

Toward an Open Access repository and data services for Plant Biodiversity research

Sierra, S. E. C., Roos, M. C., Hamann, T. (Stichting Nederlands Centrum voor Biodiversiteit Naturalis),
Berendsohn, W. G., Guentsch, A., Kirchoff, A. (Botanic Garden and Botanical Museum Berlin-Dahlem, Freie Universität Berlin),
Kirkup, D. (Royal Botanic Gardens Kew)
and Janssen, T. (National Botanic Garden of Belgium)

EDIT Benefits

EDIT has provided modern tools for taxonomists to access, share and publish taxonomic information, and in this way, has strengthened and expanded the network of the plant taxonomic community.

Concept

A wealth of knowledge on plant biodiversity (i.e. descriptions of morphological and molecular features, diagnostic characters, distribution, ecology, morphological correlates of functional traits, traditional and economic uses, species conservation status) has been assembled over hundreds of years in major European botanical institutions. The plant biodiversity knowledge is available in the form of Floras (i.e. condensed expert accounts for given regions). The accounts in Floras concern the highest quality refereed and validated data on identity and relations, descriptions and reference specimens, traits and characters, identification keys, plant uses, ecological properties and distributions over time. Generating complete Flora information is labour intensive and time consuming to collate; hence represents considerable long-term financial investment. Flora information is mainly available in the form of printed books and scattered journal publications, rarely as marked up text files, or increasingly, as locally created and curated databases.

Modern society demands up-to-date, reliable and accurate plant biodiversity information, but this to date is highly fragmented, and almost exclusively accessed by taxonomic specialists. Data are dispersed by organisation, format (a multitude of local databases or printed formats), and research community (taxonomy, morphology, functional or phenotypic traits, ecology, conservation issues, etc.). Information is frequently not shared because of protective copyright

claims. Where data are shareable in principle, doing so requires not only overcoming technical obstacles, but also harmonising different data quality standards and applying compatible ontologies. Hence, the data are not readily available, and re-use across various subdisciplines, or for delivering new custom information products is not directly possible.

As a result of this, EDIT Work Package 3 (Integrating and Reshaping the Infrastructure Basis) started a coordinated effort among twenty European institutions to optimise and share curated plant information and skills: OpenPlantBio (Open Access repository and data services for Plant Biodiversity research).

OpenPlantBio aims at a coordinated and integrated structure to assist a broad scope of users to obtain scientifically sound information on plant diversity, by federating plant biodiversity data repositories. This distributed repository will comprise the information of the major (international) Flora series produced and/or coordinated by European institutes. The objectives of OpenPlantBio are, to: 1) enhance accessibility of the taxonomic information; 2) make this content usable for a wider scientific audience, and 3) ensure sustainability of the infrastructure.

This coordinated effort is an unprecedented initiative for the botanical sphere. It builds on existing infrastructure, technology, and earlier efforts (e. g. several e-Flora initiatives, EDIT Work Package 5: EDIT Platform for Cybertaxonomy: Tools, Sharing, Networking and Integration; see Figure 1). For additional information, see related EDIT Success Story "Using the EDIT Platform for Cybertaxonomy to elaborate and disseminate complex floristic information". It has new emphasis on consolidation, interoperability, data quality curation, preservation, data mining tools (retrieval, visualisation) and service deployment for scientists, policy makers, and citizen scientists. Most consortium members are key reference institutes for plant taxonomy and will play a vital role in providing the content that will be integrated in the repository.

OpenPlantBio is annex to existing e-infrastructure projects, makes use of their infrastructure and services, and adds unique value by focusing on federating and unifying the e-Flora data and outputting innovative information and identification systems that underpin processes in science and society.

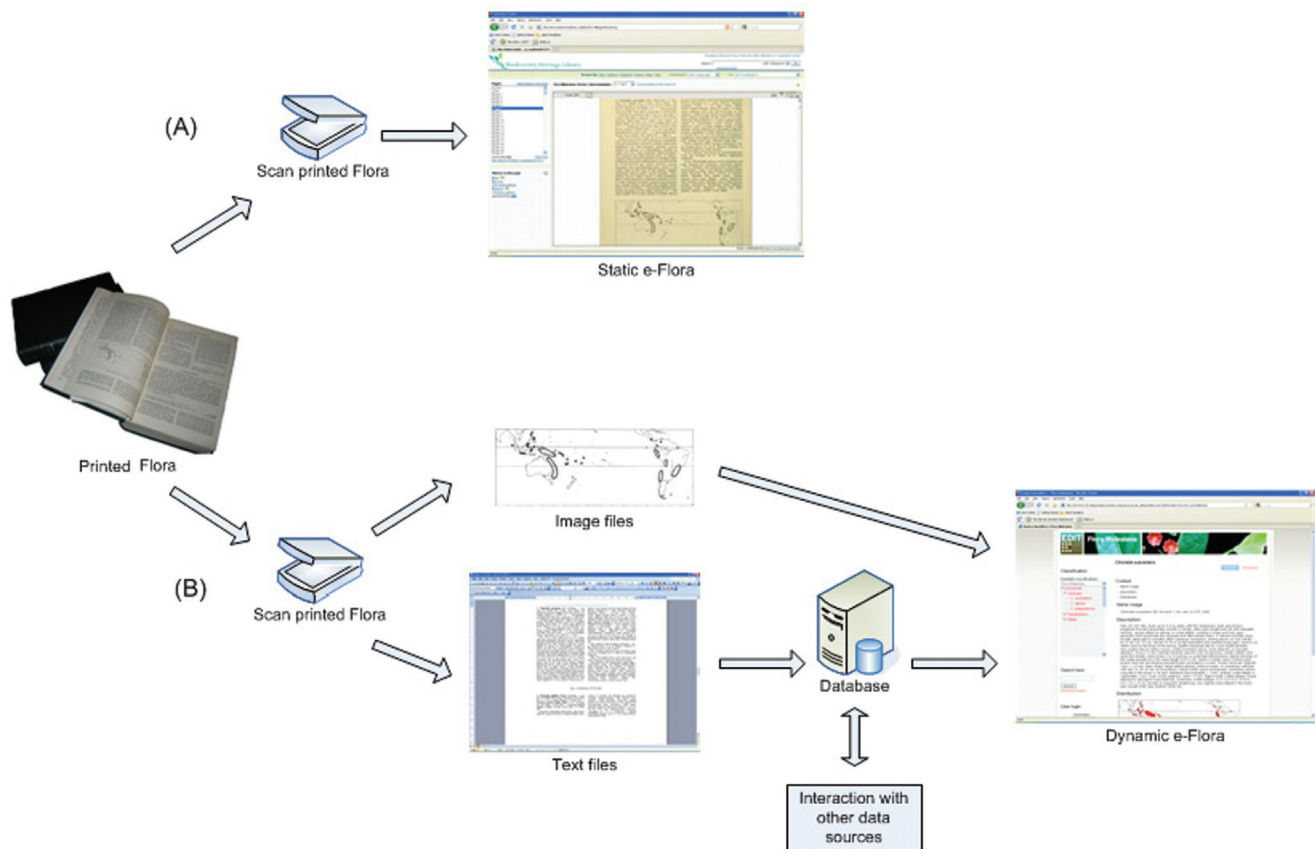


Fig. 1. From printed Floras to e-Floras: (A) from printed Flora to static e-Flora, (B) from printed Flora to locally data based dynamic e-Flora.

Outcomes

The OpenPlantBio infrastructure will help the scientific community, as well as national, regional and global organisations to have access to accurate biodiversity information, by:

- Constructing a high quality distributed repository containing the most complete synthesis of plant diversity information covering Europe, the greater part of the paleotropics, and some neotropical areas. In total it would cover ca. one third of the plant species diversity worldwide.
- Enhancing and sustaining service provisioning by the standardisation of the input and access processes and a single overarching technical structure. It will employ and refine rigorous data quality standards that are internationally approved, interoperable, functional, and that have already been used with success. To meet the need for information of flexible and transparent quality levels, it will transport tailor made plant biodiversity information to the public and to the entire environmental sciences community (see Figure 2).
- Providing a rich set of modern web-services and (search) applications in response to needs and demands. Content from a broad geographical coverage will be available and access to these data resources will be facilitated, thus, promoting data mining and exploitation by a wide user base including the environmental/natural sciences research community, the public and citizen scientists. Expert tools, particularly for identification, visualisation, special purpose queries and further use of the data, will maximize user-friendliness.
- Serving the needs of the scientific community and creating a new research environment for a virtual environmental research community (e-Science): based on the collective power of integrated and expert-curated plant biodiversity data in key European botanical research facilities.
- Promoting data sharing and collaboration between botanical, ecological and environmental research institutions and build trust (in collaborators as well as in data quality).

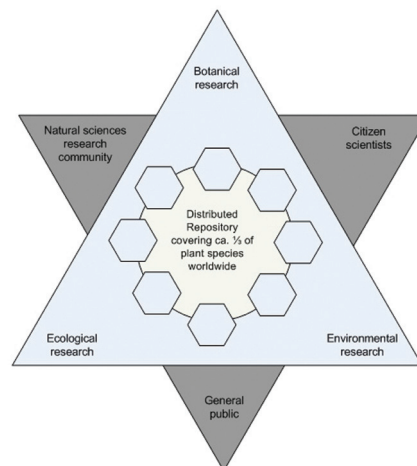


Fig. 2. Schematic overview of OpenPlantBio set-up.

- Leveraging national e-Science activities (e.g. by the data services provided) and allowing open and participatory generation and presentation of plant biodiversity knowledge.
- Providing access from raw data (specimens, observations, analyses) to integrated and curated data in a data warehouse environment (Floras, trait databases, environmental analyses). Data from various sources will be accessible to new user communities via one portal and unified access points.

- Providing a common source for developing custom plant information and identification system; products that can be used online and offline and in home/office environments as well as in the field using mobile devices and modern communication technology.

For additional information on OpenPlantBio, please contact:
roos@nhn.leidenuniv.nl; sierra@nhn.leidenuniv.nl

FACT BOX

Examples of static Flora information

Flora Malesiana: <http://www.biodiversitylibrary.org/bibliography/40744>
 Flora Zambesiaca: <http://apps.kew.org/efloras/>
 Flora of China: <http://hua.huh.harvard.edu/china/>

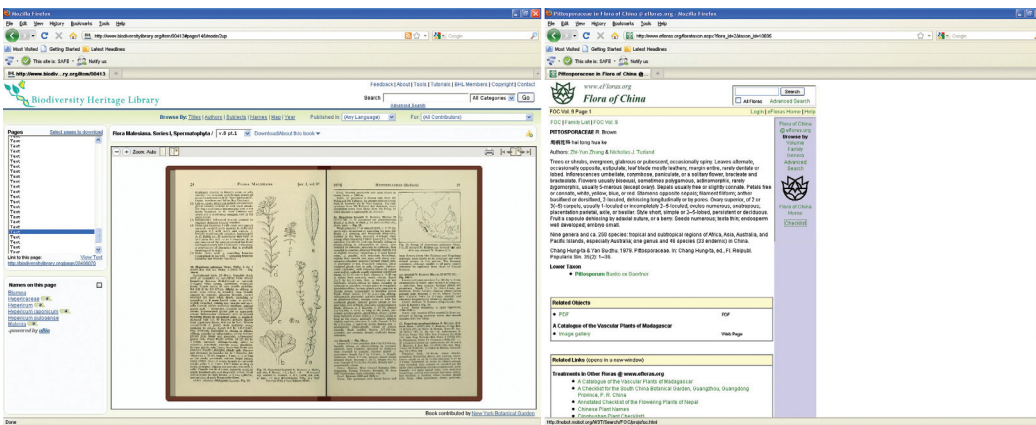


Fig. 3. Static e-Floras: Flora Malesiana at the Biodiversity Heritage Library (left) and Flora of China (right).

Examples of dynamic Flora information

Araceae taxonomy on the web: <http://www.cate-araceae.org>
 Palms of the World On-Line: <http://www.palmweb.org>
 Flore d'Afrique Centrale: <http://www.br.fgov.be/RESEARCH/DATABASES/FOCA/index.php>

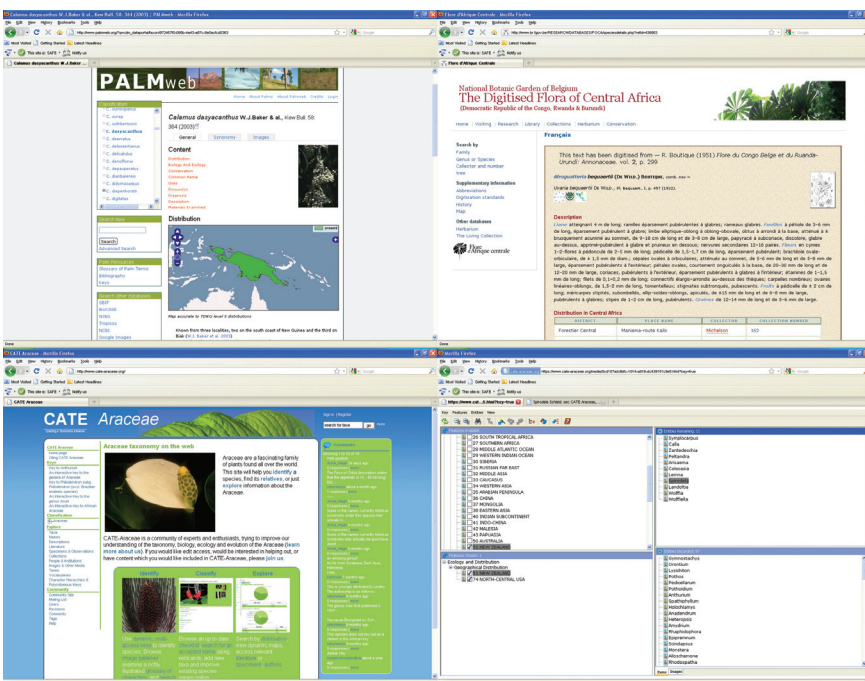


Fig. 4. Dynamic e-Floras: Palmweb (Palms of the World On-Line) showing interactive distribution map (top left), Flore d'Afrique Centrale (top right), CATE Araceae (Araceae taxonomy on the web) showing various features (bottom left) and interactive keys (bottom right).

Data providers of OpenPlantBio belong to the major natural history museums and/or centers of biodiversity research in Europe. They house some of the largest and most diverse botanical collections in the world (MNHN 11 million, NHMW 8 million, RBGK 7 million, CJBG 6 million, NCB Naturalis 5.5 million, NBGB 3.5 million, BGBM 3.5 million).