

# GallicaSnoop

Projet de collaboration BnF / Inria / Ina

# Le projet

“Développer et implanter dans Gallica, la bibliothèque numérique de la BnF et à minima dans Gallica Studio, son laboratoire numérique, un prototype expérimentant **un outil de recherche iconographique interactif**”

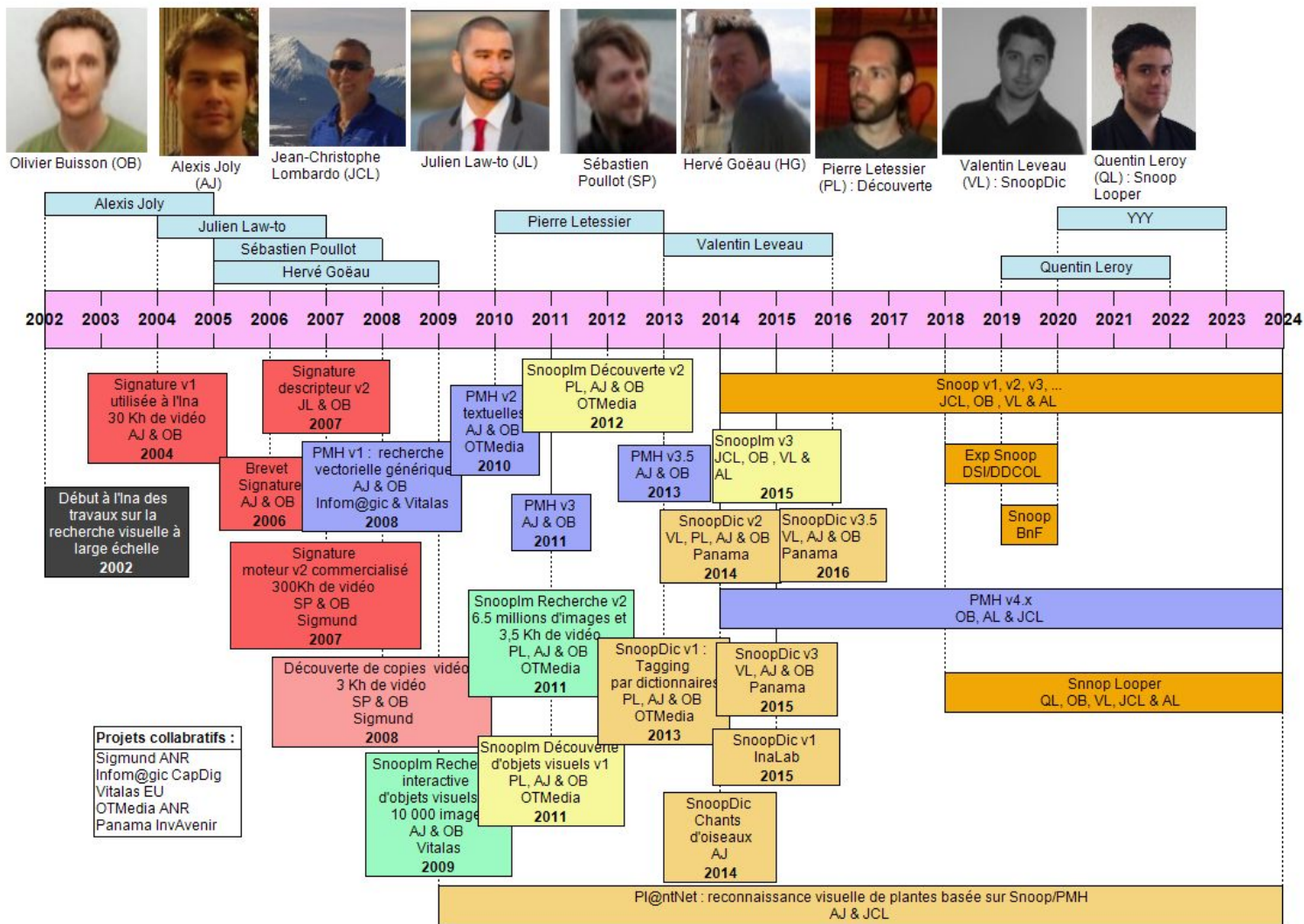
“Le problème majeur est que **les entités visuelles d'intérêt pour les utilisateurs ne sont pas connues à l'avance**. De plus, le coût humain de constitution d'une base d'apprentissage supervisée de milliers d'entités visuelles potentielles est un frein majeur.”

“**Implémenter** dans Gallica des **méthodes de recherche interactive** permettant à un utilisateur de créer à la volée des modèles d'**entités visuelles personnalisées** et de pouvoir ensuite les partager avec la communauté des autres utilisateurs.”

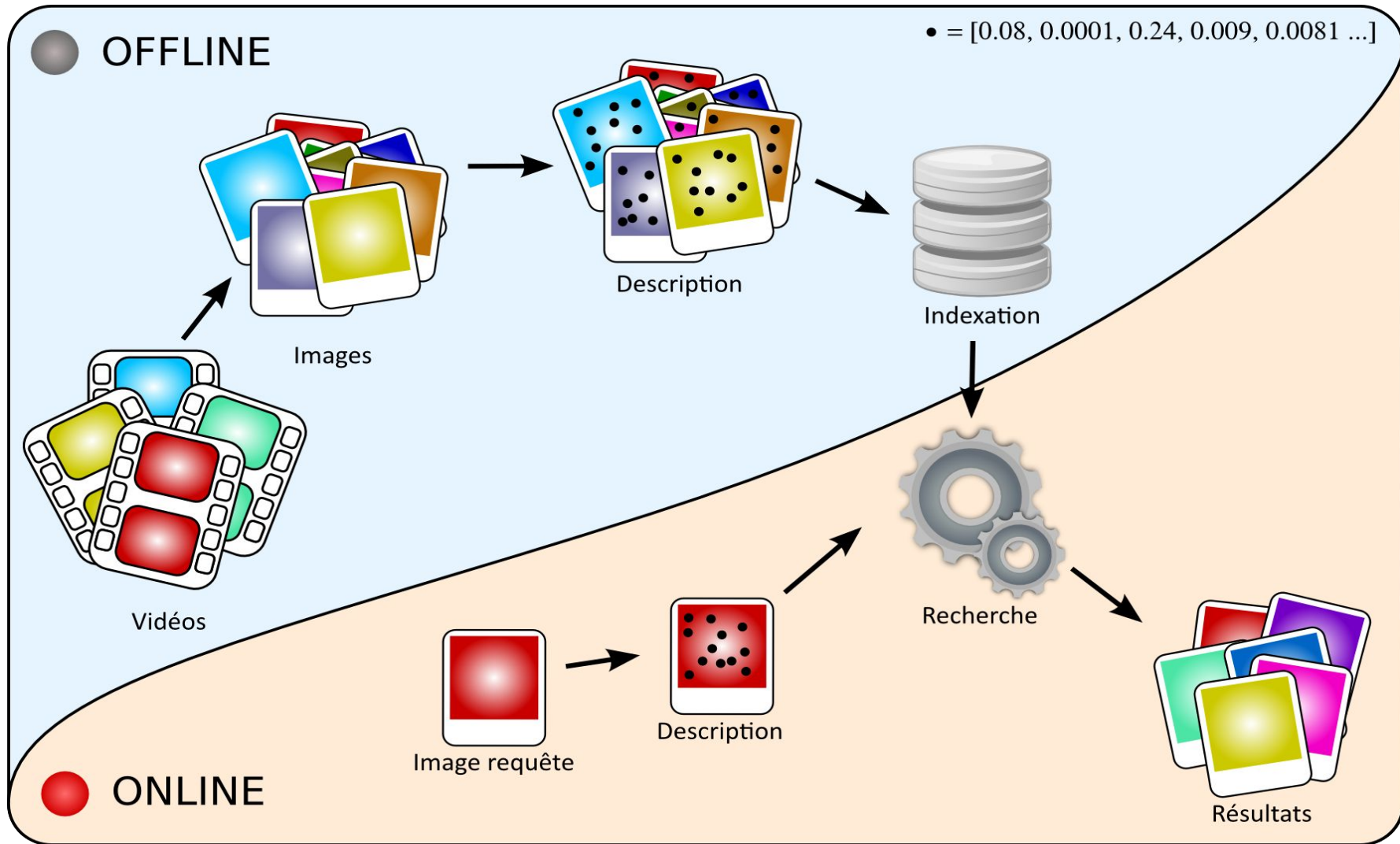
# Point de départ: Snoop

## Moteur de recherche visuelle

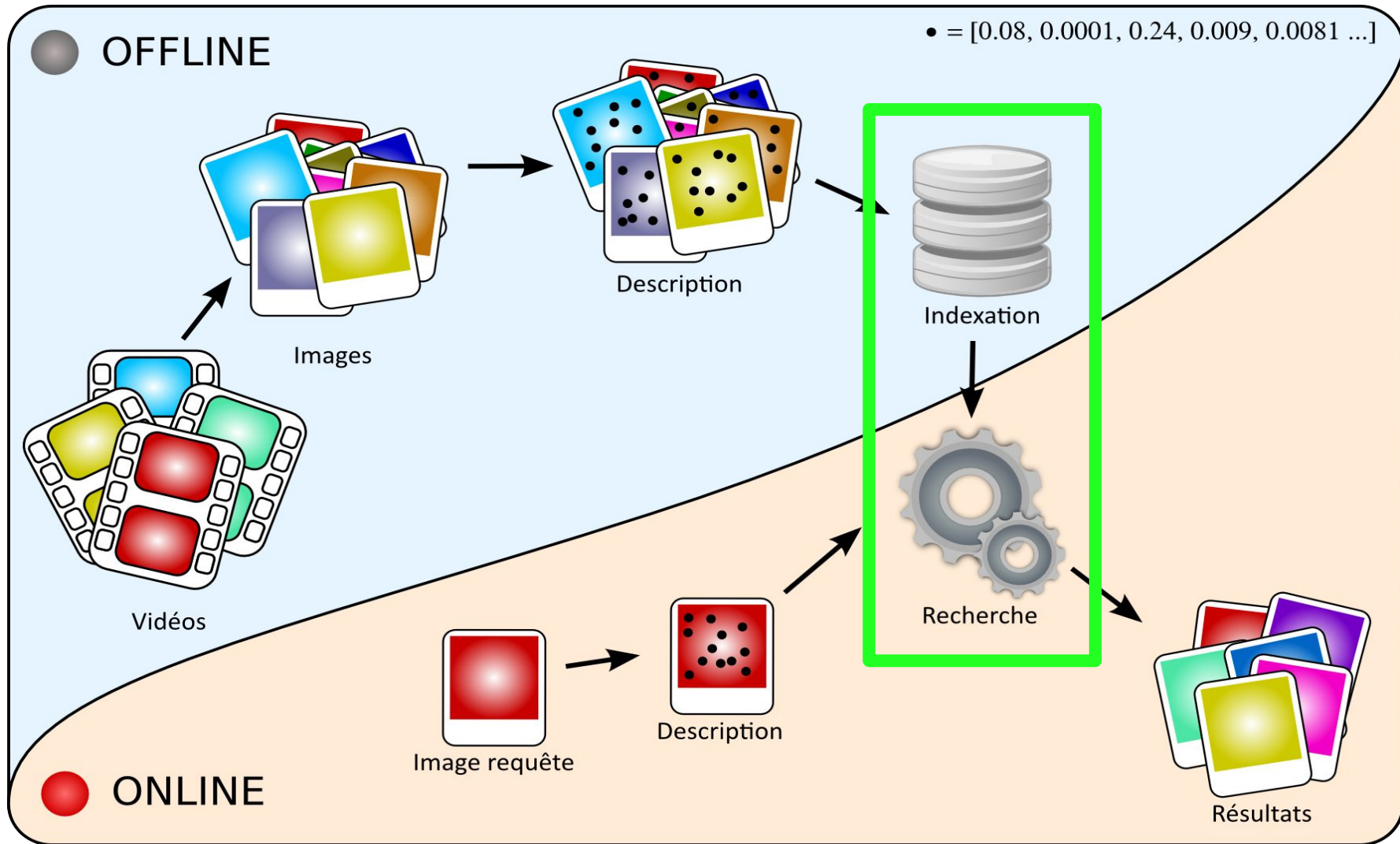
co-développé par Ina et Inria depuis 2003



# Snoop: Principe



# Snoop: Principe

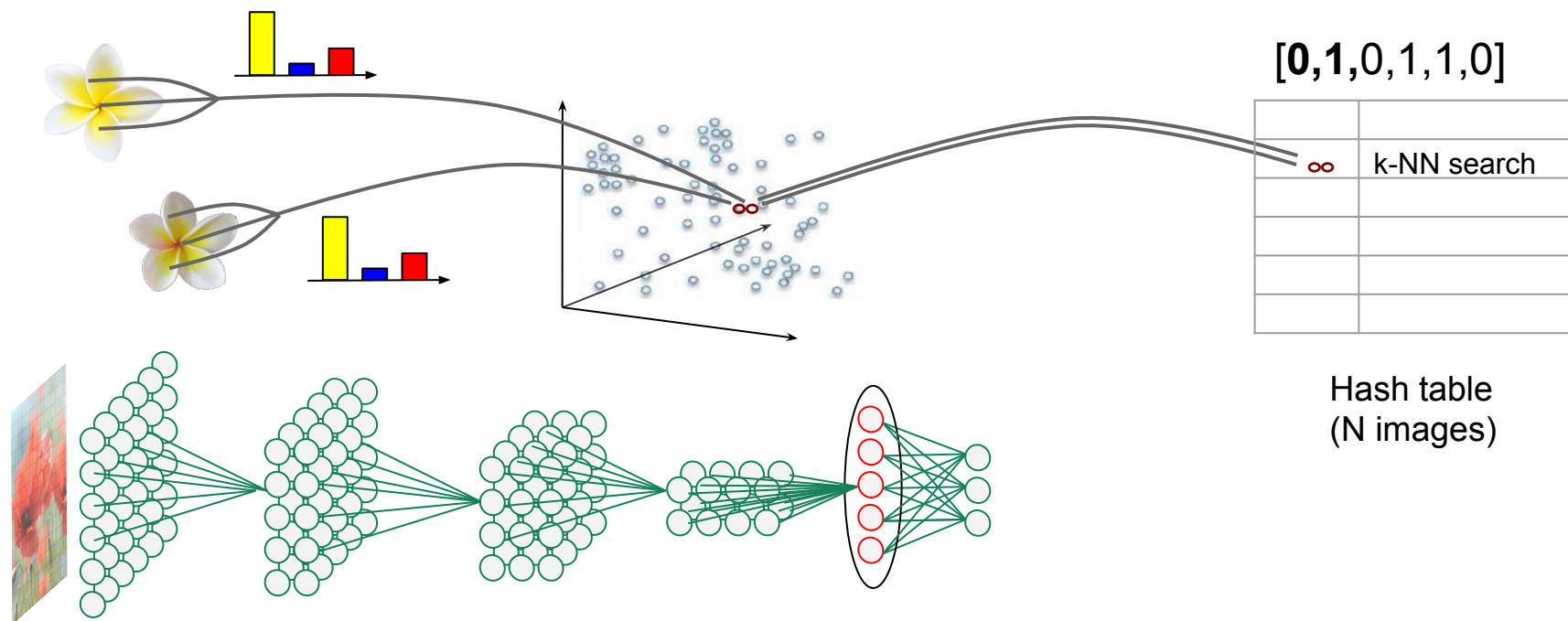


# Snoop: Comment ça marche ?

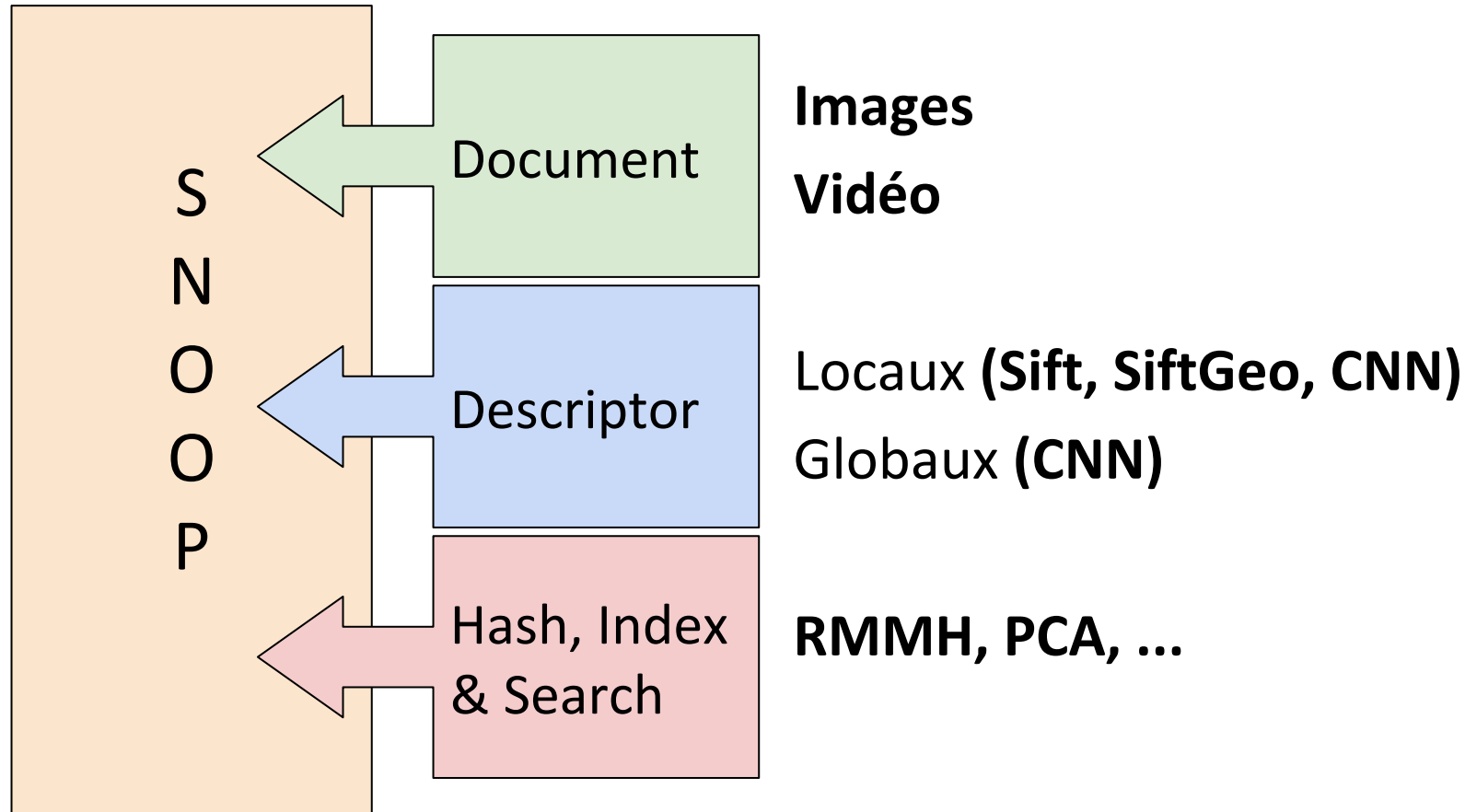
Extraction de descripteurs visuels

Représentation compacte (**hachage**) des descripteurs

Index et recherche probabiliste **générique** avec **contrôle de qualité**



# Snoop: Architecture à plugins



# Snoop: Démonstrations

- Recherche de logos



- Recherche d'oeuvre graphique



• ...




# Snoop: Découverte de contenus

**Objectif: Découvrir automatiquement** des images diffusées à travers différents média (aFP, sites web, chaînes de TV, ...).













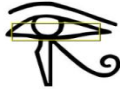





# Snoop: Hiéroglyphes

Objectif: Identifier un hiéroglyphe à partir d'un dessin à main levée

 Alpha=0.85  
Query Expansion=QE\_NONE  
nbDesc=706

file:///user/jclombar/home/Datasets/Hiero/dessins/caffe\_demos\_uploads/2015-11-23\_09:41:38.673209BLA.png

### SnoopIm Results

|  |   |   |  |  |   |  |   |
|--|---|---|--|--|---|--|---|
| <br>Result 1/16<br>Score=25.67<br>relative score=3.64%<br>x1=53<br>x2=195<br>y1=175<br>y2=223<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>    | <br>Result 2/16<br>Score=24.444<br>relative score=3.46%<br>x1=34<br>x2=133<br>y1=106<br>y2=196<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>  | <br>Result 3/16<br>Score=22.808<br>relative score=3.23%<br>x1=59<br>x2=134<br>y1=119<br>y2=195<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>    | <br>Result 4/16<br>Score=21.276<br>relative score=3.01%<br>x1=39<br>x2=197<br>y1=121<br>y2=159<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>    | <br>Result 5/16<br>Score=20.426<br>relative score=2.89%<br>x1=50<br>x2=200<br>y1=183<br>y2=223<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>   | <br>Result 6/16<br>Score=20.158<br>relative score=2.86%<br>x1=39<br>x2=149<br>y1=106<br>y2=138<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>    | <br>Result 7/16<br>Score=19.264<br>relative score=2.73%<br>x1=51<br>x2=199<br>y1=159<br>y2=198<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>   | <br>Result 8/16<br>Score=17.244<br>relative score=2.44%<br>x1=57<br>x2=198<br>y1=174<br>y2=221<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/>    |
| <br>Result 9/16<br>Score=17.204<br>relative score=2.44%<br>x1=43<br>x2=206<br>y1=116<br>y2=160<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 10/16<br>Score=15.24<br>relative score=2.16%<br>x1=53<br>x2=209<br>y1=93<br>y2=188<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 11/16<br>Score=14.154<br>relative score=2.00%<br>x1=42<br>x2=186<br>y1=104<br>y2=137<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 12/16<br>Score=14.072<br>relative score=1.99%<br>x1=37<br>x2=119<br>y1=146<br>y2=182<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 13/16<br>Score=13.59<br>relative score=1.92%<br>x1=41<br>x2=176<br>y1=157<br>y2=195<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 14/16<br>Score=13.416<br>relative score=1.90%<br>x1=38<br>x2=198<br>y1=171<br>y2=206<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 15/16<br>Score=13.006<br>relative score=1.84%<br>x1=45<br>x2=92<br>y1=181<br>y2=214<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> | <br>Result 16/16<br>Score=12.842<br>relative score=1.82%<br>x1=69<br>x2=179<br>y1=174<br>y2=193<br><input type="button" value="BBox Search"/> <input type="button" value="Full Search"/> |

# Snoop: Baleine

Objectif: Identifier des individus à partir de photos de caudales



[1191.jpg](#)



[1773.jpg](#)

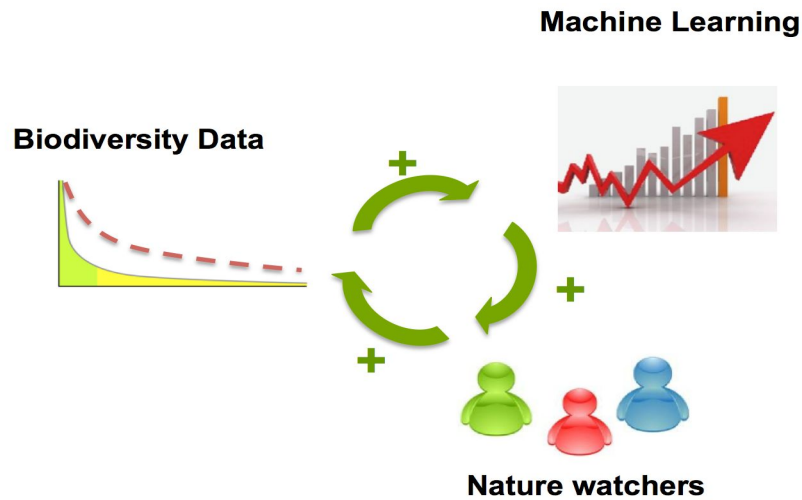


[1106.jpg](#)



[1137.jpg](#)

# **Pl@ntNet** *powered by Snoop*



- 13 million downloads
- 40-100K users per day
- 13 languages
- 200+ countries
- 25K plant species
- 115 million identification requests



- 150 million images
- 25 Tb of data
- 10 servers
- 50 API user accounts

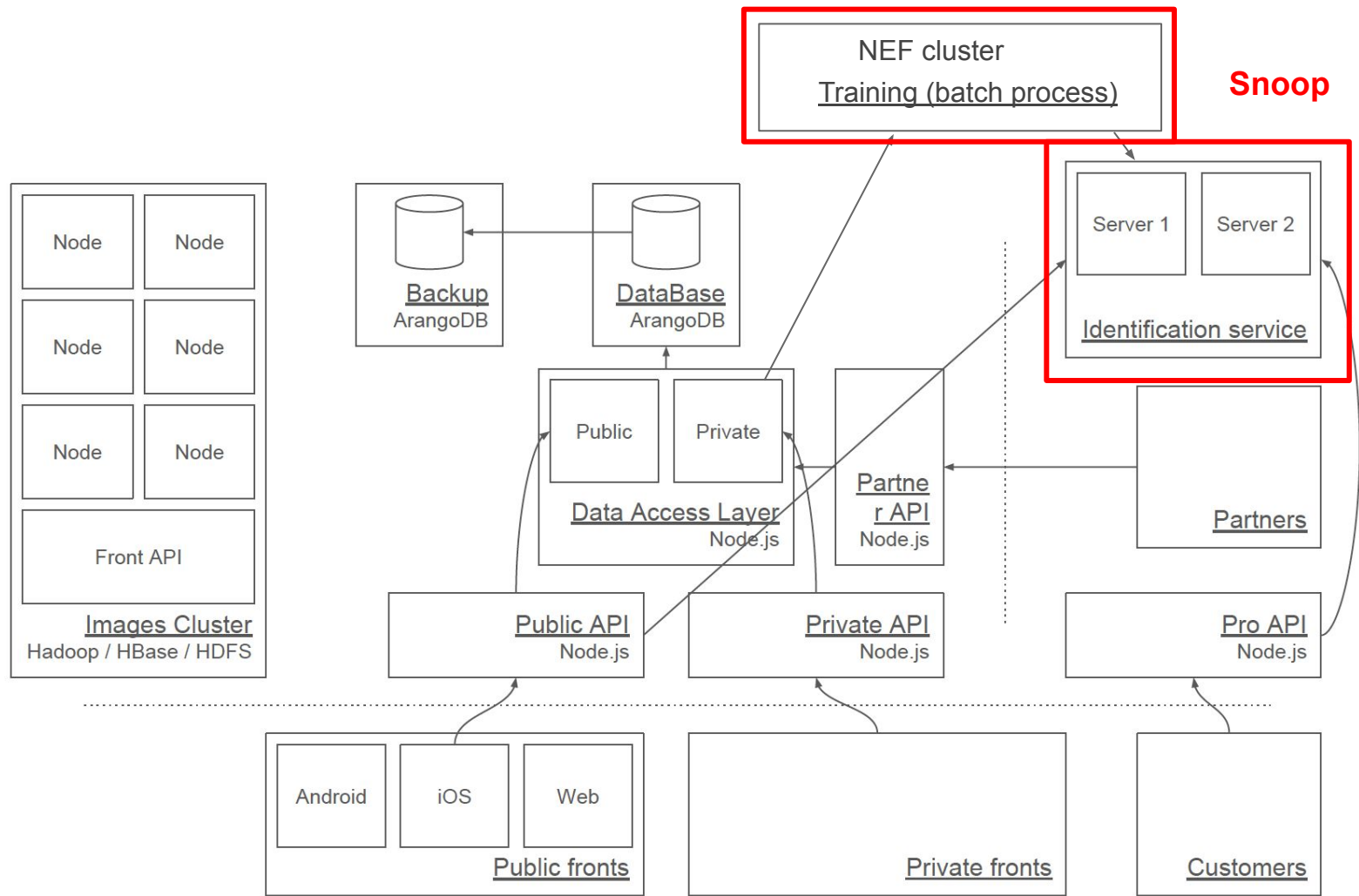


- 4 permanent researchers
- 3 engineers
- 3 PhD students
- 2 post-docs
- 4 research organisms:  
CIRAD, Inria, INRA, IRD

<https://plantnet.org>



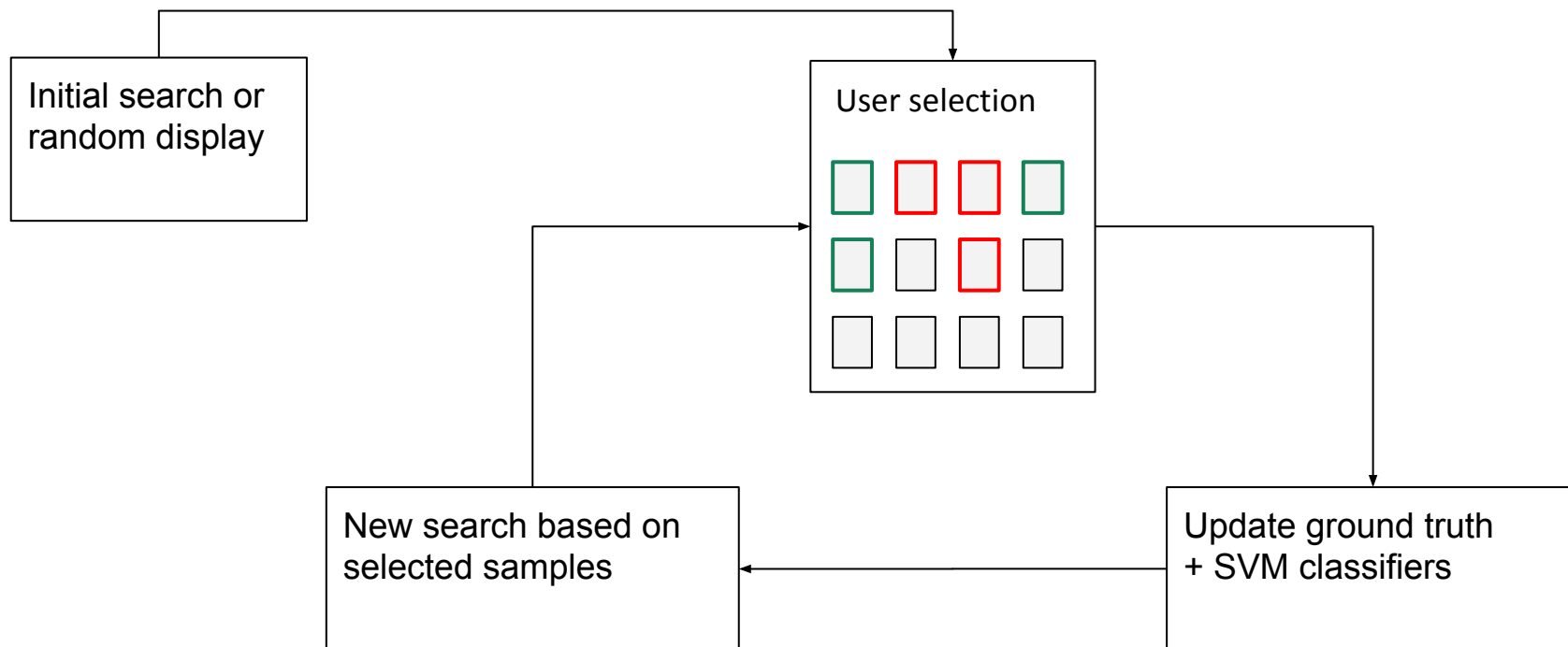
# Pl@ntNet infrastructure



# GallicaSnoop :

## Bouclage de pertinence

Permettre à l'utilisateur de définir **interactivement** des **classes visuelles**



# GallicaSnoop: Démonstrations

Extrait de la base d'images Gallica

- 1 million d'images
- 1 semaine ½ de téléchargement (web)
- 3h de calcul des descripteurs (sur un portable...)
- 10 min de calcul de l'index
- 20ms / requête externe

Récupération des méta-data (annotations existantes)

# Et ensuite ?

Dans le cadre des conventions Ministère de la Culture-Inria

## **BnF** : *GallicaSnoop*

- nouvelles méthodes de bouclage de pertinence (sélection, ...)
- description à partir de réseaux appris sur les images de Gallica:
  - à partir des vérités terrains générées par boucle de pertinence
  - intégrer les résultats du projet **CIP** dans Snoop

## **Ina** : *SnoopVideo*

Adapter la boucle de pertinence à la vidéo :

- sur plus de 100 000 heures de vidéos,
- création de bases de connaissances par des documentalistes.



# À plus long terme

- Évaluation quantitative de la recherche interactive
- Installation de GallicaSnoop en salle X à la BnF
- Intégration de Snoop dans les workflows de la BnF et de l'Ina

**Et bien sûr, ouvrir ces technologies  
aux autres musées, instituts, ...**

# Merci !

[alexis.joly@inria.fr](mailto:alexis.joly@inria.fr)

[jean-christophe.lombardo@inria.fr](mailto:jean-christophe.lombardo@inria.fr)

[obuission@ina.fr](mailto:obuission@ina.fr)

