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

Samuel Dufour-Kowalski, Jeremy Salinier, Aurelien Peronnet, Jennifer Carré, Jean-Pascal Milcent, Hervé Goëau, Alexis Joly, Jean-François Molino, Nozha Boujemaa, Pierre Bonnet, et al.

Pl@ntNet

http://www.plantnet-project.org/

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

S. Dufour-Kowalski, J. Salinier, A. Peronnet, J. Carré, J.-P. Milcent, H. Goëau, A. Joly, N. Boujemaa, P. Bonnet, J. Barbe, J.-F. Molino, D. Barthélémy

AMAP

Equipe IMEDIA

Inria

Tela Botanica

Zenith

cirad

INRA

Inria

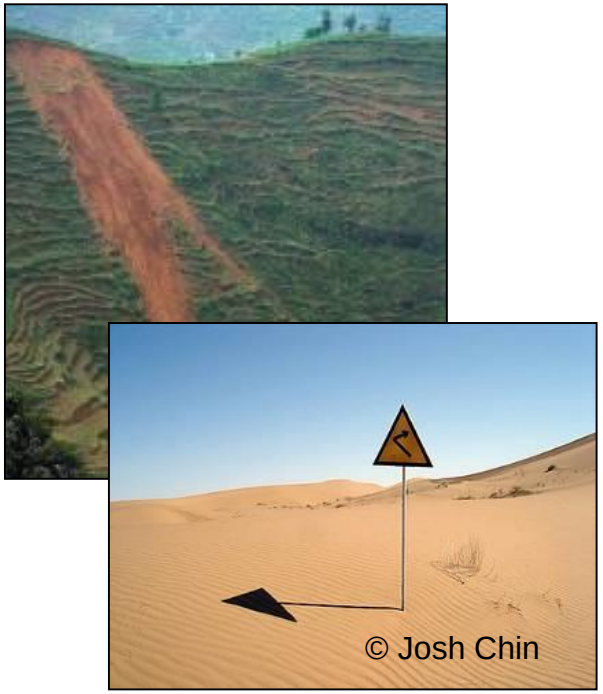
IRD

Institut de recherche pour le développement

agropolis fondation

Context & challenges

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

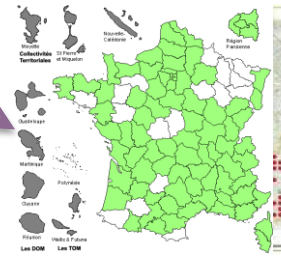
Ailanthus altissima (Mill.) Swingle

Tree of heaven
Faux verni du Japon

Árbol de los dioses
臭椿 *chòuchūn*
"Malodorous tree"

Ornamental species

Invasive species in Europe



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 …

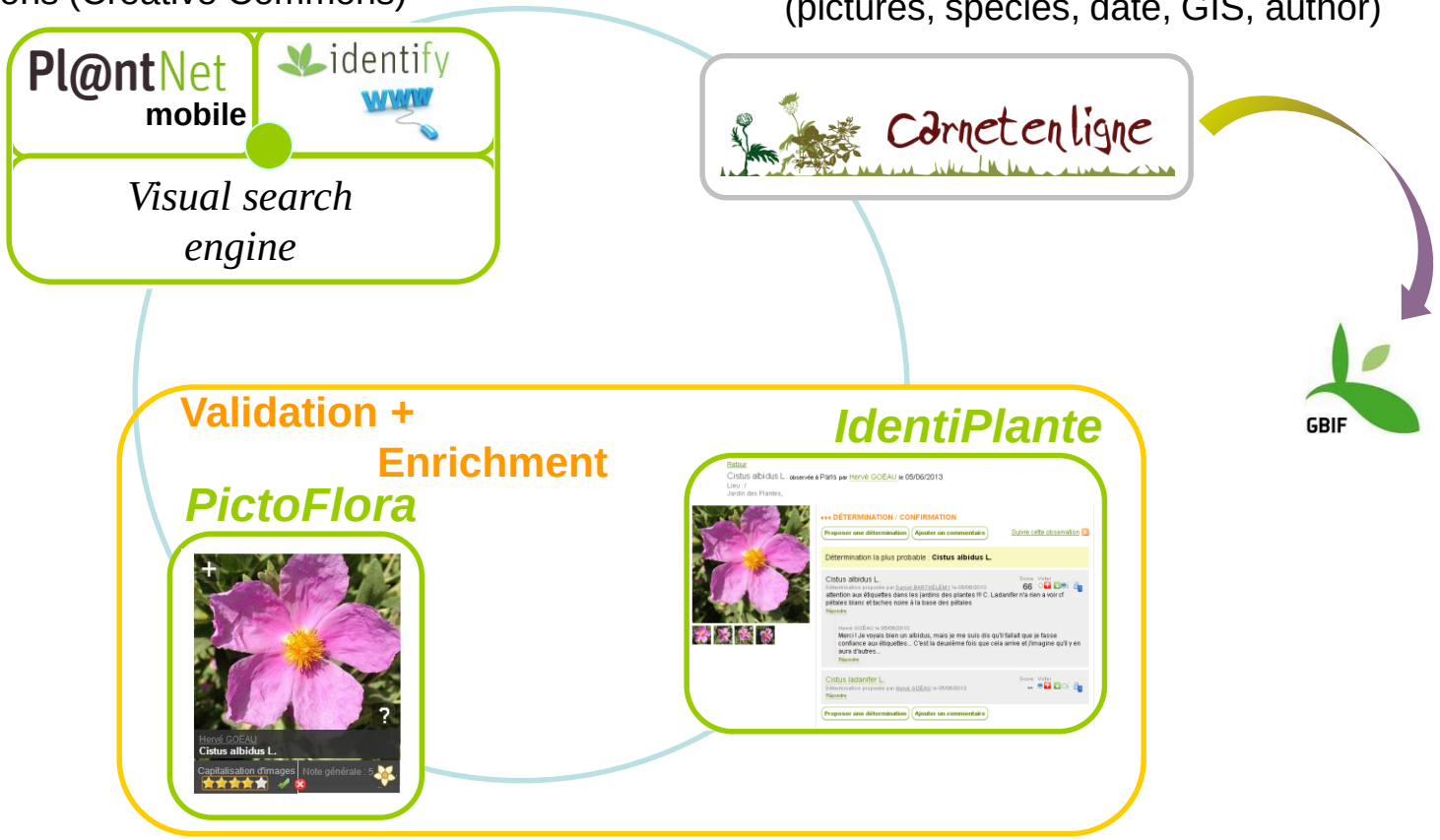
Few, small, biased datasets

- **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

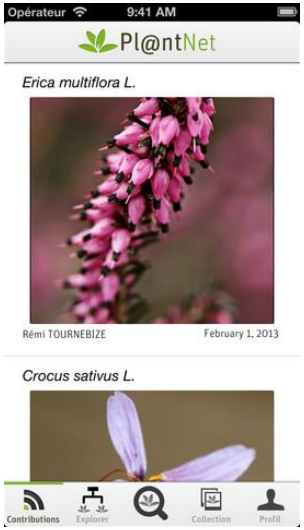
- **Collaborative Identification**
  - Identification suggestion
  - Identification vote
  - Forum

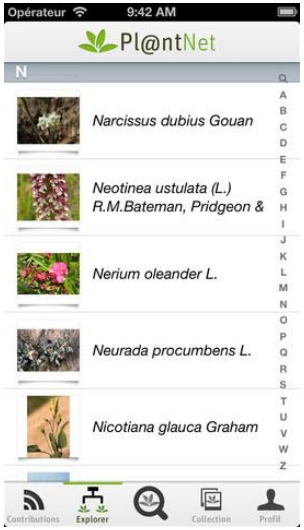
- Joly & al., 2013. Ecological informatics.

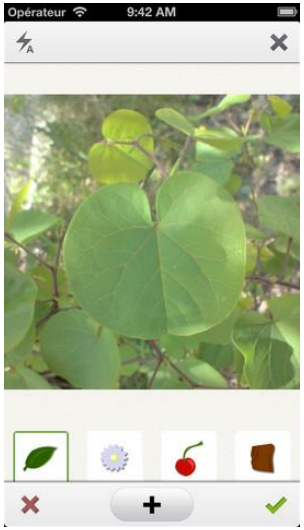
Pl@ntNet mobile app

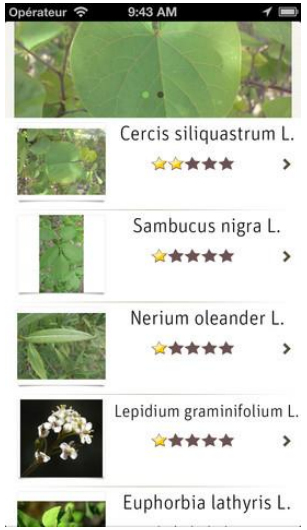
- Goëau & al., 2013. ACMM.

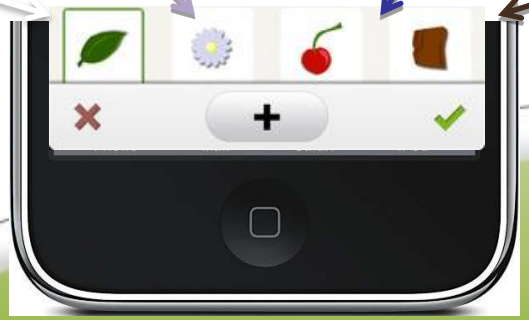
iPhone











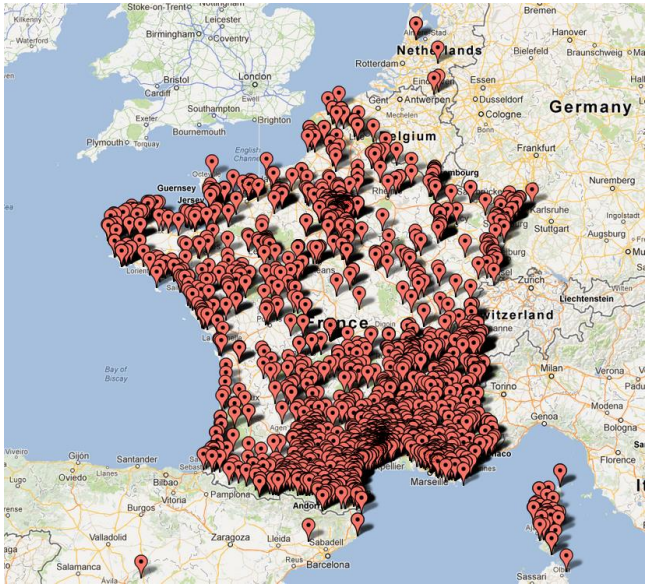
BETA

Public version

70 000 images	105 000 images
3 700 species	5 000 species

Dataset based on social network of botanists

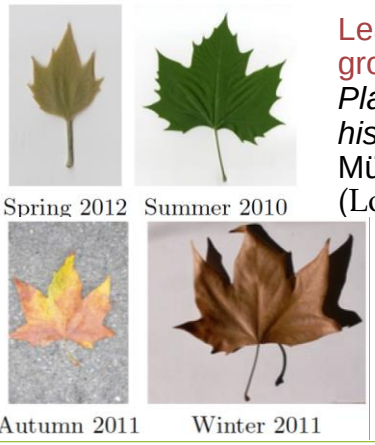
- 21 500 members
  - From amateur to expert botanists
- Hundreds of contributors with different skills
  - with their own scanners, cameras & Smartphone
- Thousands of individual botanical records
  - at different growing stage,
  - different periods of the year,
  - under different light conditions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



A huge visual diversity to canalise

Pl@ntViews dataset

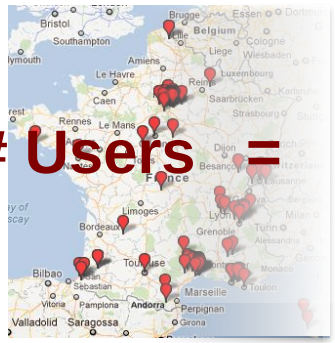
Leaf diversity



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

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

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



 # Users =

 # localities
 # seasons
 # environments
 # climate
 # ecosystems
 # devices

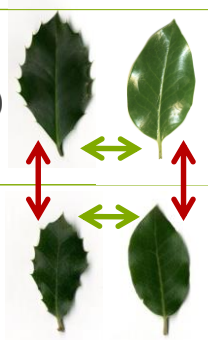
Leaflets number variability on *Fraxinus angustifolia* Vahl (Narrow-leafed Ash)

Autumnal variability of the lamina color on *Cotinus coggygria* Scop. (Eurasian smoketree)

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

Ilex aquifolium L. (European holly)

Quercus ilex L. (Holm oak)



Intra-species diversity versus visual similarities between species

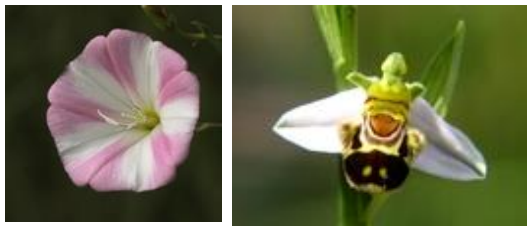
Pl@ntViews dataset

Flower diversity

COLOR

Brown White Green Rose Blue Yellow

Sym ¦ metry



Radial **Bilateral**

Structure

4 5 6 >>6

Number of petals

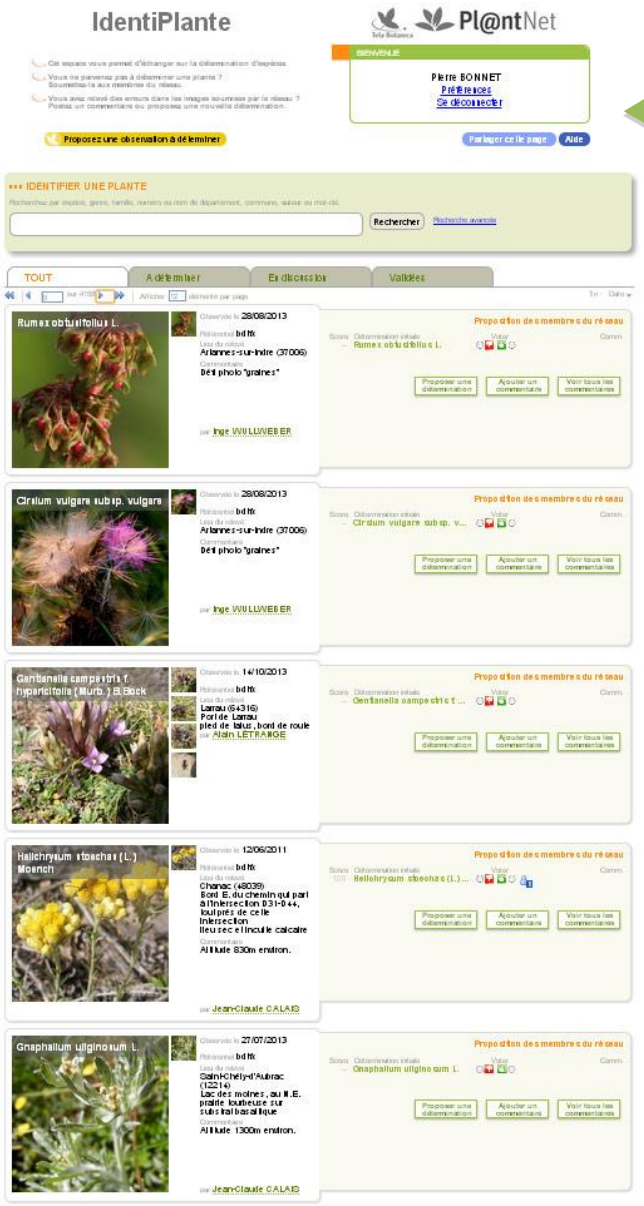
Orientation

Face **Profil**

Size

small middle big

A collaborative website for data validation and annotation

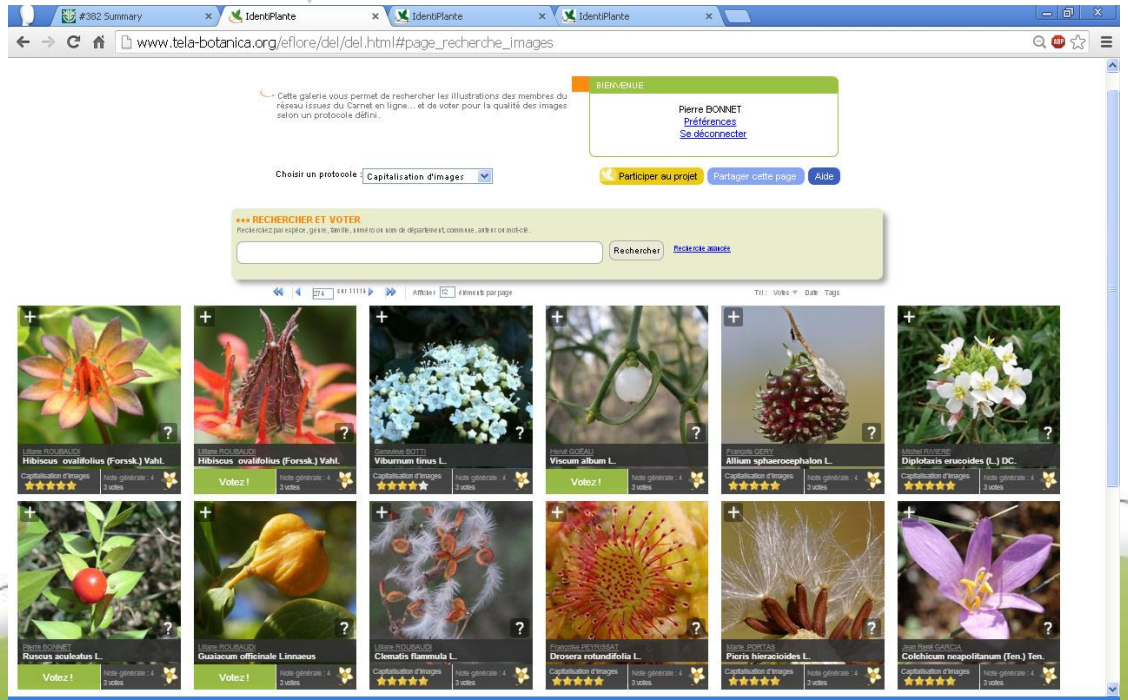


IdentiPlante

Botanical records validation

PictoFlora

Picture validation and annotation



IdentiPlante, for Identification validation

Web application

Individual URL for each Botanical record

Users can be logged but not necessary

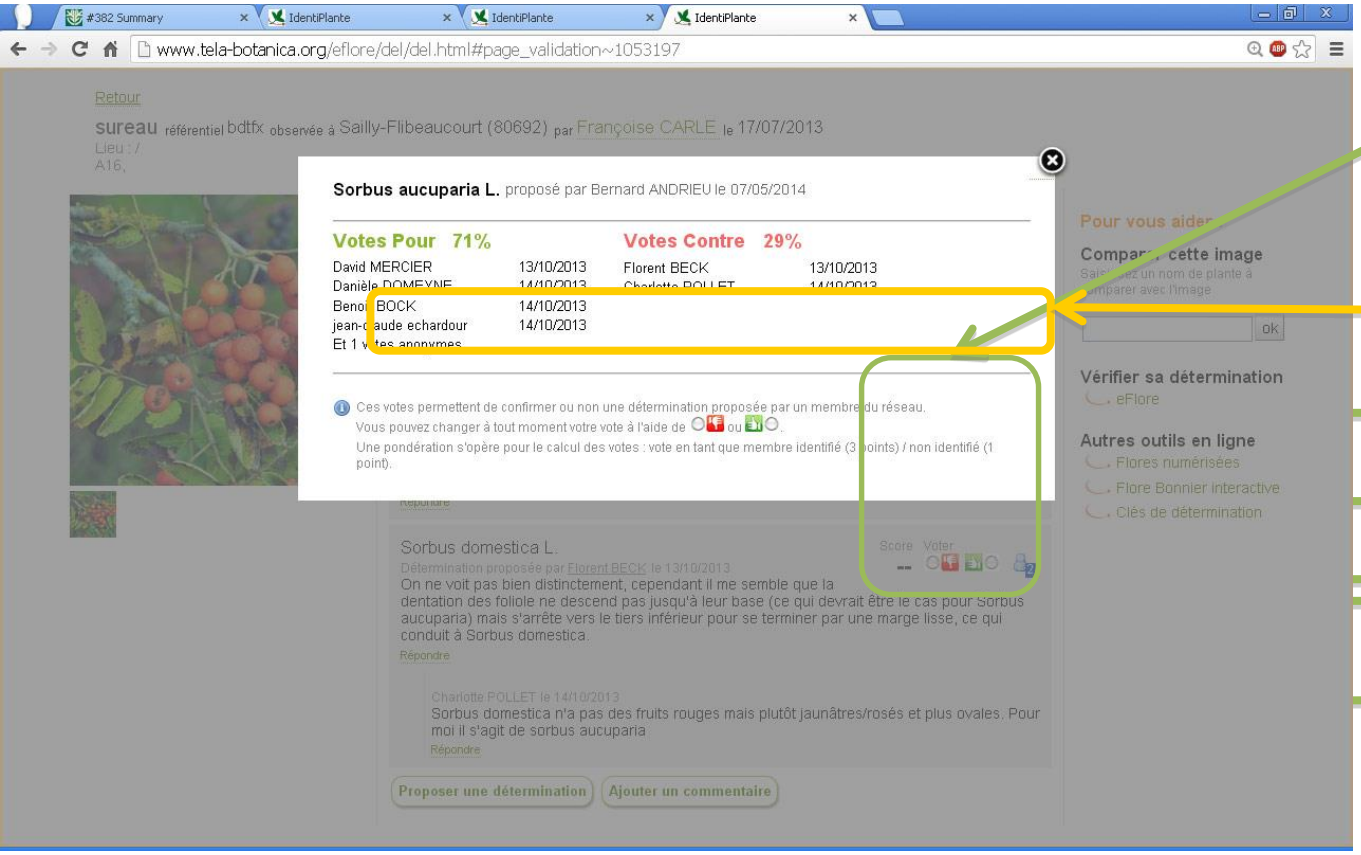
User can see any botanical record, from any contributors

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

National taxonomic indexes

National localities indexes

IdentiPlante, for Identification validation



Community members Vote for any suggestion

Define the most probable species

Several suggestions

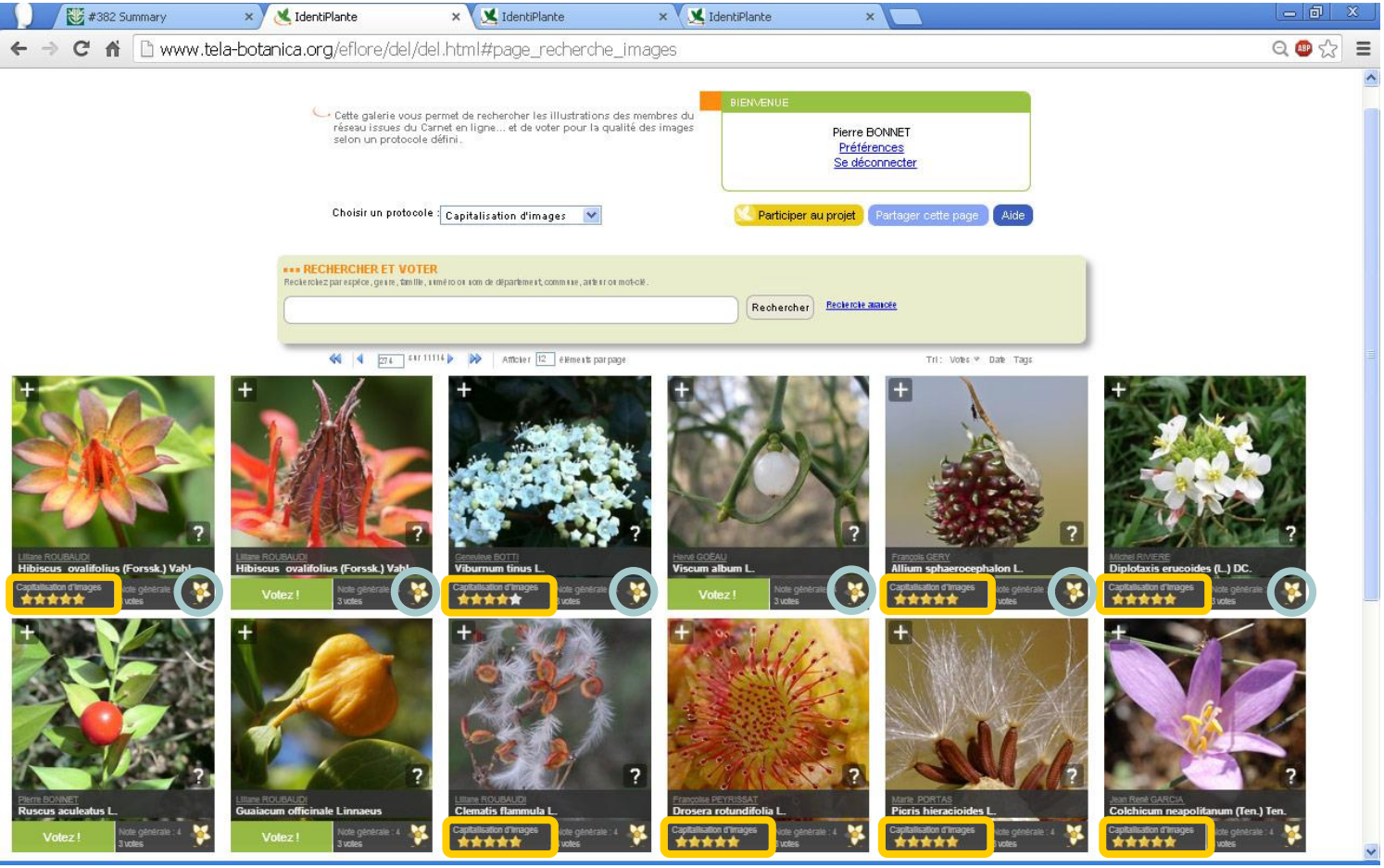
Initial identification by members of the social network or anonymous

Suggestions with vernacular name

Can be commented

… and then discussed

PictoFlora, for Tags and image quality evaluation

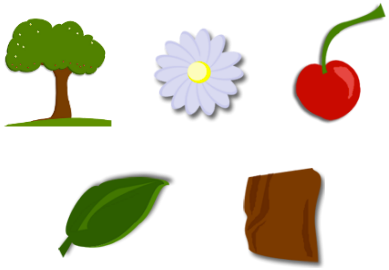


User can see its **own votes**

Each picture can be tagged

According to Visual concepts of Pl@ntNet Identification app

Or mean of votes of the community



Results

IdentiPlante

PictoFlora

The most probable species name :

- according to collaborative votes
- among the national species index

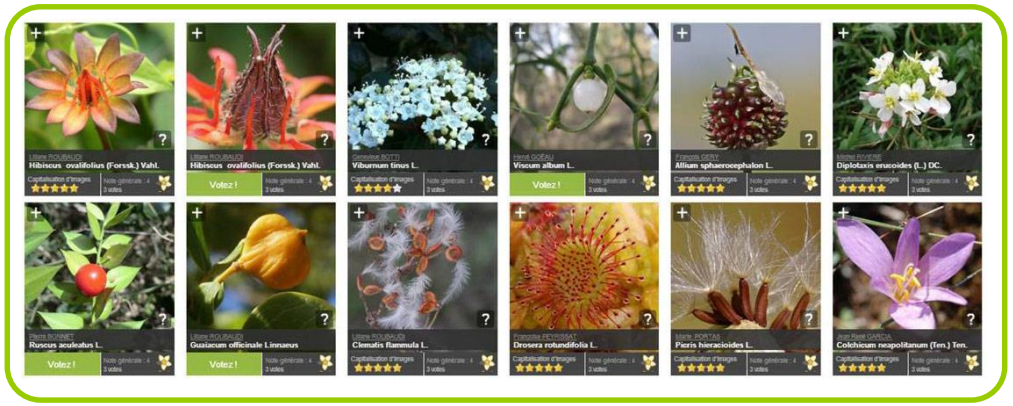
We don't use records with determination
At the family or genus level

Pictures :

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

Pl@ntView dataset

1100 users
10500 votes
7000 propostions
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

WWW.plantnet-project.org

Email : **contact@plantnet-project.org**

Thank You !!!