



From massive digitisation of Paris Herbarium to a nation-wide program



SPNHC Conference Cardiff 2014

A « conveyor belt » operation

- ▶ Building renovation is an opportunity for reconditionning, sorting and digitizing collections
- ▶ a dedicated industrial site running from 2008 to 2012



Costs

- ▶ Overall project cost: 24,5 Million €

Building 's renovation	12 000 000
Movers	900 000
Mounting specimens	3 200 000
Reconditioning, digitizing and ordering	6 700 000
Supplies	1 600 000
Storage	100 000



What we learned

- ▶ Written procedures ease negociation with contractor
- ▶ Quality control matters



Results

- ▶ 5,3 million images online ☺
- ▶ only a minimum set of attributes
 - ▶ scientific name
 - ▶ family
 - ▶ catalog number
 - ▶ folder color ≈ continent of origin

Coffea arabica L.

Rubiaceae

P04010591

AME



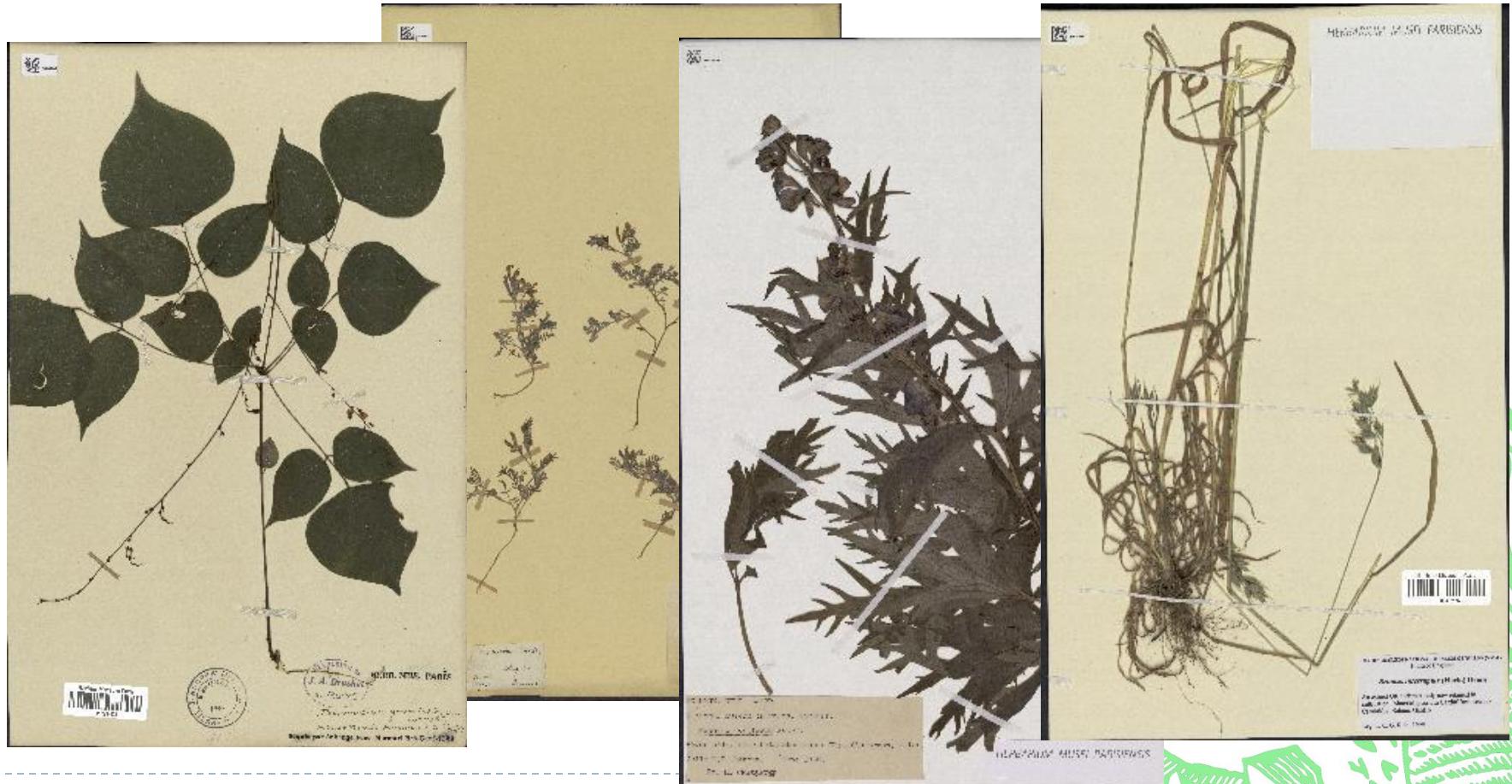
Optical Character Recognition

- ▶ Tesseract open source solution
- ▶ Months of computing

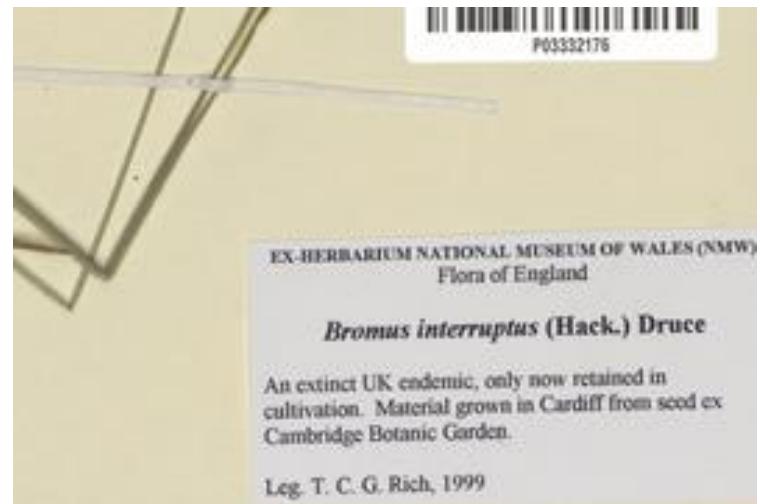
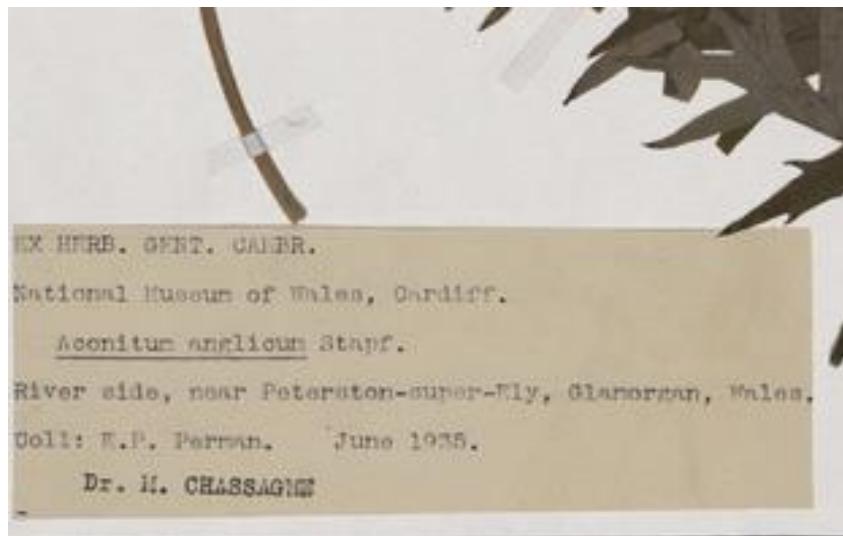
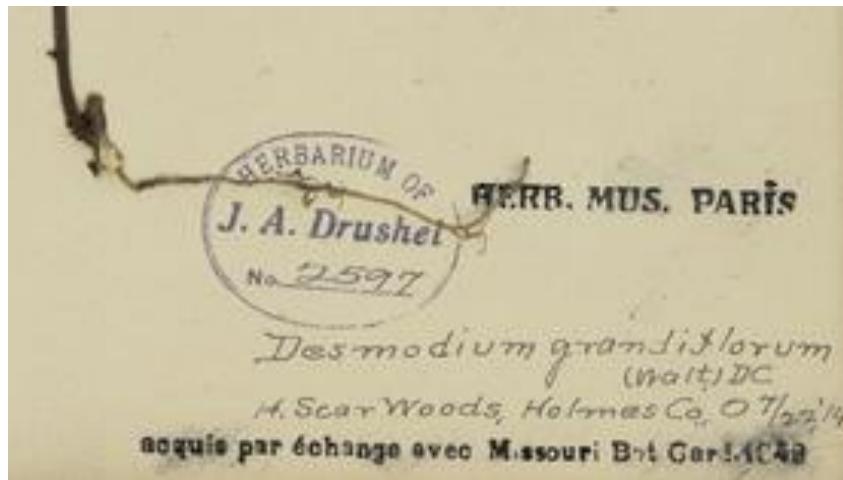


Don't expect too much from OCR

- ▶ Only 4 specimens mention 'Cardiff'
 - P02934033, P02721418, P03332176, P03458270



Zooming in the specimens « from Cardiff »

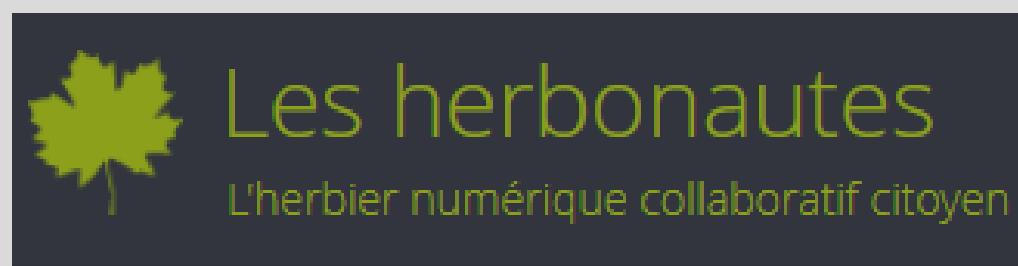


So what use of OCR ?

- ▶ As a filter for human driven-process
 - ▶ Reflora program
 - ▶ Crowdsourcing



Crowdsourcing label retranscription



- ▶ From December 2012 to May 2014

- 1405 contributors
- 69 085 specimens seen
- 783 311 contributions
- *to compare with 49613 transcriptions at herbarium*



Propose tasks you can complete



Fleurs des champs

A la rencontre des fleurs des champs.

100%

Objectif :
20000 contributions

[Participer](#)



Les Ficus

Pour ceux qui ont un peu peur des écritures manuscrites. Une mission facile pour s'habituer...

60%

Objectif :
20000 contributions

[Participer](#)



Les plantes subantarctiques

Mêmes aux confins de l'Antarctique, dans les îles parmi les plus isolées au monde, se cachent certaines plantes : oui mais lesquelles ? Cette mission permettra de les découvrir...

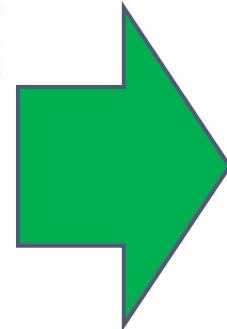
100%

Objectif :
22000 contributions

[Participer](#)



Make it fun



chlorofil

Nom: fifi lagrenouille
Ville: Lannion
Membre depuis: 1 an
Niveau: 6

12 missions 91 contributions 46 specimens

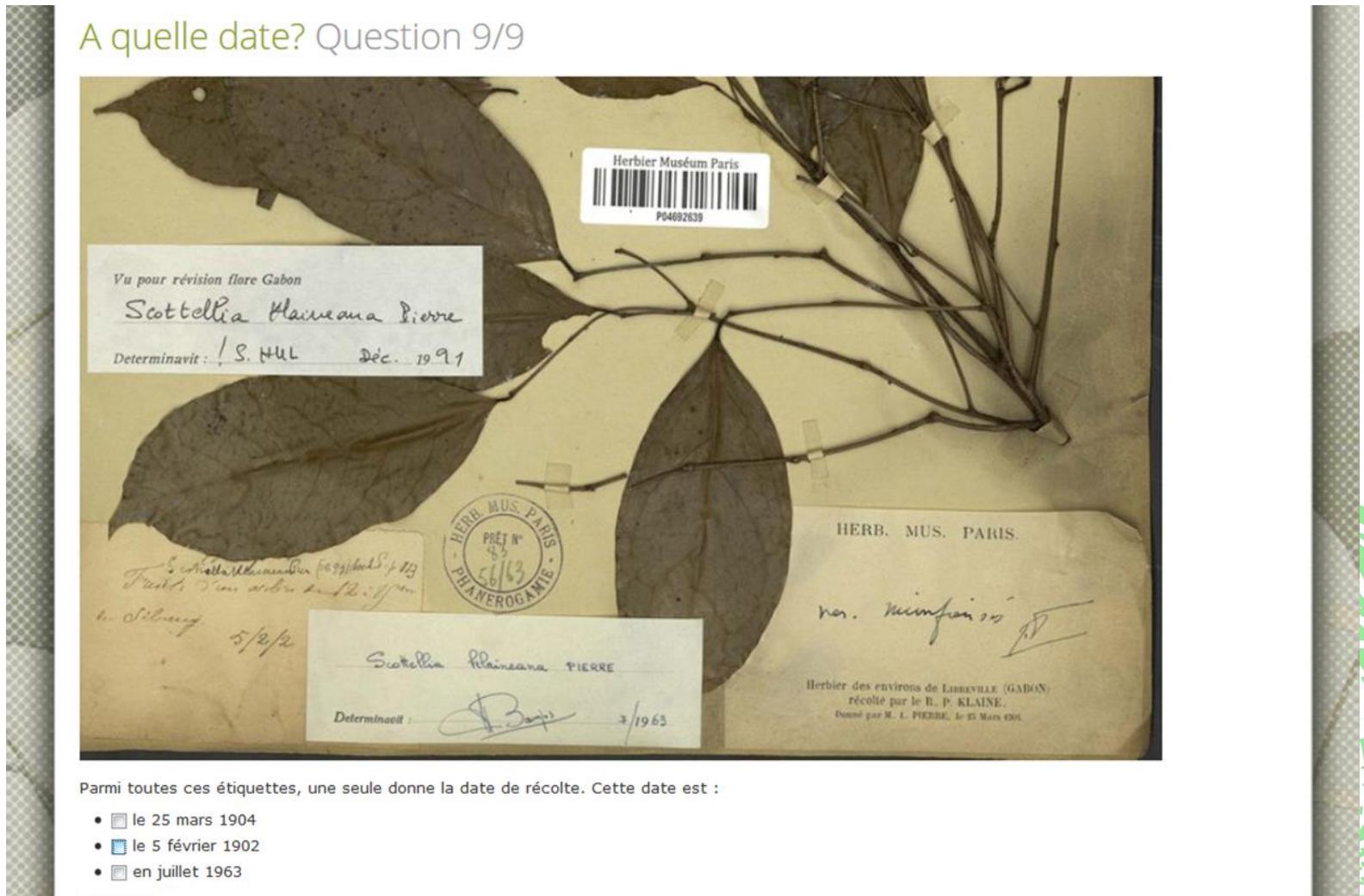
Cool Sherlock Podium Eclaireur Finisseur



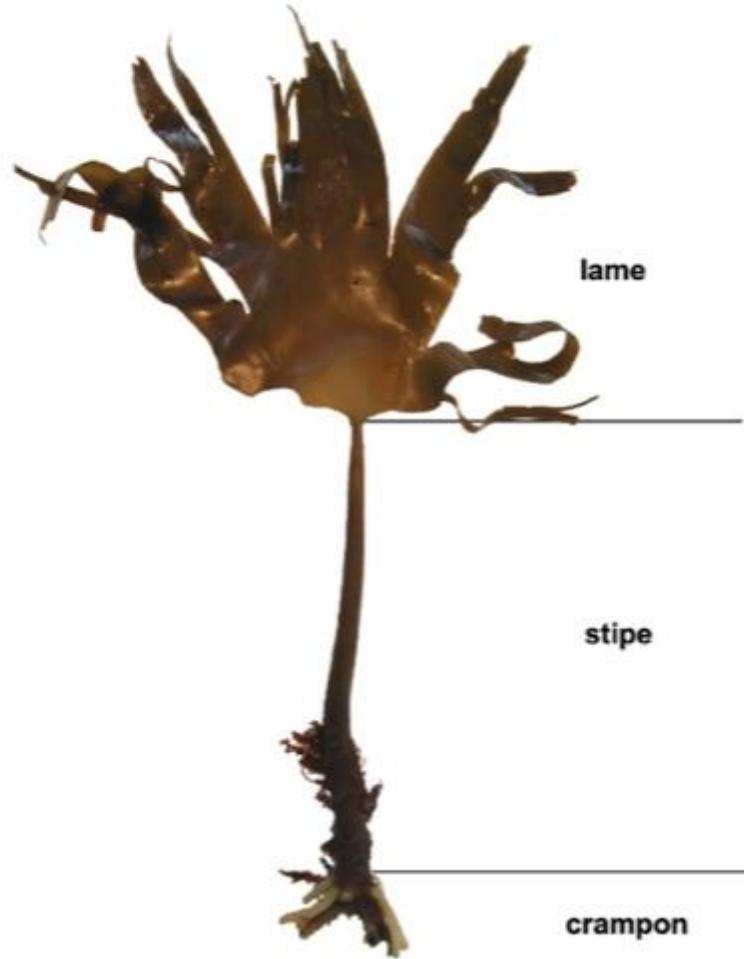
SPNHC2014 Digitisation symposium

Train the crowd

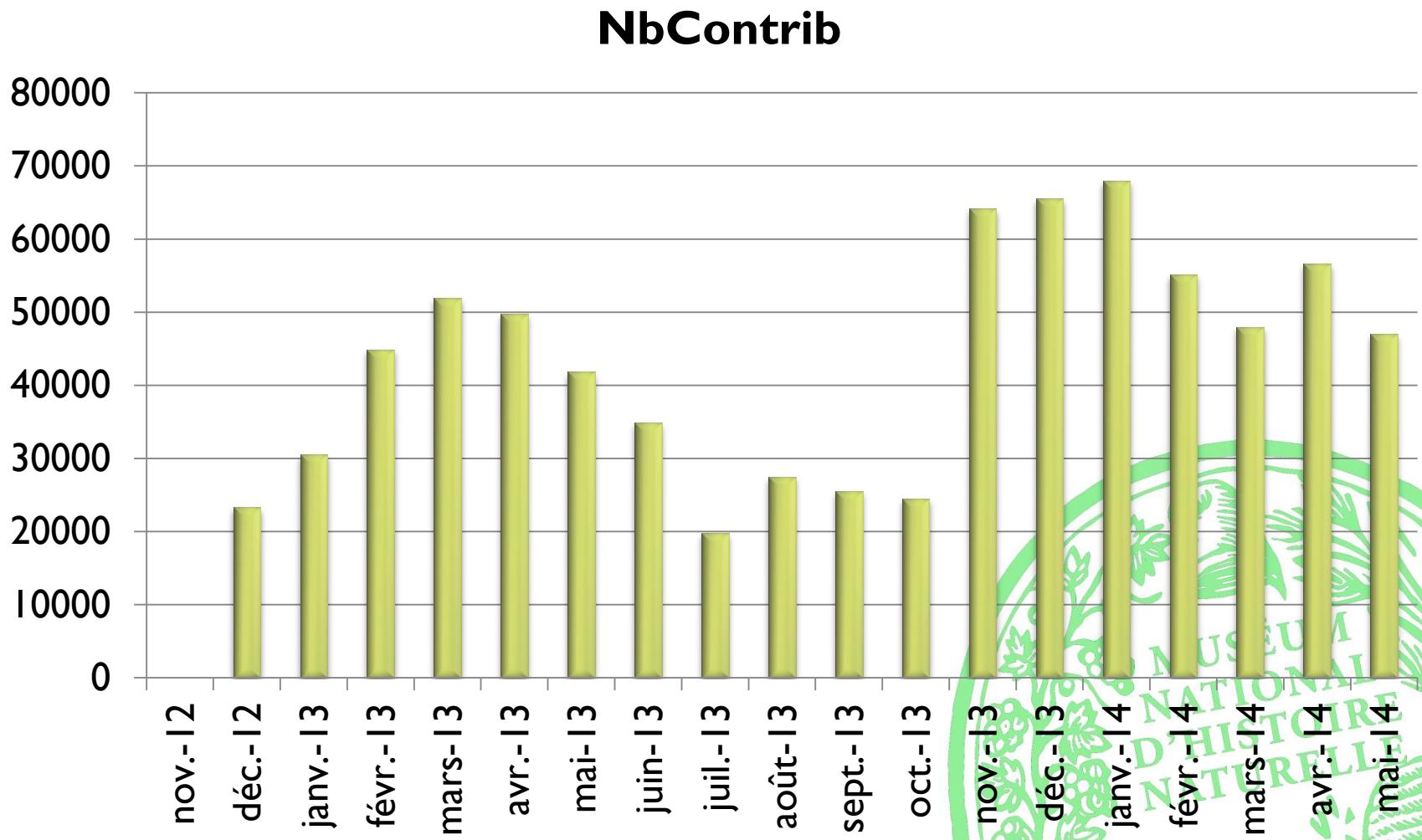
A quelle date? Question 9/9



Don't forget science



A lasting keen interest



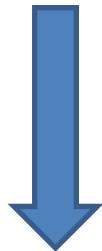
To a nation-wide program : recolnat

- ▶ Paris projects
 - ▶ Estimate the cost of herbarium digitisation
 - ▶ Show that the general public can contribute
- ▶ A consortium of institutions setup a nation-wide program
 - ▶ Granted in june 2013



E-ReColNat project

- ▶ IT layer of the TGIR RECOLNAT
- ▶ Purpose : Extract and make available data from the French naturalist collections
- Background: 100 million specimens collected in France during the last 350 years (millions of years in paleontology and geology)
- Scope: more than 60 partners and several disciplines: botany, zoology, paleontology, geology
- Method: Scan Documentation and shared massive web2.0

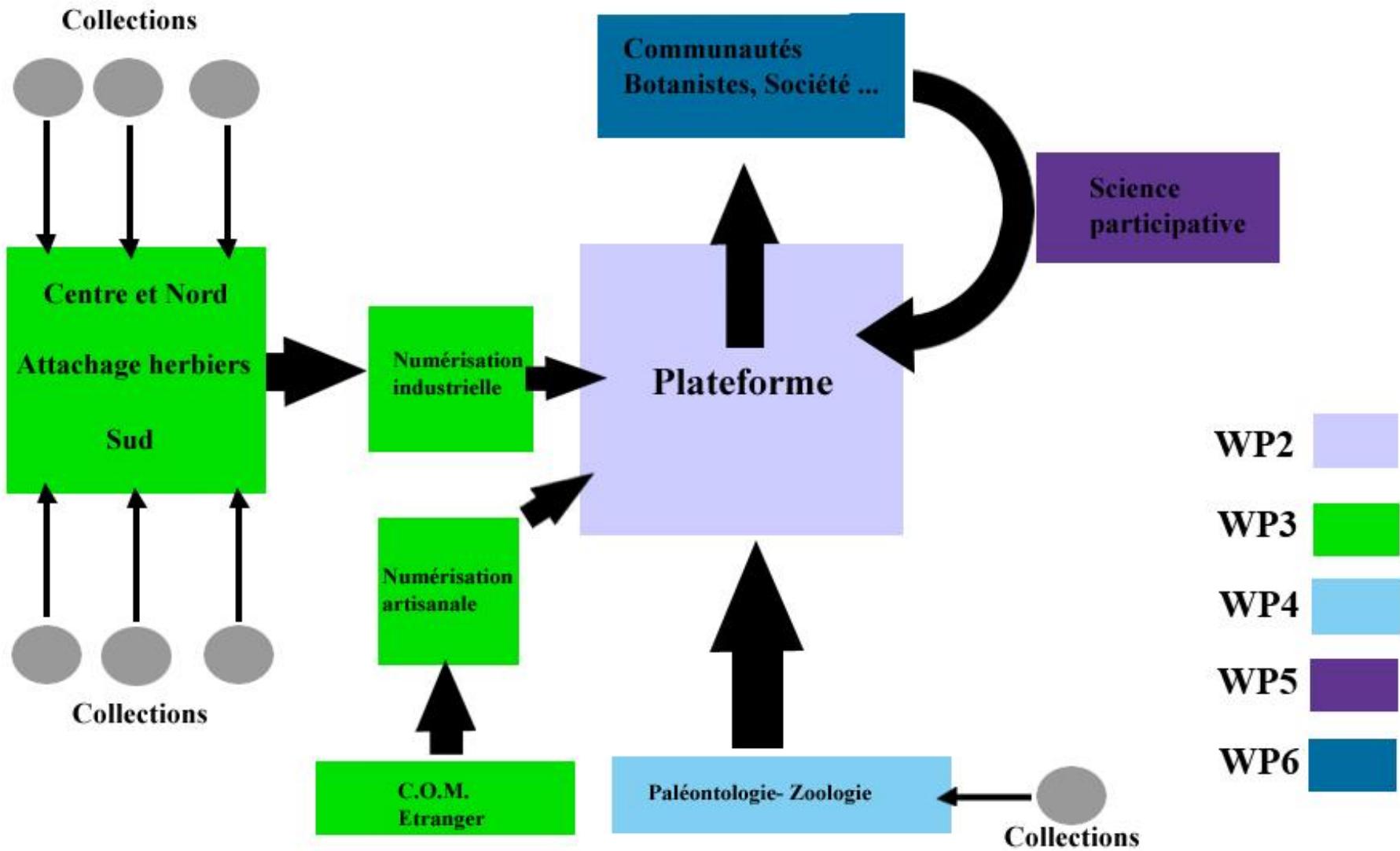


- ▶ **•Dynamic of the biodiversity**
- ▶ **•Information on Global Change**

SPNHC2014 Digitisation symposium



Workflow and packages



Any questions ?

- ▶ Marc Pignal 
- ▶ pignal @ mnhn.fr
- ▶ Simon Chagnoux
- ▶ chagnoux @mnhn.fr



<http://www.webdoc-herbier.com/#!91>

<http://lesherbonautes.mnhn.fr>



@LesHerbonautes

