-based biodiversity systems such as GBIF (Global Biodiversity Information Facility, http://www.gbif.org) and Encyclopedia of Life (EoL, http://www.eol.org). LSIDs are electronic tags that are unique for every taxon concept and make it possible for different systems and databases to share metadata linked to the same taxon concept (e.g. a species name used in a particular version of the consensus web revision). The CATE project began just prior to the international European Union-funded EDIT (European Distributed Institute of Taxonomy), a Network of Excellence project designed to stimulate greater integration of European taxonomic institutions with the object of providing more effective delivery of taxonomic information and expertise to problems of biodiversity conservation. CATE is working closely with a key part of EDIT, the software initiative entitled 'Platform for Cybertaxonomy' (http://wp5.e-taxonomy.eu/EDIT-Architecture.html), and this collaboration will help to ensure generic interoperability between CATE and other systems used and promoted by EDIT.

The target audience envisaged consists of two broad user types; taxonomists who both produce and need access to information, and non-taxonomists whose primary interest is to obtain access to information. The management of the e-taxonomy website (the consensus taxonomy and any alternative hypotheses) will be divided into two main areas: content management handled by the taxonomic community through an editorial body, and software management which because of the costs involved in server hosting and maintenance is probably only viable for institutions such as natural history museums and herbaria.

The software aims to make it possible for any user to propose changes to the revision through an online procedure analogous to electronic submission of scientific manuscripts. These will be reviewed by members of the interested community, both as unsolicited responses (the proposals will be visible on the site) and as reviews specifically sought from expert moderators. The coordination of this process will be in the hands of editors also drawn from the taxonomic community. If accepted, proposals will be incorporated into the next version of the consensus revision. Otherwise, they will remain on the site visible as alternative hypotheses, attributed clearly to their authors. Other kinds of contribution, such as less complex observations and images not implying taxonomic changes, are also strongly encouraged and the protocols for this are already being implemented on the two taxon websites. This provides a way for a larger pool of biologists and natural historians, in addition to taxonomists, to contribute information, images or ideas that do not require community approval, but which are interesting and help to build a better picture of the taxa. This type of content can potentially be incorporated into future versions of the consensus, whenever the editorial body considers this desirable.

The most characteristic feature of the CATE project is thus its ambitious goal to provide software to make it possible for the taxonomic community to carry out editing and peer-review of the consensus e-revisions online.

THE *ARACEAE* WEBSITE

The foundation for the *Araceae* site is the framework provided by the World Checklist of Monocotyledons (www.rbgkew.org.uk/wcsp/monocots), in which a comprehensive checklist of currently accepted species of the family is incorporated. The list and associated information (synonyms, literature references, geographical and ecological notes) were transformed into web pages, one for each accepted species. This created a complete set of species web pages for the family with minimal information content.

The next stage of the process was to create web pages for all the genera and these were based on the generic monograph of Mayo et al. (1997) with some recent additions and changes. A new identification key was written especially for the genera using the software LUCID3 (http://www.lucidcentral.com/lucid3/); this is the first interactive illustrated key to all genera of *Araceae* and is now online.

Building web pages for the more than 3 000 species of *Araceae* represents a bigger challenge. The strategy of the project is to create the basic taxonomic 'substrate' on which the taxonomic community can work, gradually improving the content through successive versions. So the primary goal of the project's content team is to upload acceptable existing descriptive taxonomic information for as many species as possible during the current period of funding (ends in December 2008). The first pages mounted covered the genus *Arum*, based on Peter Boyce's monograph (Boyce 1993), and the next two genera, currently in progress, are the two biggest, *Anthurium* and *Philodendron* — the work on these latter genera depends heavily on collaboration with Dr Tom Croat (Missouri Botanical Garden) and Marcela Mora (University of Alabama). Treatments of Asian genera are also being provided in collaboration with Peter Boyce, Wong Sing Yen, and Dr Alistair Hay, who are deeply involved with the monograph of *Araceae* for the Flora Malesiana. Pages for species of *Pothos*, *Hapaline* and *Colocasia* are now online and more are expected in the coming months.

The project has committed itself to the preparation of illustrated identification keys in LUCID3 software for all the species of three genera (*Arum, Anthurium, Philodendron*). The *Arum* key served as a pilot for the larger genera; LUCID3 key construction requires building a matrix of characters and taxa and the project provided an opportunity to experiment with illustrating character states (mostly as vector image diagrams). For the very large genera *Anthurium* and *Philodendron*, keys are an essential tool for the user, but in neither case have comprehensive keys been attempted in modern times. The preparation of the key matrices is time-consuming, requiring close collaboration with specialists and frequent checking and consultation of herbarium material. The aim is not 100 % success in identifying a single species, but rather to narrow down the field of choice to 10 species or less, which can then be searched using the web pages. It should be borne in mind that at this initial stage, the CATE website is not a new revision of the genera of the family, but an attempt to upload the current state of knowledge and provide useful tools to navigate and improve the content as an ongoing collective revisionary process. The keys play a vital role both in navigating the taxonomy and in providing users with basic means of identification. Our intention is that all genera will eventually have interactive keys provided on the website.

The main approach of the *Araceae* website has been from a monographic viewpoint, but it is becoming increasingly clear that many taxonomists will find it most convenient to take a regional perspective in contributing information. This aspect of the web revision highlights another innovative aspect of e-taxonomy. If the taxonomists of a family like *Araceae* can agree to work within a single overall framework, such as the species-page foundation already implemented on the CATE website, then every floristic treatment will directly improve the global taxonomy and be immediately accessible to all. An opportunity arose to work from a floristic perspective in September 2007, when Lucinda Lay joined the Kew team for 12 months to work on species pages and an interactive key to the African species, estimated at 190 taxa. We hope that other colleagues will be able to contribute floristic subsets of species information in due course, from various parts of world.

In later stages the project has developed a parallel website (Araceae Network, http://scratchpad.cate-araceae.org/) to act as a forum for taxonomists and others interested in contributing to the CATE site. Araceae Network was made possible because of the EDIT project mentioned earlier, which provides opportunities for any taxonomist or group of taxonomists to set up their own e-taxonomic site, hosted by London's Natural History Museum (see http://www.editwebrevisions.info/scratchpads/). The justification for such a forum is that of the need to bring about a supportive environment on the internet through which the taxonomic community can debate and reach agreement on topics which are important for developing a consensus taxonomy. These include technical subjects like character state terminology and organizational ones such as how the editorial and review roles needed for the CATE website can be legitimately and fairly assigned so as to promote a progressive and inclusive climate for taking the web revisions forward into the future.

REFERENCES

Boyce PC. 1993. The genus Arum. HMSO/Royal Botanic Gardens, Kew.

Godfray HCJ. 2002. Challenges for taxonomy. Nature 417: 17–19.

Godfray HCJ, Clark BR, Kitching IJ, Mayo SJ, Scoble MJ. 2007. The web and the structure of taxonomy. Systematic Biology 56: 943–955.

Mayo SJ, Bogner J, Boyce PC. 1997. The genera of Araceae. Royal Botanic Gardens, Kew.

Scoble MJ, Clark BR, Godfray HCJ, Kitching IJ, Mayo SJ. 2007. Revisionary taxonomy in a changing e-landscape. Tijdschrift voor Entomologie 150: 305–317.

Websites:

http://scratchpad.cate-araceae.org/
http://wp5.e-taxonomy.eu/EDIT-Architecture.html
http://www.bgbm.org/TDWG/CODATA/Schema/
http://www.cate-araceae.org
http://www.cate-sphingidae.org
http://www.diversitycampus.net/Projects/TDWG-SDD/index.html
http://www.editwebrevisions.info/scratchpads/
http://www.eol.org
http://www.gbif.org
http://www.lucidcentral.com/lucid3/
http://www.rbgkew.org.uk/wcsp/monocots
http://www.tdwg.org/activities/tnc/tcs-schema-repository/