



HAL
open science

Pl@ntNet, une plate-forme innovante d'agrégation et partage d'observations botaniques

Daniel Barthélémy, Nozha Boujemaa, Jean-François Molino, Alexis Joly,
Hervé Goëau, Vera Bakić, Souheil Selmi, Julien Champ, Jennifer Carre,
Mathias Chouet, et al.

► To cite this version:

Daniel Barthélémy, Nozha Boujemaa, Jean-François Molino, Alexis Joly, Hervé Goëau, et al..
Pl@ntNet, une plate-forme innovante d'agrégation et partage d'observations botaniques. International
Conference 'Botanists of the Twenty-first Century', UNESCO, Sep 2014, Paris, France. pp.191-197.
hal-01837361

HAL Id: hal-01837361

<https://hal.science/hal-01837361v1>

Submitted on 5 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Organisation
des Nations Unies
pour l'éducation,
la science et la culture



Quels botanistes pour le 21^e siècle ?

Métiers, enjeux et opportunités



Basé sur les actes de la Conférence
internationale de l' UNESCO
Quels botanistes pour le 21^e siècle ?
Métiers, enjeux et opportunités

Septembre 2014, Paris, France



United Nations
Educational, Scientific and
Cultural Organization



Botanists

of the twenty-first century

Roles, challenges and opportunities

Based on the proceedings of the UNESCO International conference
“Botanists of the twenty-first century: roles, challenges and opportunities”
held in September 2014 in Paris, France

Edited by Noëline R. Rakotoarisoa, Stephen Blackmore and Bernard Riera

Published in 2016 by the United Nations Educational, Scientific and Cultural Organisation
7, place de Fontenoy, 75352 Paris 07 SP, France

© UNESCO, 2016

All rights reserved.

ISBN 978-92-3-100120-8



This publication is available in Open Access under the Attribution-ShareAlike 3.0 IGO (CC-BY-SA 3.0 IGO) license (<http://creativecommons.org/licenses/by-sa/3.0/igo/>). By using the content of this publication, the users accept to be bound by the terms of use of the UNESCO Open Access Repository (<http://www.unesco.org/open-access/terms-use-ccbysa-en>).

Original title: Botanists of the twenty-first century: roles, challenges and opportunities – Based on the proceedings of the UNESCO International conference “Botanists of the twenty-first century: roles, challenges and opportunities” held in September 2014 in Paris, France

The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO concerning the legal status of any country, territory, city or area or of its authorities or concerning the delimitation of its frontiers or boundaries.

The ideas and opinions expressed in this publication are those of the authors and are not necessarily those of UNESCO and do not commit the Organisation.

Editors: Noëline R. Rakotoarisoa, Stephen Blackmore and Bernard Riera
Cover photo: Shutterstock/Madlen

Composed and printed in the workshops of UNESCO
The printer is certified Imprim'Vert®, the French printing industry's environmental initiative.

Printed in France



Warning: The articles in this book are either in French or English depending on the language used by the author. Each article is summarized in the other language.

Attention: Les articles présentés dans cet ouvrage sont en français ou en anglais suivant la langue utilisée par l'auteur. Chaque article est résumé dans l'autre langue.

With the support of



Convention on
Biological Diversity



The Global Partnership
for Plant Conservation





Présentation
de Daniel
Barthélémy,
Centre de
Coopération
Internationale
en recherche
Agronomique
pour le
Développement
(CIRAD),
Montpellier,
France

Pl@ntNet, une plate-forme innovante d'agrégation et partage d'observations botaniques

Auteurs

Daniel Barthélémy, Nozha Boujemaa, Jean-François Molino, Alexis Joly, Hervé Goëau, Vera Bakić, Souheil Selmi, Julien Champ, Jennifer Carré, Mathias Chouet, Aurélien Peronnet, Christel Vignau, Samuel Dufour-Kowalski, Antoine Affouard, Julien Barbe, Pierre Bonnet

Résumé

Pl@ntNet est un réseau humain s'appuyant sur une infrastructure informatique, permettant l'identification, l'agrégation et le partage d'observations botaniques à très grande échelle. Cette initiative mobilise différentes institutions de recherche dans divers champs scientifiques (informatique, agronomie, écologie) et de larges réseaux associatifs de naturalistes ; elle a permis au cours des 20 derniers mois la collecte de plusieurs dizaines de milliers d'observations de plantes sur le territoire européen. Celles-ci sont collectées à travers un système web ou mobile d'aide à l'identification des plantes par l'image, au sein duquel chacun des participants peut partager des observations déterminées ou non, avec différents niveaux d'informations. Ces observations, sont ensuite révisées collaborativement à travers 2 applications web, l'une dédiée à la révision de la détermination des observations (IdentiPlante), l'autre à l'estimation de la qualité visuelle des images illustrant l'observation (PictoFlora). Les observations révisées et / ou validées, viennent alors enrichir la base d'observations publiques exploitée par le système d'aide à l'identification par l'image. Bien que cette infrastructure logicielle soit encore récente (lancement des applications mobiles en février 2013), elle a été exploitée au cours des 20 derniers mois par près de 300 000 utilisateurs à travers le monde. Les modalités de fonctionnement de celle-ci, ainsi que les perspectives d'évolution sont présentées dans cet article, qui se conclut par une discussion sur les changements actuels que les nouvelles technologies de l'information permettent d'opérer en Botanique.



Back to
Contents
*Retour au
Sommaire*

Abstract

Pl@ntNet is a human network based on IT infrastructure, enabling identification, aggregation and sharing of botanical observations at a large scale. This initiative mobilizes various research institutions in several scientific fields (computer science, agronomy, ecology) and wide associative networks of naturalists. It has over the past 20 months collected tens of thousands of observations of plants on the European territory. These are collected through a web or mobile application systems based on an image plant identification approach, in which each participant can share observations, with different levels of information. These observations are then collaboratively evaluated through two web applications, one dedicated to the review of the taxonomic determination (IdentiPlante), the other one to estimate the visual quality of images (PictoFlora). Revised and/or validated observations enrich the public database used by the image-based identification system. Even if this software infrastructure is new (launch of the mobile application in February 2013), it has been exploited in the past 20 months by nearly 300,000 people worldwide. Operating procedures and outlooks are presented in this article, which concluded with a discussion of the current changes that new information technologies allow to operate in Botany.



Introduction

Bien que le besoin d'expertise en botanique n'ait jamais été aussi important pour faire face aux grands défis de notre société (changement climatique, crises alimentaires récurrentes, érosion de la biodiversité sans précédent), l'accès à cette expertise est de plus en plus difficile compte-tenu de la réduction du nombre de botanistes professionnels au cours de ces dernières décennies (Bidwell, 2013 ; Whitfield, 2012). L'un des verrous majeurs pour s'investir en botanique est l'identification des plantes, qui fait appel à une large expérience et à la connaissance fine du vocabulaire usuel dans cette discipline. Cette étape d'identification, constitue bien souvent une barrière importante que beaucoup de novices ne réussissent pas à franchir rapidement, et qui laisse alors place à un découragement important. L'usage de clés dichotomiques ou multi-entrées, textuelles ou illustrées, se heurte souvent à la nécessité de bien maîtriser la terminologie descriptive du groupe auquel appartient l'espèce recherchée (cf. Fougères, Gymnosperme, *Orchiadaceae*, *Cyperaceae*, etc.). Certains termes ne se trouvent en effet utilisés que pour des groupes taxonomiques précis. Un novice qui abordera une nouvelle et vaste flore se retrouvera alors bien souvent démuné devant le volume et la diversité des mots et expressions employés pour différencier des taxons entre eux. Les guides papiers, souvent très largement illustrés, et structurés par couleurs et/ou formes des principaux organes (fleurs, feuilles, fruits) ont, quant à eux, connu une attractivité importante ces dernières décennies notamment grâce au fait qu'ils s'affranchissent en grande partie d'un vocabulaire trop avancé. Ils se heurtent cependant à des difficultés pour passer à l'échelle, et permettre d'identifier plusieurs milliers de taxons avec fluidité. Leur transfert sur support numérique offre de nouvelles perspectives (Farnsworth et al., 2013). Dans ce contexte, l'initiative Pl@ntNet s'est ainsi attelée au développement d'outils numériques innovants (i) d'aide à l'identification taxonomique par l'image, (ii) de révision collaborative de la qualité des données, (iii) de gestion de gros volumes d'observations botaniques. Le fonctionnement, l'originalité et l'apport de chacun d'eux sont présentés dans les sections suivantes, avant d'en discuter les perspectives d'évolution, favorisant l'engagement d'un plus grand nombre d'acteurs dans cette démarche.

L'un des verrous majeurs pour s'investir en botanique est l'identification des plantes, qui fait appel à une large expérience et à la connaissance fine du vocabulaire usuel dans cette discipline.

Matériels et méthodes

Pl@ntNet s'appuie sur des expertises, méthodes et jeux de données complémentaires, qui ont été développés depuis 2009 pour un même objectif : faciliter l'aide à l'identification des plantes sur le territoire de France métropolitaine, ainsi que dans différentes régions tropicales, pour le plus grand nombre de personnes possibles. Dès le début du projet Pl@ntNet (<http://www.plantnet-project.org/>), les scientifiques et volontaires mobilisés pour parvenir à cet objectif ont ainsi produit des jeux de données de plus en plus riches et complexes pour évaluer et faire progresser les méthodes actuelles de reconnaissances visuelles, afin de les appliquer à l'identification des plantes. Une image est en effet vectrice d'un très grand nombre d'informations, puisqu'il est possible à travers une photo de fleur par exemple d'en dénombrer les pétales ou sépales, d'en identifier la / les couleurs, les formes et dispositions des structures qui la composent. Un lot de quelques photos d'une plante qui, réalisées à différentes échelles et sous différents angles de vue, peut aisément en permettre l'identification par un botaniste (Leggett & Kirchoff, 2011). Il est ainsi de plus en plus fréquent de voir sur les réseaux sociaux (cf. Twitter ou FaceBook) des demandes d'aides à l'identification s'appuyant sur un petit jeu d'images et quelques informations complémentaires.

Fort de ce constat, nous avons constitué des jeux de données de différentes natures pour évaluer les techniques les plus avancées de reconnaissance visuelle. Les feuilles étant les motifs visuels les plus largement traités dans ce domaine nous avons constitué un premier jeu de données, intitulé Pl@ntLeaf, constitué de scans, photos de feuilles sur fond uniforme et photo de feuilles directement insérées sur la plante (Goëau et al., 2011a). Ce jeu de données bien que de taille modeste, a été produit de manière collaborative avec l'aide de 17 contributeurs, permettant d'illustrer 70 espèces ligneuses de la flore française. Son exploitation a permis de constater que les méthodes d'analyses visuelles, basées sur des approches généralistes par point d'intérêts permettaient d'approcher les performances des méthodes ad hoc dédiées à ce seul motif visuel (Goëau et al., 2011b).

Ce travail s'est poursuivi avec l'animation d'un projet collaboratif plus vaste intitulé « capitalisation d'images » au sein de l'infrastructure du réseau Tela Botanica (<http://www.tela-botanica.org/>). Celui-ci a abouti 2 ans plus tard à la production d'un jeu de données, constitué d'observations végétales, multi-images (plusieurs images du même motif visuel d'une plante) et multi-vues (plusieurs motifs visuels distincts d'une même plante). Les différentes vues mobilisées étaient alors celles de « ports », « fleurs »,

Les projets de sciences participatives menés à travers le monde et plus particulièrement en Europe ont prouvé la très grande capacité de ce type d'initiative pour produire rapidement et en masse, des jeux de données nouveaux permettant de répondre à de nouvelles questions en écologie.

« feuilles » (scan ou photo), « fruits », et « tiges ». Ce jeu de données intitulé Pl@ntViews, a été produit par 327 contributeurs et permet l'illustration de 250 espèces ligneuses et herbacées (Goëau et al., 2013). Il s'agit de l'un des premiers jeux de données libres (licence CC-BY-SA), aussi vaste (en nombre d'espèces) et diversifié (en nombre d'images), permettant de refléter la très grande hétérogénéité visuelle des plantes grâce à la mobilisation d'un grand nombre d'appareils photo numériques exploités tout au long de l'année, par un grand nombre de personnes.

En parallèle de la constitution de jeux de données de plus en plus vaste, et continuellement révisés, nous avons travaillé sur le développement d'outils informatiques dédiés à la navigation au sein de banques visuelles de plantes, avec un focus particulier sur la recherche des noms des plantes. Ces outils, web ou mobiles, ont alors été enrichi d'un nombre croissant d'observations pour arriver en septembre 2014 à la constitution d'un jeu de données de 108 000 images illustrant 4 390 espèces de la flore de France métropolitaine. La présentation de ce jeu de données, et de l'infrastructure à travers laquelle il a été produit et exploité est effectuée dans la section suivante.

Résultats

Les projets de sciences participatives menés à travers le monde et plus particulièrement en Europe ont prouvé la très grande capacité de ce type d'initiative pour produire rapidement et en masse, des jeux de données nouveaux permettant de répondre à de nouvelles questions en écologie. L'animation des réseaux humains impliqués dans le projet Pl@ntNet a ainsi permis de confirmer cela, en permettant la constitution d'un jeu de données structuré, unique en son genre, parmi les plus vaste et diversifié au niveau mondial. Celui-ci se compose de 108 000 images illustrant 4 390 espèces, réparties comme suit : 18 427 images ports, 43 810 de fleurs, 29 432 de feuilles, 10 403 de fruits, 6 096 de tiges. Ce jeu d'images porte sur X familles de trachéophytes référencées dans la BDTFX. Il est actuellement principalement produit à travers 2 outils logiciels distincts :



- le carnet en ligne (CEL), un outil web de gestion d'observations individuelles (<http://www.tela-botanica.org/page:cel>)
- Pl@ntNet-mobile, (Goëau et al., 2014) un outil mobile de navigation interactif [fig. 1] et d'aide à l'identification des plantes par l'image (<http://identify.plantnet-project.org/fr/>). Contrairement au CEL, cet outil permet de parcourir l'ensemble des données agrégées à travers le carnet en ligne, mais ne sont visibles que celles révisées collaborativement à travers les outils de révision IdentiPlante (<http://www.tela-botanica.org/appli:identiplante>) et PictoFlora (<http://www.tela-botanica.org/appli:pictoflora>). Il intègre un système d'aide à l'identification automatique de plantes à partir de photos qui s'appuie sur un moteur de recherche visuel novateur [Figure 1]. Ce système permet à l'utilisateur de soumettre jusqu'à 5 photographies d'une même plante selon différentes vues détaillées de ses organes (fleur, feuille, fruit, écorce). Les photos sont alors comparées automatiquement avec les images de la base de données botanique développée. L'application permet alors de trouver le nom de l'espèce d'une plante à partir d'une liste de résultats, si celle-ci est suffisamment illustrée [Figure 2]. Le nombre d'espèces traitées par cette application, ainsi que le nombre d'images exploitées, s'accroissent continuellement grâce aux contributions à ce projet. Elle fonctionne d'autant mieux que les images soumises sont centrées sur un organe ou une partie de plante bien précise. Depuis 2011, le moteur de recherche visuelle développé est évalué à travers la tâche d'identification des plantes de la campagne d'évaluation internationale ImageCLEF (2011-2013) (Goëau et al., 2013b) puis LifeCLEF (2014) (Joly et al., 2014a).

En travaillant sur (i) la constitution d'un jeu de données structuré, (ii) le développement d'outils de recherche innovants, (iii) la constitution d'une communauté de volontaires, l'initiative Pl@ntNet a permis l'agrégation d'un volume très important d'observations botaniques (plus de 600 000 observations actuellement en cours d'analyse), issues des requêtes d'identification de la communauté d'utilisateurs. L'infrastructure mise en œuvre a été exploitée par une très vaste typologie d'utilisateurs (de novices à chercheurs confirmés), dans près de 150 pays du monde, par plus de 300 000 personnes. Ce constat permet d'envisager une nouvelle forme de collecte de données botaniques dans le futur, s'appuyant sur des réseaux humains plus diversifiés qu'actuellement, et impliquant notamment un plus grand nombre d'acteurs de la société civile (Joly et al., 2014b).



Figure 1 : Acquisition d'une photo de fleur, pour la réalisation d'une requête visuelle à travers l'application Pl@ntNet-mobile (version Android).

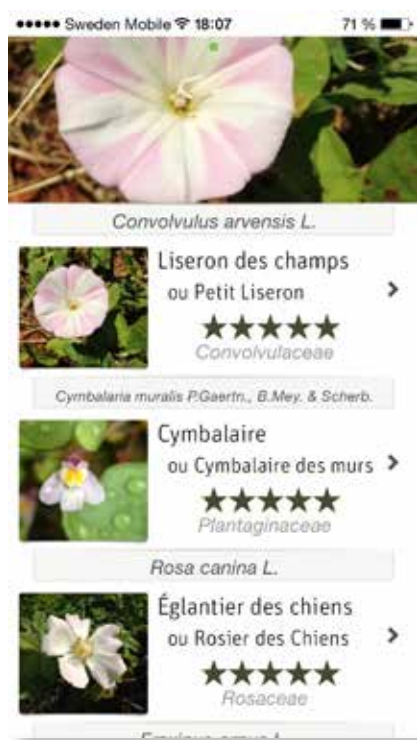


Figure 2 : Résultat d'une requête visuelle, à travers l'application Pl@ntNet (version IOS). La photo en haut de l'interface, est l'une des photos de la requête utilisateur, les espèces situées sous cette image, sont illustrées par les photos les plus proches de la requête.

Discussion

L'essor des sciences citoyennes a contribué dans de nombreuses disciplines à collecter des données dans des volumes et gammes spatio-temporelles, jamais égalées jusque-là (Newman G. et al., 2012). Compte-tenu de la difficulté relative de l'identification taxonomique en botanique, avec un nombre de plusieurs milliers d'espèces par pays, la mise en œuvre d'une démarche scientifique, engageant le grand public sur de grands ensembles taxonomiques est plus difficile que pour d'autres disciplines (ornithologie, biologie marine, etc.). Pour pallier à cela, différentes initiatives visent à développer des approches novatrices d'aide à l'identification des plantes. Xper (Ung, V. et al., 2010) est l'une d'entre elle, et permet la constitution de bases descriptives de manière collaborative. LeafSnap (Kumar N. et al., 2012) fût quant à elle la première approche mobile basée sur des méthodes de reconnaissance visuelle. Elle a permis d'ouvrir la porte à d'autres initiatives développant ou exploitant des techniques de « machine learning », telles que « FlowerChecker », « Folia » (Cerutti et al., 2011), et « Pl@ntNet ». Ces approches qui permettent d'impliquer rapidement des novices en botaniques, s'avèrent particulièrement encourageantes pour leur développement dans des contextes éducatifs. Au-delà des contextes floristiques limités dans lesquels ont été mobilisées ces approches, elles offrent des opportunités énormes dans le futur, pour l'agrégation en masse de données. Un enjeu majeur pour la communauté scientifique, sera la mise en place de flux de validation appropriés de ces données d'origines variées, mais qui devrait permettre d'impliquer un plus grand nombre d'acteurs dans le suivi de l'évolution de la biodiversité et de la phénologie des plantes, à des échelles inégalées jusque-là.

Remerciements

Ce projet est soutenu par Agropolis Fondation (<http://www.agropolis-fondation.fr/>) qui en a fait son premier projet étandard. Les auteurs souhaitent également remercier l'ensemble des volontaires qui ont participé à ce travail notamment en partageant des milliers d'observations botaniques à travers l'infrastructure réalisée.

Références bibliographiques

- Bidwell A. (2013). *The Academic Decline: How to Train the Next Generation of Botanists*. US News, Nov. 12, 2013.
- Cerutti G., Tougne L., Vacavant A., Coquin D., (2011). *A Parametric Active Polygon for Leaf Segmentation and Shape Estimation*. In Advances in Visual Computing, vol. 6938, Lecture Notes in Computer Science serie, ed. Bebis, G. and Boyle, R. and Parvin, B., Koracin, D. and Wang, S. and Kyungnam, K. and Benes, B. and Moreland, K. and Borst, C. and DiVerdi, C. and Yi-Jen, C. and Ming, J., Springer Berlin Heidelberg, pp. 202-213.
- Farnsworth E. J., Chu M., Kress W. J., Neill A. K., Best J. H., Pickering J., Stevenson R. D., Courtney G. W., VanDyk J. K., and Ellison A. M. (2013). *Next-Generation Field Guides*. BioScience, 63(11), 891-899.
- Goëau, H., Bonnet, P., Joly, A., Affouard, A., Bakic, V., Barbe, J., Dufour, S., Selmi, S., Yahiaoui, I., Vignau, C., (2014). *Pl@ntnet mobile 2014: Android port and new features*. In: Proceedings of International Conference on Multimedia Retrieval, p. 527. ACM.
- Goëau H., Bonnet., Joly A., Bakic V., Barthélémy D., Boujemaa N. & Molino J.-F., (2013a). *The ImageCLEF 2013 Plant Identification Task*. CLEF 2013 Working Notes, Valencia, Spain, 2013.
- Goëau H., Bonnet., Joly A., Bakic V., Barthélémy D., Boujemaa N. and Molino J.-F., 2013b. *The ImageCLEF Plant Identification Task 2013*. Proceedings of the 2nd ACM International Workshop on Multimedia Analysis for Ecological Data, MAED 2013, Barcelona, Spain.
- Goëau H., Bonnet P., Joly A., Boujemaa N., Barthélémy D., Molino J.-F., Birnbaum P., Mouysset E. & Picard M., 2011a. *The CLEF 2011 plant images classification task*. CLEF 2011 working notes, Amsterdam, The Netherlands, 2011, pp. 1-23.
- Goëau H., Joly A., Yahiaoui I., Bonnet P. & Mouysset E., 2011b. *Participation of INRIA & Pl@ntNet to ImageCLEF 2011 plant images classification task*. CLEF 2011 working notes, Amsterdam, The Netherlands, 2011, pp 1-13.
- Joly A., Goëau H., Bonnet P., Bakic V., Barbe J., Selmi S., Yahiaoui I., Carrée J., Mouysset E., Molino J.F., Boujemaa N., Barthélémy D., 2014b. *Interactive plant identification based on social image data*. Ecological Informatics. Vol. 23, Sept. 2014, Pages 22–34.
- Joly A., Goëau H., Glotin H., Spampinato C., Bonnet P., Vellinga W.P., Planque R., Rauber A., Fisher R., Müller H. 2014a. *Lifeclef 2014: Multimedia life species identification challenges*. In: Information Access Evaluation. Multilinguality, Multimodality, and Interaction, pp. 229-249. Springer.
- Kumar N., Belhumeur P., Biswas A., Jacobs D., Kress J. Lopez, I., and Soares, J., 2012. *Leafsnap: A computer vision system for automatic plant species identification*. Computer Vision-ECCV 2012, pp. 502-516, 2012. Springer.
- Leggett R., Kirchoff B. K., 2011. *Image use in field guides and identification keys: Review and recommendations*. AoB Plants 2011.
- Newman G., Wiggins A., Crall A., Graham E., Newman S., and Crowston K., 2012. *The future of citizen science: emerging technologies and shifting paradigms*. Frontiers in Ecology and the Environment 10: 298–304.
- Ung, V., Dubus, G., Zaragüeta-Bagils, R., and Vignes-Lebbe, R., 2010. *Xper2: introducing e-taxonomy*. Bioinformatics, vol. 26, Number 5, pp. 703-704. Oxford Univ Press.
- Whitfield J., 2012. *Superstars of botany: Rare specimens*. Nature News, Nature Publishing Group, Apr 25, 2012.