Rodriguésia 66(4): 983-987. 2015
http://rodriguesia.jbrj.gov.br
DOI: 10.1590/2175-7860201566404

Hard Copy to Digital: Flora Neotropica and the World Flora Online

William Wayt Thomas[1,2] *& Melissa Tulig*[1]

Abstract

One of the greatest challenges in achieving the goals of the World Flora Online (WFO) will be to make available the huge amount of botanical information that is not yet available digitally. The New York Botanical Garden is using the Flora Neotropica monograph series as a model for digitization. We describe our efforts at digitizing Flora Neotropica monographs and why digitization of hardcopy descriptions must be a priority for the WFO project.

Key words: Electronic monographs, open access, Flora Neotropica, monographs.

Resumo

Um dos maiores desafios para alcançar as metas do projeto World Flora Online (WFO), será a disponibilizar a enorme quantidade de informações botânicas que ainda não estão disponíveis digitalmente. O New York Botanical Garden está utilizando a série de monografias da Flora Neotropica como um modelo para a digitalização. Nós aqui descrevemos nossos esforços na digitalização das monografias da Flora Neotropica e porque a digitalização das descrições impressas deve ser uma prioridade para o projeto WFO.

Palavras-chave: Monografias eletrônicas, open access, Flora Neotropica, monografias.

Introduction

The World Flora Online (WFO) was developed as part of the United Nation's Global Strategy for Plant Conservation with the goal of providing "an online flora of all known plants," One of the greatest challenges in achieving the goals of the WFO will be to make available the huge amount of botanical information that is not yet available digitally. Here, we describe our efforts at digitizing Flora Neotropica monographs, discuss why that series is our first priority, and why digitization of hardcopy descriptions must be a priority for the WFO project.

World Flora Online

Objective I of the United Nation's Global Strategy for Plant Conservation, states that by 2020 plant diversity should be expertly characterized and documented in a standardized format. In response, a consortium of global research institutions, endorsed by the international botanical community, will create a definitive, verified online resource documenting all known plants of the world. It is called the World Flora Online (WFO). This consortium of professionals will create open-access one-stop searching of world flora with verified information, including new and previously published data, and coordinated with links to other plant database and catalog Web sites.

The core information in the World Flora Online will be vetted by professional botanists and will include (*highest priority):

- A comprehensive list of names for accepted species*, with synonyms placed with each correct name.
- Published modern descriptions, as well as the protologue* (the original published description when the species was first described in scientific literature).
- Literature citations for each name*, with reference to the type specimen for each name.
- Images of each species, including digital images of herbarium specimens (including connections to the type images in JSTOR Global Plants) as well as photographs of living plants.
- Links to other portals with plant information (example: the herbarium specimen data for

[1] The New York Botanical Garden, Bronx, NY 10458-5126, USA.
[2] Author for correspondence: wthomas@nybg.org

each species through the Global Biodiversity Information Facility (GBIF), among others).

- Geographic distribution of each species.
- Conservation status of each species.

Plant Diversity of the Neotropics

The Neotropics have functioned as an engine for generating plant diversity (Antonelli *et al.* 2015) but why do they deserve attention as a botanical priority, and as a test case for digitization? The Neotropics are bounded by the Tropics of Cancer and Capricorn. Although these boundaries are artificial, they approximate biological boundaries limiting the distribution of tropical species outside the tropics and vice versa. In Mexico, the Tropic of Cancer follows the break between the arid and semi-arid climates to the north and the humid and semi-humid climates to the south (Toledo *et al*. 1997). The Caribbean buffers the West Indies and maintains a tropical climate north into southernmost Florida. Farther north, however, the vegetation must withstand rare freezing; consequently, many species simply are not found north of the Everglades. To the south, the Tropic of Capricorn runs through the Atacama Desert of northern Chile, a significant barrier to plant life. To the east, the Tropic is found at the latitude of the city of São Paulo, just south of the southern limit of the cerrado. The Atlantic coastal forest continues south to the end of the Serra do Mar at about 30° south (Thomas 2005).

The flora of tropical America includes many plant groups that are exclusively or primarily found in the New World, such as the Bromeliaceae (Smith & Downs 1974, 1977, 1979) and Gunneraceae (Mora-Osejo *et al*. 2011). In addition, there are many families for which genera or groups of genera are mostly American, such as the Lecythidaceae (Prance & Mori 1979; Mori & Prance 1990), *Mimosa*-Fabaceae (Barneby 1991), and Miconieae-Melastomataceae (Michelangeli *et al.* 2009 onward). Although there are exceptions (Pennington *et al.* 2004), in many cases (see Smith *et al.* 2004), therefore, a revision of a Neotropical group will functionally be a monograph of that group - this is even more so if the small number of non-tropical outliers are also treated.

So, how many species are there in the Neotropics, and what are our prospects of making their descriptions available through the World Flora Online? Estimates of species diversity are fraught with uncertainty, but do provide a useful yardstick and, for the Neotropics, new information means a new estimate is needed. Earlier estimates (Prance 1977; Raven 1988; Thorne 2002) of the number of Angiosperm species on earth were approximately 250,000, with about 36 percent (90,000) found in tropical America (Forero & Mori 1995; Maas & Westra 1998; Raven 1988; Thomas 1999). Govaerts (2001, 2003) estimated the World's species diversity of seed plants at a much higher 420,000 species. Since The Plant List (2013) is based on an actual count of species, it offers a more detailed estimate of the World's plant species. It includes 951,140 species names of Angiosperms. Of these 304,419 are accepted species names and another 216,375 unresolved. If 20–50 percent of the unresolved names are found to be accepted, the estimate of the number plant species in the World rises to 350,000–400,000. With the estimate for the World's species having increased by almost 100,000, it is likely that the estimate for the Neotropics should be higher also - if the Neotropical flora comprises a third of the World's flora, it would total at least 120,000 species.

Flora Neotropica monographs - A Model for Digitization

Like many botanical journals, Flora Neotropica monographs were published in print form over many years. The idea of a Flora Neotropica project was first envisioned in 1957 by an advisory committee of UNESCO and, in 1964, the Organization for Flora Neotropica was created at a meeting held at the Institute of Botany of São Paulo at the invitation of the Institute's director, Dr. Alcides R. Teixeira. It also led to the creation of a commission to stimulate and assist in the preparation and publication of all the naturally occurring fungi and plants of tropical America in a monograph series entitled "Flora Neotropica." *Flora Neotropica* was based at The New York Botanical Garden and, since its inception, all costs of the monograph series have been assumed by the Garden. The series is reserved for manuscripts that, because of their length, cannot be published in their entirety in other journals (Forero & Mori 1995; Thomas 2005).

Since the first Flora Neotropica monograph appeared in 1965, an additional 113 volumes have been published. Of these, 87 were on vascular plants (16 were on fungi, and 11 on bryophytes). These 87 monographs treat 16,730 scientific names

and recognize 8008 species, of which 2087 (26%) were described by the authors of the monographs. These numbers demonstrate that monographers are those best able to recognize and circumscribe the species in the groups they study - they also perceive rare and cryptic species passed over by non-specialists and those which can be critical for conservation planning (Marhold *et al.* 2013).

In order to increase contributions of monographs, especially from botanists in the Neotropics, monograph submissions are accepted in Portuguese, Spanish and French. Flora Neotropica monographs are published through The New York Botanical Garden Press and have always been available for sale. So that costs of printed monographs is minimized, in recent years the New York Botanical Garden Press has reduced original print runs and is making the monographs available through Print-on-Demand publication. Currently, Flora Neotropica is available only as hardcopy or with subscription access to JSTOR (<http://www.jstor.org>), making them generally unavailable to those not affiliated with a university or those from developing countries. Even for those with access, the PDFs available are only of images and do not contain any embedded text, making them unsearchable.

Hard Copy to Digital: Flora Neotropica monographs

The goals of WFO include providing published modern descriptions of each species, online. Most of the descriptions of the World's plant species have been published in print form. What proportion of the descriptions of the World's species already exist digitally as searchable text, and what proportion only exist as hardcopy, or as digital page images? The WFO Consortium has estimated that no more than 160,000 species have digital descriptions, meaning that there are an estimated 190,000 species for which descriptions must be transformed from hardcopy to digital searchable descriptions - a large task.

With efforts like the Biodiversity Heritage Library (<http://www.biodiversitylibrary.org/>), access to much of the earlier literature, especially that out of copyright, has been made available to the public online. Many of these publications contain protologues (original descriptions), that are often the only available description for a given species. BHL will be an excellent resource for the WFO. Access and re-use of recent literature with treatments and synonymy based on current classifications using modern techniques (DNA analyses, etc.), however, is generally restricted by copyright. Organizations such as Plazi argue that taxonomic descriptions do not fit the definition of copyright and can be marked up and re-used (Agosti and Egloff 2009), but there is still some hesitation with how copyright laws in different countries will be applied to this type of re-purposing of publications.

Increasingly, semantically-enhanced, born-digital publications are competing with the print-only (or print plus paywall) approach. Open access publishing models promoted by publishers like Pensoft will revolutionize publishing (Miller *et al.* 2012), but there remains a large amount of literature still in copyright and mostly available online as PDFs behind subscription services or individually available by download from the publisher for a set price per article.

In light of the WFO and because of its importance botanically, we are using Flora Neotropica monographs as a test case for converting a print plus paywall publication to marked up text and providing free access to these descriptions online. Even for those with access to digital copies of Flora Neotropica, the PDFs available are typically only page images and contain only some automatically generated, but not corrected or marked up text, making them unreliably searchable both individually and across all volumes. Each volume of Flora Neotropica was published with different licensing agreements with scientists and artists: database of generic- and species-level descriptions has therefore been generated from images of each volume.

To convert each volume of Flora Neotropica to text, PDFs of the images are sent through ABBYY FineReader Professional OCR (optical character recognition) software. This software also highlights potentially misspelled words or incorrect character conversion to make it easier to quickly proofread. A botanical dictionary of plant morphology terms and taxonomic names was loaded into the software to minimize the number of potential misspellings. Since all volumes remained fairly standard in font and formatting, issues with conversion to text have been consistent across volumes and the software has been trained to recognize reccurring problems. Project staff manually edit any OCR transcription errors or formatting issues discovered during the process. Once the text is cleaned up, the information is

parsed and mapped to the DarwinCore standard, adopted by the Biodiversity Information Standards (TDWG) association (<http://rs.tdwg.org/dwc/>).

Once complete, DarwinCore-Archive (DwC-A) files will be produced and shared via the NYBG Integrated Publishing Toolkit (IPT) for harvesting by the World Flora Online Consortium portal, Global Biodiversity Information Facility (GBIF), the Encyclopedia of Life (EOL), the Digital Public Library of America (DPLA) or any interested party. They will also be available on the NYBG website, with links to cited NY specimen records in the Virtual Herbarium.

Conclusion

To reach the stated goals of the WFO by 2020, digitizing hardcopy plant descriptions is critically important and must be embraced by all members of the Consortium. We have started with Flora Neotropica because of the biological importance of the series and because we can clarify the issues of copyright and licensing agreements.

Acknowledgments

We thank the Alfred P. Sloan Foundation and Google, Inc., for their valuable support of this project.

References

Agnarsson, I. & Kuntner, M. 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. Systematic Biology 56: 531-539. DOI: 10.1080/10635150701424546.

Agosti, D. & Egloff, W. 2009. Taxonomic information exchange and copyright: the plazi approach. BMC Research Notes 2: 53. DOI: 10.1186/1756-0500-2-53.

Antonelli, A.; Zizka, A.; Silvestro, D.; Scharn, R.; Cascales-Miñana, B. & Bacon, C.D. 2015. An engine for global plant diversity: highest evolutionary turnover and emigration in the American tropics. Frontiers in Genetics 6: 130. DOI: 10.3389/fgene.2015.00130.

Barneby, R. 1991. Sensitivae Censitae - A description of the genus *Mimosa* Linnaeus (Mimosaceae) in the new world. Memoirs of the New York Botanical Garden 65: 1-835.

Forero, E. & Mori, S.A. 1995. The organization for Flora Neotropica. Brittonia 47: 379-393.

Maas, P.J.N. & Westra, L.Y.T. 1998. Familias de plantas neotropicales. Koeltz Scientific Books. Koenigstein, Germany. 315p.

Marhold, K.; Stuessy, T.; Agababian, M.; Agosti, D.; Alford, M.H.; Crespo, A.M.; Crisci, J.V.; Dorr, L.J.; Ferencová, Z.; Frodin, D.; Geltman, D.V.; Kilian, N.; Linder, H.P.; Lohmann, L.G.; Oberprieler, C.; Penev, L.; Smith, G.; Thomas, W.; Tulig, M.; Turland, N. & Zhang, X.-C. 2013. The future of botanical monography: report from an international workshop, 12-16 March 2012, Smolenice, Slovak Republic. Taxon 62: 4-20.

Michelangeli, F.A.; Almeda, F.; Goldenberg, R.; Judd, W.S.; Becquer-Granados, E.R. & Tulig, M. 2009 onward. PBI Miconieae: a complete web-based monograph of the tribe Miconieae (Melastomataceae). The New York Botanical Garden, Bronx, NY. Available at <http://sweetgum.nybg.org/melastomataceae>. Access on 3 September 2015.

Miller, J.; Dikow, T.; Agosti, D.; Sautter, G.; Catapano T.; Penev, L.; Zhang, Z.-Q.; Pentcheff, D.; Pyle, R.; Blum, S.; Parr, C.; Freeland, C.; Garnett, T.; Ford, L.S.; Muller, B.; Smith, L.; Strader, G.; Georgiev, T. & Bénichou, L. 2012. From Taxonomic Literature to Cybertaxonomic Content. BMC Biology 10: 87. DOI: 10.1186/1741-7007-10-87.

Mora-Osejo, L.E.; Pabón-Mora, N. & González, F. 2011. Gunneraceae. Flora Neotropica 109: 1-166.

Mori, S.A. & Prance, G.T. 1990. Lecythidaceae - part II, The Zygomorphic-flowered new world genera (*Couroupita*, *Corythophora*, *Bertholletia*, *Couratari*, *Eschweilera*, & *Lecythis*). Flora Neotropica 21: 1-376.

Pennington, R.T. & Dick, C.W. 2004. The role of immigrants in the assembly of the South American rainforest tree flora. Philosophical Transactions of the Royal Society B: Biological Sciences 359: 1611-1622.

Pimm, S.L.; Clinton, N.J.; Joppa, L.N.; Roberts, D.L. & Russell, G.J. 2010. How many endangered species remain to be discovered in Brazil? Natureza e Conservação 8: 71-77. DOI: 10.4322/natcon.00801011.

Prance, G.T. 1977. Floristic inventory of the tropics: where do we stand? Annals of the Missouri Botanical Garden 64: 659-684.

Prance, G.T. & Mori, S.A. 1979. Lecythidaceae - part I, the actinomorphic-flowered new world Lecythidaceae (*Asteranthos*, *Gustavia*, *Grias*, *Allantoma*, & *Cariniana*). Flora Neotropica 21: 1-270.

Raven, P. H. 1988. Tropical floristics tomorrow. -Taxon 37: 549- 560.

The Plant List 2013. Version 1.1. Published on the Internet. Available at <http://www.theplantlist.org/>. Access on 1 September 2015.

Smith, L.B. & Downs, R.J. 1974. Pitcairnioideae (Bromeliaceae). Flora Neotropica 14: 1-658.

Smith, L.B. & Downs, R.J. 1977. Tillandsioideae (Bromeliaceae). Flora Neotropica 14: 663-1492.

Smith, L.B. & Downs, R.J. 1979. Bromelioideae (Bromeliaceae). Flora Neotropica 14: 1493-2142.

Smith, N.; Mori, S.A.; Henderson, A.; Stevenson, D.W. & Heald, S. 2004. Flowering plants of the Neotropics. Princeton University Press, Princeton. 594p.

Thomas, W.W. 1999. Conservation and monographic research on the flora of Tropical America. Biodiversity and Conservation 8: 1007-1015.

Thomas, W.W. 2005. Flora Neotropica - Monographs as inventories. *In*: Species Plantarum 250 years: Proceedings of the Species plantarum symposium held in Uppsala August 22-24, 2003. Acta Universitatis Upsaliensis, Symbolae Botanicae Upsalienses 33: 187-192.

Thorne, R.F. 2002. How many species of seed plants are there? Taxon 51: 511-522.

Artigo recebido em 08/09/2015. Aceito para publicação em 16/10/2015.