Collaborative validation of visual data through the Pl@ntNet identification system

To cite this version:
Samuel Dufour-Kowalski, J. Salinier, A. Peronnet, J.-F Molino, Jennifer Carré, et al.. Collaborative validation of visual data through the Pl@ntNet identification system. TDWG 2013 - Annual Conference, Oct 2013, Florence, Italy. hal-03885113

HAL Id: hal-03885113
https://hal.inrae.fr/hal-03885113
Submitted on 5 Dec 2022

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.
Collaborative validation of visual data through the Pl@ntNet identification system

Accurate knowledge of **plants** (distribution and ecology) is essential for **sustainable agriculture** and **biodiversity conservation**.

But accessing basic information about plants is still challenging.

**Botanical data** is:
- *decentralized* and **heterogeneous**
- *complex* (un-structured tags, empirical measurements,...)
- *sparse* and **incomplete**
  - huge & unknown number of species
  - “long tail distribution” (1 record per species !)
Towards bridging the taxonomic gap

Identifying and naming plants is a very difficult task

Plant names are the KEY to access and to enrich botanical information on plants

Ailanthus altissima (Mill.) Swingle

Invasive species in Europe

Tree of heaven
Faux verni du Japon
Árbol de los dioses
Malodorous tree

Ornamental species
Towards bridging the taxonomic gap

Possible solutions

- **Collaborative Information Systems**
  Sharing and speeding up integration of raw data

- **Large audience Identification Tools**
  Multimedia image retrieval techniques...

But...

**Few, small, biased datasets**

- Information system
- Identification tool

- Validation / Data quality?
Pl@ntNet Workflow

- Image sharing and retrieval app for plant identification
- Shared observations (Creative Commons)
- Botanical obs. management system (pictures, species, date, GIS, author)
- Collaborative images annotation system
  - Tags (flowers, leaves, etc.)
  - Quality evaluation
- Collaborative Identification
  - Identification suggestion
  - Identification vote
  - Forum

Visualization engine

Validation + Enrichment

PictoFlora

Moteur d'indexation visuel

Pl@ntNet mobile

Pl@ntNet mobile app

- Goëau & al., 2013. ACMM.

Public version

- 70,000 images
- 3,700 species

BETA

- 105,000 images
- 5,000 species
Dataset based on social network of botanists

- 21,500 members
  - From amateur to expert botanists
- Hundreds of contributors with different skills
  - with their own scanners, cameras & Smartphone
- Thousands of individual botanical records
  - at different growing stage,
  - different periods of the year,
  - under different light conditions (raining, sunny, ...)

A huge visual diversity to canalise
Autumnal variability of the lamina color on *Cotinus coggygria* Scop. (Eurasian smoketree)

Leaf at different growing stage of *Platanus x hispanica* Mill ex. Münchh. (London plane)

Growing stage: two compound leaves from the same tree! *Gleditsia triacanthos* L. (Honey Locust)

Lobe number and deep of leaf lobes on *Ficus carica* L. (Common fig)

Shooting conditions and used devices, *Acer platanoides* L. (Norway mapple)

Leaflets number variability on *Fraxinus angustifolia* Vahl (Narrow-leafed Ash)

Leaf diversity

# Users = # localities # seasons # environments # climate # ecosystems # devices
Pl@ntViews dataset

Flower diversity

**COLOR**
- Brown
- White
- Green
- Rose
- Blue
- Yellow

**Symmetry**
- Radial
- Bilateral

**Structure**
- Number of petals
  - 4
  - 5
  - 6
  - >>6

**Orientation**
- Face
- Profil

**Size**
- small
- middle
- big
A collaborative website for data validation and annotation

IdentiPlante
Botanical records validation

PictoFlora
Picture validation and annotation
Identiplante, for Identification validation

Web application

Individual URL for each Botanical record can be logged but not necessary

User can see any botanical record, from any contributors

Botanical record = Image(s) + Taxa name + Place + Date + Contributor name

National taxonomic indexes

National localities indexes
Identiplante, for Identification validation

Initial identification
Based on vernacular name
Several suggestions by members of the social network
Suggestions can be commented… and then discussed

Community members vote for any suggestion

Define the most probable species
PictoFlora, for Tags and image quality evaluation

User can see its own votes
Each picture can be tagged
According to Visual concepts of PI@ntNet Identification app
Results

The most probable species name:
- according to collaborative votes
- among the national species index

We don’t use records with determination
At the family or genus level

PictoFlora

Pictures:
- With one tag only
- A mean of more than 3 stars

Pl@ntView dataset

- 1100 users
- 10500 votes
- 7000 propositions
- 900 Comments

- 850 users
- 63 000 tags
- 137 000 votes

70 000 images / 3 700 species
Future directions

- Invest in user profile (for a specific region, or group of taxa)
- Use all the data according to their quality
- Use some automatic algorithm to tag data
- Use of meta data in the identification and the validation process (localisation and/or date).
- Apply this workflow on other botanical (or non botanical) datasets
Thank You !!!