e-Flora Malesiana: state of the art and perspectives

M.C. Roos[1,2], W.G. Berendsohn[3], S. Dessein[4], T. Hamann[1],
N. Hoffmann[3], P. Hovenkamp[1], T. Janssen[4], D. Kirkup[5],
R. de Kok[5], S.E.C. Sierra[1], E. Smets[1,7],
C. Webb[6] and P.C. van Welzen[1]

[1]Netherlands Centre for Biodiversity – Naturalis,
Section National Herbarium of The Netherlands, Leiden University,
PO Box 9514, 2300 RA Leiden, The Netherlands
[2]roos@nhn.leidenuniv.nl (corresponding author)
[3]Freie Universittät Berlin - Botanic Garden and Botanical Museum Berlin-Dahlem
Königin-Luise-Str. 6-8, DE-14195 Berlin, Germany
[4]The National Botanic Garden of Belgium
National Botanic Garden of Belgium, Domein van Bouchout, BE-1860 Meise, Belgium
[5]Royal Botanic Gardens, Kew
Richmond, Surrey TW9 3AE, U.K.
[6]The Arnold Arboretum, c/o Harvard University Herbaria
125 Arborway, Boston, MA, 02130-3500, United States of America
[7]Laboratory of Plant Systematics, K.U.Leuven
Kasteelpark Arenberg 31 box 2437, BE-3001 Leuven, Belgium

ABSTRACT. An overview is presented of available e-taxonomic products and ongoing projects contributing to Flora Malesiana. This is presented in the context of a strong plea to strengthen the implementation of state-of-the-art e-taxonomy tools to speed up the generation and publication of Flora Malesiana information.

Keywords. EDIT, effective collaboration, European e-Floras Initiative, Flora Malesiana

Introduction

The increasing array of electronic taxonomy tools available for the elaboration and dissemination of floristic information has brought many advantages. It enables a shift from the traditional Flora concept as a static, printed account to a dynamic and interactive format, allowing for rapid updating and multiple uses of information. It provides taxonomists with means for instant interactive and remote co-operation, including continuous processes for evaluating preliminary results by peers and updating existing information. It also allows taxonomists to forge better links with their user-communities by making the products of their research more tailor made and accessible via internet (e.g., identification lists, specimen databases to monographs, biodiversity and Non-Timber Forest Products assessments, and analyses of spatial patterns of biodiversity).

In order to strengthen and speed up Flora Malesiana activities, Roos & Hovenkamp (2009) suggested that the Flora Malesiana community needed to adopt a more pragmatic and flexible attitude—flexible in terms of formats and publication strategy, and standards of robustness and confidence in the results.

The Flora Malesiana project has recently gained new momentum by starting a new website (www.floramalesiana.org), by adopting e-taxonomy tools and by joining a broadly supported initiative that wants to promote the adoption of richly interactive and truly collaborative systems for the production and presentation of floras: The European e-Floras (enhanced-Floras) initiative.

The European e-Floras initiative

Based on the results of an EDIT e-Flora Platform Workshop organised in Leiden (January 2010) the Board of the Foundation Flora Malesiana has decided during the most recent board meeting in Singapore to work towards an e-Flora Malesiana platform, to be developed in co-operation with the European e-Floras initiative. Participating institutions of the EDIT workshop in January also discussed the possibility of starting collaboration activities between Flora projects and the EDIT Platform for Cybertaxonomy by means of taxonomic exemplar groups (http://www.e-taxonomy.eu/node/748).

During a second workshop in Berlin (March 2010) three exemplary groups were set up: Flore d'Afrique centrale (National Botanic Garden of Belgium; Dessein et al., 2010; http://www.br.fgov.be/RESEARCH/DATABASES/FOCA/index.php), Flora Malesiana (Netherlands Centre for Biodiversity; http://160.45.63.201/dataportal/preview/flora-malesiana) and Kew's African Floras (Royal Botanic Gardens Kew).

During a third workshop organised in Brussels (September 2010) participants agreed to establish an "European e-Floras Initiative" to enhance communication between (e-)Flora authors, e-taxonomy initiatives and users of content and thus to increase research efficiency, reduce redundant efforts and speed up the preparation of up-to-date, high-quality content on plant biodiversity in formats permitting rapid updates and multiple uses (http://www.e-taxonomy.eu/node/859).

At present, the European e-Floras initiative is supported by 23 leading institutes from Europe, Asia, Africa and Australia involved in Flora production.

Why an e-Flora Malesiana?

The Flora Malesiana progress has lost some of its momentum since the beginning of the century. The overall progress in terms of the number of species covered is too slow (about 75% of all species in Malesiana still need to be treated) to make a meaningful contribution towards sustainability and conservation. Reasons for this are (i) the decreasing number of taxon experts that are available and devoted to write taxonomic treatments; (ii) the scattered generation of information and its particular

usage (i.e., just for one specific project and format) in differently oriented projects; (iii) duplication of efforts; and (iv) different formats. In order to use the available expertise and resources effectively, a shift in the conventional work processes is needed.

Since 1950, Flora Malesiana has published its products as family treatments (in total 204 up to now). At present, several volumes are no longer available as hard copy (Series I Volume 1, 4–6, 9 Part 2–3 and Series II Volume 1 Part 1–3). Only the most recent volumes are available in electronic formats that can be more easily imported to a database than printed texts:

Text: Series I Volume 13;

Colour images: Volume 14;

Digital files: Volume 15 onwards, for Series II only Volume 3 completely available as digital files;

CD-ROMs: Leguminosae: Caesalpinioideae and Mimosoideae, Orchids of New Guinea Vol. I–VI, Orchids of the Philippines Vol. I.

An e-Flora will complement printed formats with electronic editions, offer unlimited access and instant updates, and increase the cost-effectiveness substantially by streamlining the production work flow. It will allow structured data entry, interactivity, multimedia, and enhanced accessibility. Moreover, it will also allow collaboration between several e-Flora projects (e.g., consortium members of the European e-Floras Initiative), users of taxonomic data, and other databases. This increases the possibilities for innovative scientific co-operation, also between other research fields (ecology, ethnobotany, etc.) and attracts worldwide contributors using services offered by the e-Flora.

How can FM work be strengthened?

Flora Malesiana (FM) work can be strengthened by increasing its global accessibility, facilitating efficient remote collaboration, making use of databases to safeguard data, changing the work flow in data preparation and presentation, and creating institutional commitment.

1. Increasing FM global accessibility

FM Website. The visibility of FM on the World Wide Web will be increased by bringing available information and data online via its website (www.floramalesiana. org). The current website will gradually be brought up to date both in design and contents, and will be established as the information exchange portal for FM.

It will feature information on the latest FM meetings (including web feed formats used to publish frequently updated works), and the FM Editorial Committee, collaborators, and other contributors. Furthermore, it will provide access to the FM e-Flora, and also include links to other relevant e-Floras, e-initiatives and/or databases, e.g., Biodiversity Heritage Library (BHL), Cyclopedia of plant collectors, European Distributed Institute of Taxonomy (EDIT), Creating a Taxonomic e-Science (CATE),

Journal Storage (JSTOR), specimen databases (e.g., plants.jstor.org), etc. FM printed volumes and CD-ROMs will be available through the website.

Interactive Key to the Malesian Seed Plants. The Interactive Key to the Malesian Seed Plants is a user-friendly electronic identification aid (DELTA key) for the plants in SE Asia. The first version was published on the web (www.kew.org/herbarium/keys/ fm) and in CD-ROM format in 2004. It covers all seed plant families of the Malesian region, and is supported by c. 1000 pictures and family portraits. In the second version, the plan is to enlarge the key by including the plant families from Thailand and Indo-China, by including all genera done in the Flora Malesiana as possible answers, and by linking the key to the electronic version of the Flora.

List of remaining families. It is imperative that contributors and users can have access to available FM information and data. An inventory of the families still not allotted to (teams of) specialists and work that has been done and still needs to be done for Flora Malesiana (in terms of genera treatments) is being made available through the FM website. It will provide an overview of the different contributors, the working teams and the taxa they work on (with an indication how well verified or preliminary they are). The inventory will be periodically updated by the different contributors via the website contact (roos@nhn.leidenuniv.nl).

Checklist. For missing treatments a checklist of data will be made available by extracting the necessary data from the Kew World Checklist. This will form a backbone by incorporating missing parts of the FM area, provide an overview of missing taxa, connect information, and encourage people to start working on those taxa. It will also give an overview of the resources on the FM area that are still needed.

FM Bulletin. The FM Bulletin (http://www.nationaalherbarium.nl/fmbull/biblio.htm) contains the bibliography on Malesian botany, fieldwork and other field trips, etc. It was started in 1947 by C.G.G.J. van Steenis and through its indices is a goldmine of information on SE Asian floristics and taxonomy. At present the Bullletin has been reduced to the bibliographies on mosses and vascular plants and is only available online. These contain only the bibliographies published from volume 11 (6) 1995 to 2009, but not those of 13(4) through 14(3) as yet.

Cyclopedia of plant collectors. The Cyclopedia of plant collectors (http://www. nationaalherbarium.nl/fmcollectors) contains data on collectors in the Southeast Asian Archipelago, also known as Malesia (comprising Brunei, Indonesia, Malaysia, Papua New Guinea, Phillippines, Singapore, and Timor Leste). The data were collected by Mrs. Van Steenis-Kruseman and they are digitised from FM ser. 1, part 1, 5, and 8. Pictures of collectors and samples of handwriting and signatures are often included. The data on collecting trips are more or less complete up to 1974. Later data on collectors and their trips can be obtained from the FM Bulletin. The website is especially useful when one digitises hand written specimen labels from the earlier collectors.

2. Facilitating efficient remote collaboration

Collaboration can have different forms. A measure taken in 1989 to speed up and revitalise the FM project was to establish family teams. At present, the most common way is to divide work (either taxonomically or geographically), finish each part independently and perhaps build a common (set of linked) data base(s) (see, e.g., an early example, the Euphorbiaceae: www.nationaalherbarium.nl/euphorbs) and compile an overall treatment at the end. However, to make full use of the e-possibilities, it is imperative to exploit the potential of internet communication tools that allow instant communication and sharing of data, such as online fora, Internet Relay Chat (IRC), Cloud computing services (e.g., Google Docs), and social networking tools, during the whole process of generating data and publishing information.

By making use of community e-tools for taxonomy (i.e., Scratchpads, EDIT Platform for Cybertaxonomy) the efficiency of the taxonomic work processes (data preparation and publication) could be increased. Scratchpads (http://scratchpads. eu) is a social networking application that enables communities of researchers to manage, share and publish taxonomic data online. It helps to increase visibility of ongoing projects, and creates interaction and synergy between remote working groups. The Taxonomic Editor - EDITor is part of the EDIT Platform for Cybertaxonomy (Berendsohn 2010). It is a desktop application that can be used to edit data stored in a standards-based community store (CDM-Store). It edits data in either a remote source, or a local data source embedded in the application. The EDITor allows collaborators to manage their data. Other CDM-based applications of the Platform allow the production of printed versions in flexible format or direct and up-date output to a website (CDM Data Portal). The latest version can be found at http://wp5.e-taxonomy.eu/cdm-setups/ taxonomic_editor).

3. Making use of e-databases to safeguard FM data

The defragmentation of taxonomic data and the use of common standards will increase the sustainability of FM work. The Common Data Model (CDM) is a data format for every conceivable type of data produced by taxonomists in the course of their work (http://wp5.e-taxonomy.eu). It enables professionalised taxonomic software development and allows for common standards that create sustainability. Furthermore, it can be used to exchange information with other taxonomic databases such as BRAHMS (http://dps.plants.ox.ac.uk/bol), and also non-taxonomic databases like TRY (http://www.try-db.org). The CDM-based EDIT Platform for Cybertaxonomy will facilitate the generation of species lists without generic contradictions (with the advantage that other kinds of projects could also use it), but it also allows to store alternative taxonomic classifications, e.g. entire monographs and flora treatments. This will greatly assist the consolidation of the taxonomic research results in problem areas.

We envisage that the FM e-Flora will become available in several different formats: (i) for online use on personal computers and mobile phones; (ii) as stand-alone versions on CD/DVD-ROM or as down-loadable applications for smart phones, tablets, and other mobile devices; and (iii) as printed volumes using a print-on-demand system that will always use the latest version of the e-Flora.

4. Changing the work flow in data preparation and presentation

The use of e-tools facilitates the production of new content for printed and electronic publications (instead of using printed sources to produce digitised content). However, to port FM data, which is currently only available in print or in text format, to the online tools, markup is a prerequisite, i.e., the insertion of markers designating specific content types in the text (e.g., a generic name, a distribution record, etc.). FM floristic information is at present being digitised using XML (eXtensible Markup Language; see http://www.w3.org/standards/xml). FM volumes that are already available in a digital format can be marked up straightaway, whereas earlier volumes need to be scanned first in high quality.

Mark-up is generally performed in Microsoft Word, using a combination of automated procedures and manual corrections. Automation (e.g., Word-macros) may be used to speed up the mark-up process of highly structured texts. Manual corrections are required, e.g., when taxonomists have used various types of shorthand notations to combine similar species names into one paragraph in printed floras, or when typographical or text-recognition-errors interfere with the automated mark-up. The resulting XML-files are imported into the EDIT CDM using a specific import scheme. Figures are prepared for use with the marked-up files and the CDM, but are located on a separate image server. The mark-up process and preparation of the images takes roughly a month per volume to complete (25 MS pages/day). A preview of the future FM e-Flora portal as a CDM Dataportal implementation can be found at http://160.45.63.201/dataportal/preview/flora-malesiana. A finalised version of this portal will be made available through the FM website in 2011.

5. Creating institutional commitment

Institutional commitment from Flora Malesiana institutions and support of the European e-Floras Initiative activities is highly desirable and needed to speed up activities of the Flora Malesiana project. The institutional commitment that is necessary to achieve this is not limited to providing dedicated staff, but should also include the necessary infrastructure, including connections to available high-throughput internet facilities: TEIN3 and TEIN4 (http://www.tein3.net). Staff evaluation criteria should include also contributions towards data base maintenance, conforming to the recent MoU on Evaluation Criteria for Taxonomic Work as adopted by EDIT institutions.

References

Berendsohn, W.G. (2010) Devising the EDIT Platform for Cybertaxonomy. In: Nimis, P.L. & Vignes-Lebbe, R. (eds) *Tools for Identifying Biodiversity: Progress and Problems*, pp. 1–6. Italy, Trieste: Edizioni Università di Trieste.

Dessein, S., Janssen, T., Groom, Q., Robbrecht, E., Roos, M. & Sierra, S. (2010) E-Floras for Africa: state of the art and perspectives. In: Jeannoda, V.H., Razafimandimbision, S.G. & Block, P. de (eds) *Abstracts. XIXth AETFAT*

Congress - Madagascar, 25–30 April, 2010, p. 133. Scripta Botanica Belgica 46, Meise: National Botanic Garden of Belgium.

Roos, M. & Hovenkamp, P. (2009) Flora Malesiana in the coming decade. *Blumea* 54: 3–5.

http://wp5.e-taxonomy.eu/cdm-setups/taxonomic_editor

http://dev.e-taxonomy.eu/trac/wiki/TaxonomicEditor

http://dps.plants.ox.ac.uk/bol

http://scratchpads.eu

http://wp5.e-taxonomy.eu

http://www.br.fgov.be/RESEARCH/DATABASES/FOCA/index.php

http://www.e-taxonomy.eu/node/748

http://www.e-taxonomy.eu/node/859

http://www.floramalesiana.org

http://www.kew.org/herbarium/keys/fm

http://www.nationaalherbarium.nl/fmbull/biblio.htm

http://www.nationaalherbarium.nl/fmcollectors

http://www.tein3.net

http://www.try-db.org

http://www.w3.org/standards/xml

http://160.45.63.201/dataportal/preview/flora-malesiana