Projet CNRS-Mastodons
ANalyse d'IMages fondée sur des Informations TEXtuelles

Bruno Crémilleux, **Pierre Gançarski**, Mathieu Roche,
Christian Sallaberry, Maguelonne Teisseire *et al.*

Strasbourg – novembre 2014

Plan

Vision du "Big Data"

(DNA, le 5 novembre 2014)

Plusieurs questions se posent dans le cadre d'un **suivi des dynamiques territoriales** :

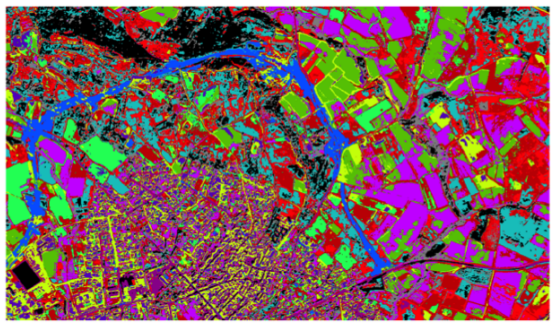
- Comment identifier les *trajectoires de transformations*, et comment les documenter au mieux ?
- Comment procéder à des analyses rétrospectives de ces trajectoires ?
- Comment fournir aux décideurs une *mémoire territoriale* synthétique et lisible ?

Exemple 1 : Quand les **images** mettent en relief la **concrétisation d'un projet d'aménagement du territoire**...

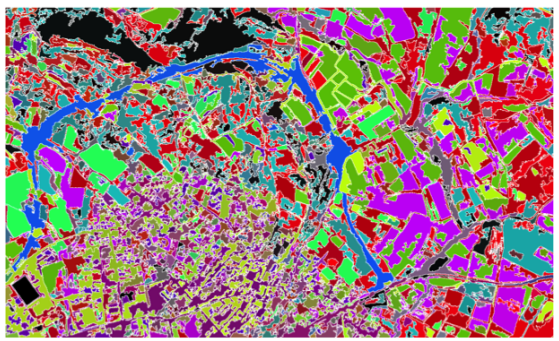
10 july 2012 14 september 2012 15 march 2013

Exemple 1 : Quand les **images** mettent en relief la **concrétisation d'un projet d'aménagement du territoire**...

a.



b.



c.

Exemple 2 : Quand les **textes** mettent en relief l'**échec d'un projet d'aménagement du territoire**...

Midi Libre AGDE ALES AVIGNON BAGNOLS BEZIERS CARCASSONNE LODEVE LUNEL MENDE MILLAU MONTPE

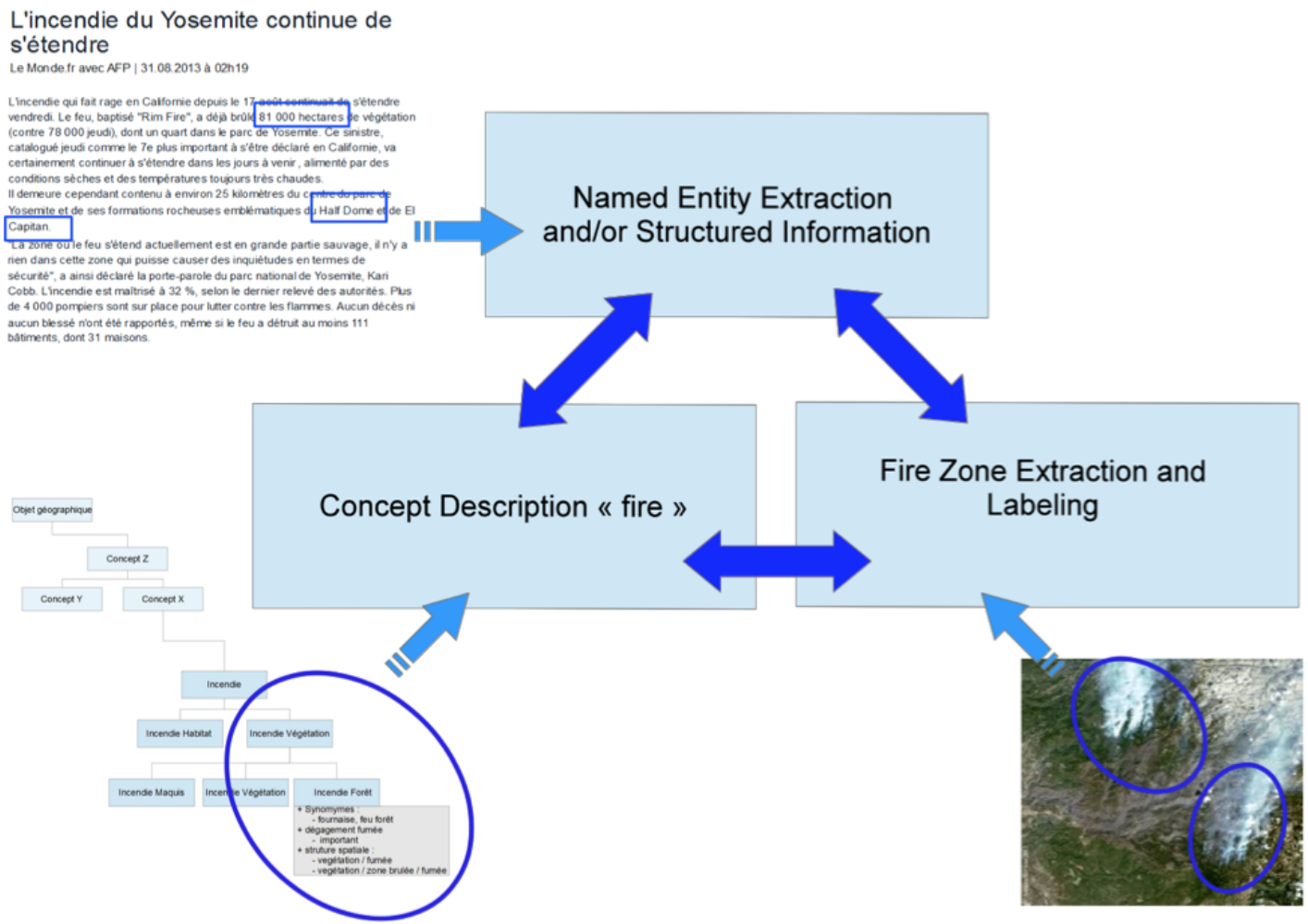
L'hinterland du port de Sète à Poussan a-t-il du plomb dans l'aile ?

il y a 421 jours 12 Pa.C.

Recommander Partager 27 personnes recommandent ça. Soyez le premier parmi vos amis. TWITTER g+1 0

La commission d'enquête a émis un avis favorable au Schéma de cohérence territoriale du bassin de Thau. Mais avec quelques

Idée : Utiliser toutes les informations disponibles quel qu'en soit le média



Comment intégrer la spécificité des informations portées par ces données massives et hétérogènes.

1. Occupation des sols : généralement visible sur les images de télédétection
2. Nommage des objets géographiques : Donné par des documents textuels
3. Usage des sols : rarement visible sur les images, mais peut être précisé par des documents textuels

→ **Comment nommer** un objet géographique dans une image ou **situer** une entité géographique
→ **Comment relier** une vision structurelle à une vision fonctionnelle

Comment intégrer la spécificité **temporelle** des informations portées par ces données massives et hétérogènes.

1. Etude préalable : majoritairement des documents textuels (enquêtes publiques, comptes rendus ...
2. Construction : généralement visible sur les images de télédétection ;
3. Mise en service : visible sur les images, mais également via des articles de journaux, de blogs, reportages télévisuels

→ **Comment relier** une vision continue (séquence d'images) à une vision discrète (événement)

Exploiter des **données textuelles** massives et hétérogènes permettant de compléter l'**analyse des images satellites**

Deux axes principaux :

- Extraire les informations des textes (entités nommées), leurs thèmes et leurs relations spatiales (positions relatives)
- Relier ces informations à des informations (objet) extraits des images ou liées à celles-ci (concepts géographiques)

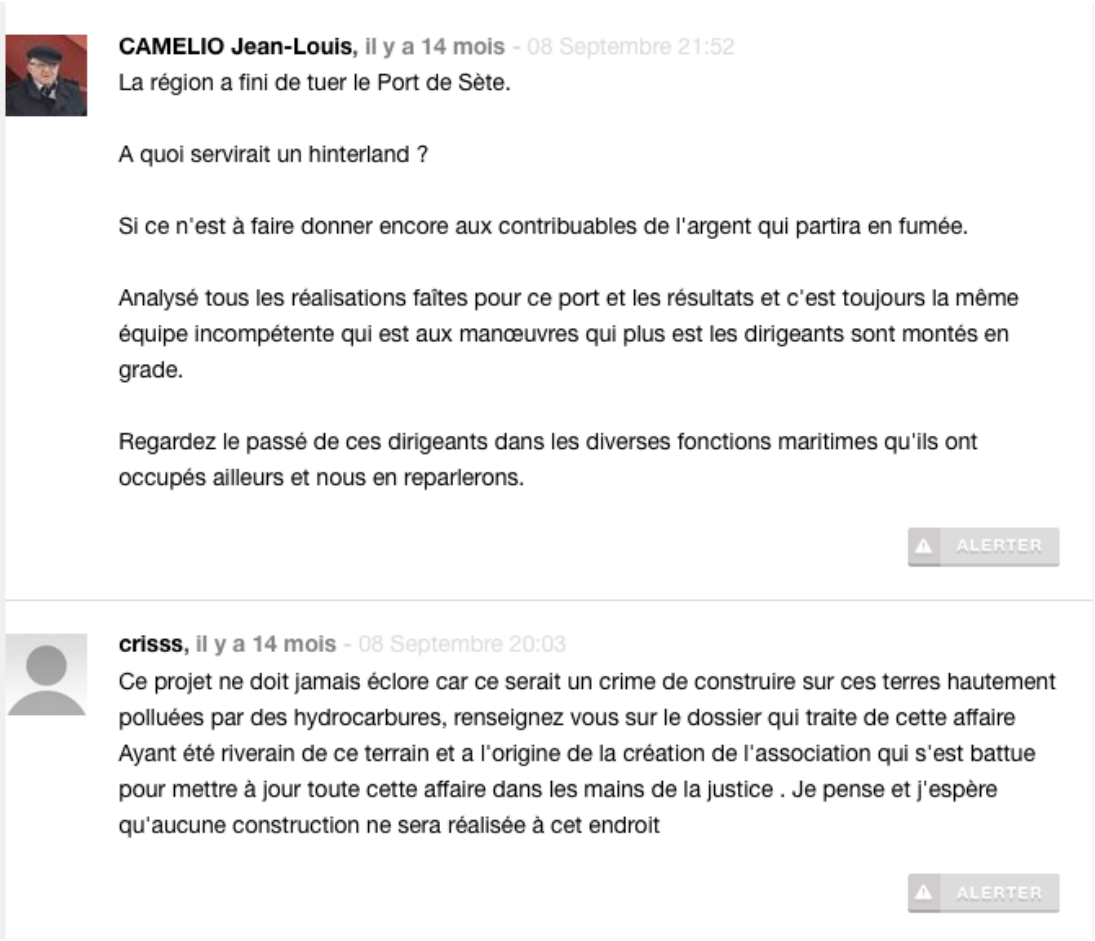
1. Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier – **LIRMM (Montpellier)**
   Domaine : Traitement Automatique du Langage Naturel et Fouille de Données
2. Territoires, Environnement, Télédetection et Information Spatiale – **TETIS (Montpellier)**
   Domaine : Informations Géospatiales et Fouille de Données
3. Laboratoire des sciences de l'ingénieur, de l'informatique et de l'imagerie – **ICube (Strasbourg)**
   Domaine : Analyse d'Image et Traitement Automatique du Langage Naturel
4. Groupe de Recherche en Informatique, Image, Automatique et Instrumentation de Caen – **GREYC (Caen)**
   Domaine : Fouille de Données et Traitement Automatique du Langage Naturel
5. Laboratoire d'Informatique de L'Université de Pau et des Pays de l'Adour – **LIUPPA (Pau)**
   Domaine : Recherche d'Information Géographique et Traitement Automatique du Langage Naturel

15 membres permanents participent activement au projet.

→ **Échanges de savoir-faire** sur des méthodologies, état de l'art, prototypes logiciels et ressources.
→ **Définitions collectives** des besoins, scénarios, entrées/sorties des modules et chaîne globale de traitement.

3 post doctorants (Labex NUMEV, Equipex GEOSUD, ATER) ont été intégrés au projet dès son démarrage.

Traitement des données textuelles : Détection automatique de **sentiments** dans les textes [Kergosien *et al.* 2014]



CAMELIO Jean-Louis, il y a 14 mois - 08 Septembre 21:52
La région a fini de tuer le Port de Sète.

A quoi servirait un hinterland ?

Si ce n'est à faire donner encore aux contribuables de l'argent qui partira en fumée.

Analysé tous les réalisations faîtes pour ce port et les résultats et c'est toujours la même équipe incompétente qui est aux manœuvres qui plus est les dirigeants sont montés en grade.

Regardez le passé de ces dirigeants dans les diverses fonctions maritimes qu'ils ont occupés ailleurs et nous en reparlerons.

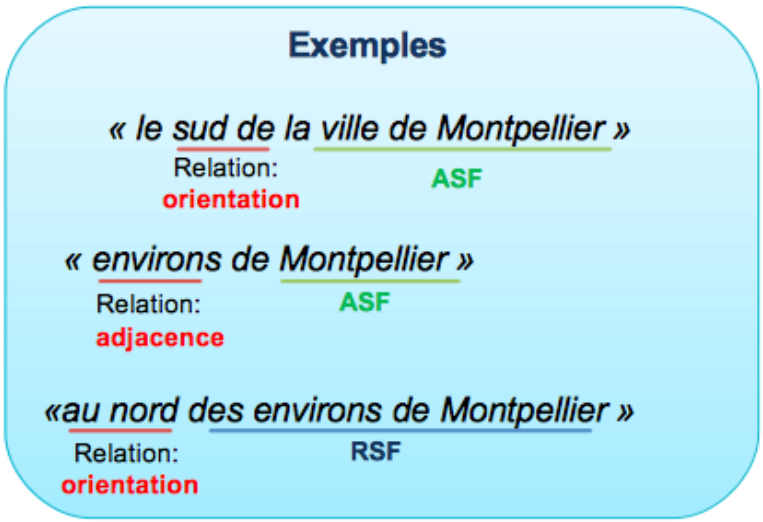
ALERTER

crisss, il y a 14 mois - 08 Septembre 20:03
Ce projet ne doit jamais éclore car ce serait un crime de construire sur ces terres hautement polluées par des hydrocarbures, renseignez vous sur le dossier qui traite de cette affaire Ayant été riverain de ce terrain et a l'origine de la création de l'association qui s'est battue pour mettre à jour toute cette affaire dans les mains de la justice . Je pense et j'espère qu'aucune construction ne sera réalisée à cet endroit

ALERTER

Traitement des données textuelles : Identification automatique **d'Entités Spatiales (ES)** dans les textes [Gaio *et al.*, 2012]

- utilisation de **patrons d'extraction** pour identifier
  - des indicateurs spatiaux (orientation, distance, adjacence, inclusion, figure géométrique)
  - des Entités Spatiales Absolues et Relatives



→ F-mesure autour de 67% (ESA) / 74% (ESR)

- *précision élevée pour les ESR,*
- *limite des outils de TAL pour le traitement de ces données complexes.*

Traitement des données textuelles

Identification automatique **d'Entités Spatiales (ES)** dans les textes

- utilisation de **méthodes statistiques** pour
  - traiter les masses de données
  - enrichir les méthodes symboliques

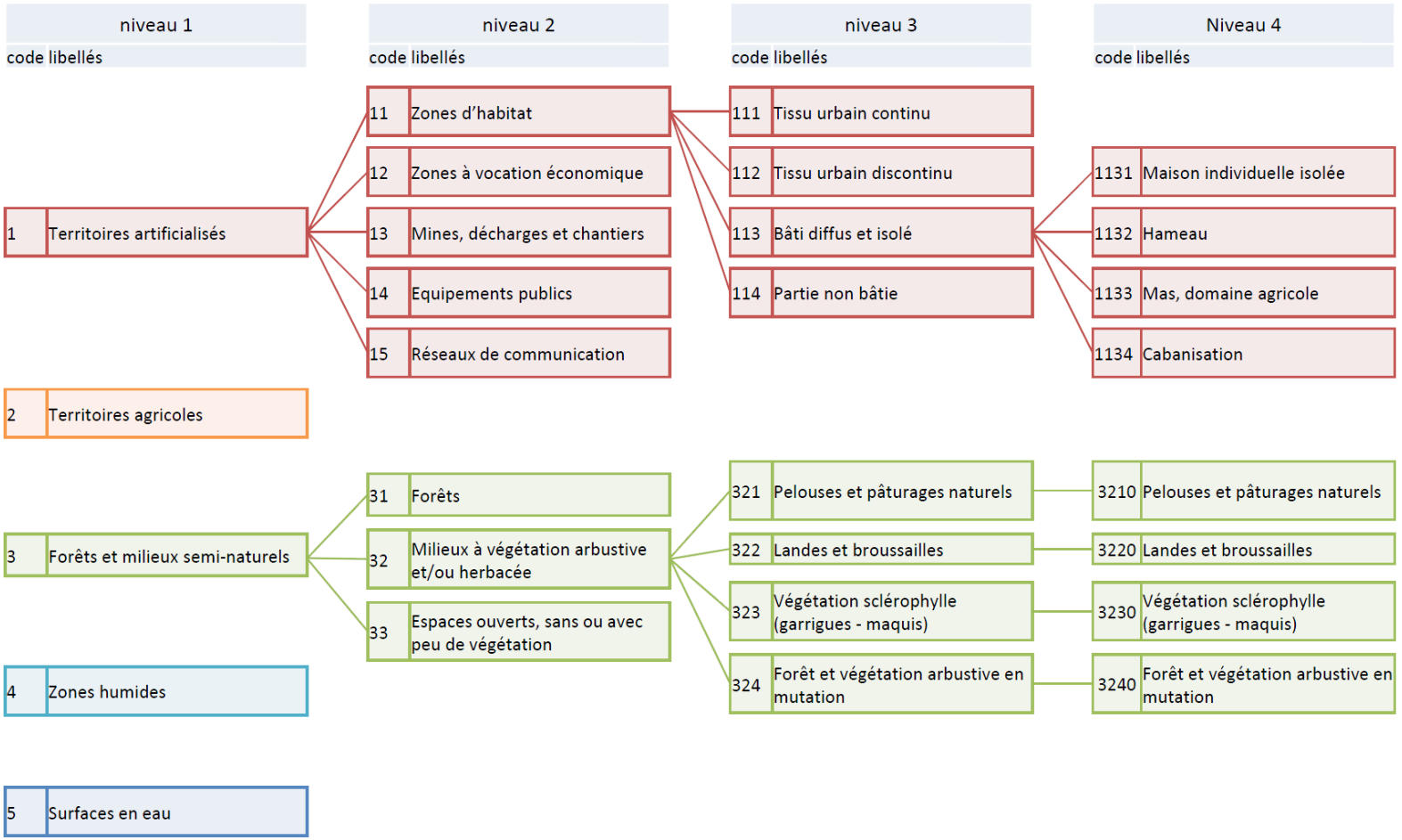
Principaux résultats

→ **Identification d'ES et coordonnées associées**
→ **Identification de thèmes :**

- par méthodes statistiques : **topic models** (*Montpellier et Strasbourg*)
- par méthodes sémantiques à large spectre : **Babelnet** (*Strasbourg*) et **Agrovoc** (*Montpellier*)
- par méthodes sémantiques spécialisées : **Geonto** (*Pau*)

Traitement des données images

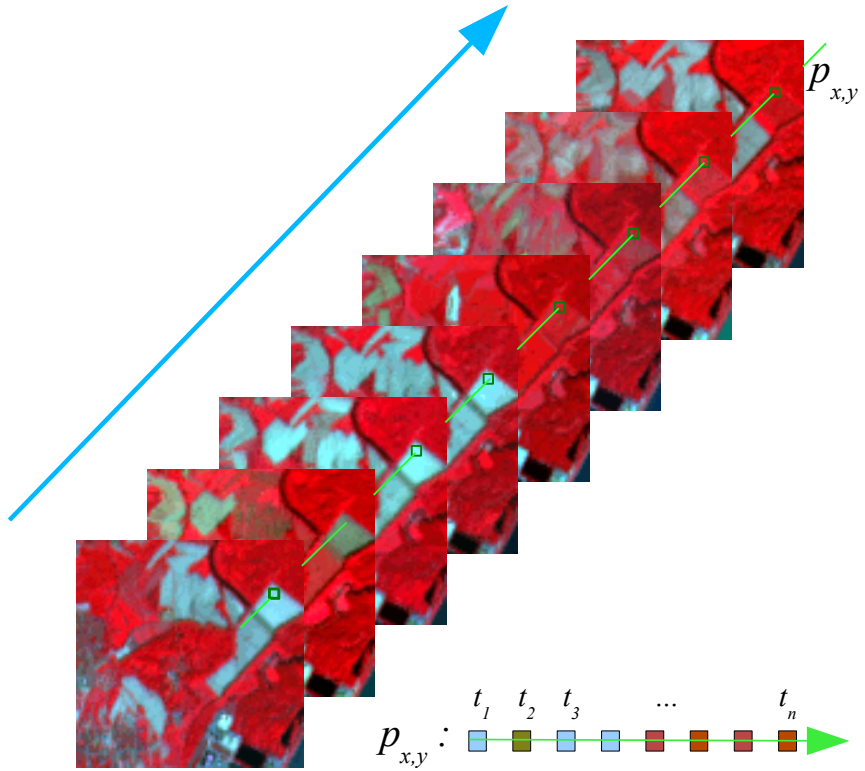
Définition d'une "ontologie" du domaine



→ **Comment la "projeter"** dans l'image

Extraction des objets géographiques et de leur évolution

- Soit une série d'images ordonnée suivant le temps

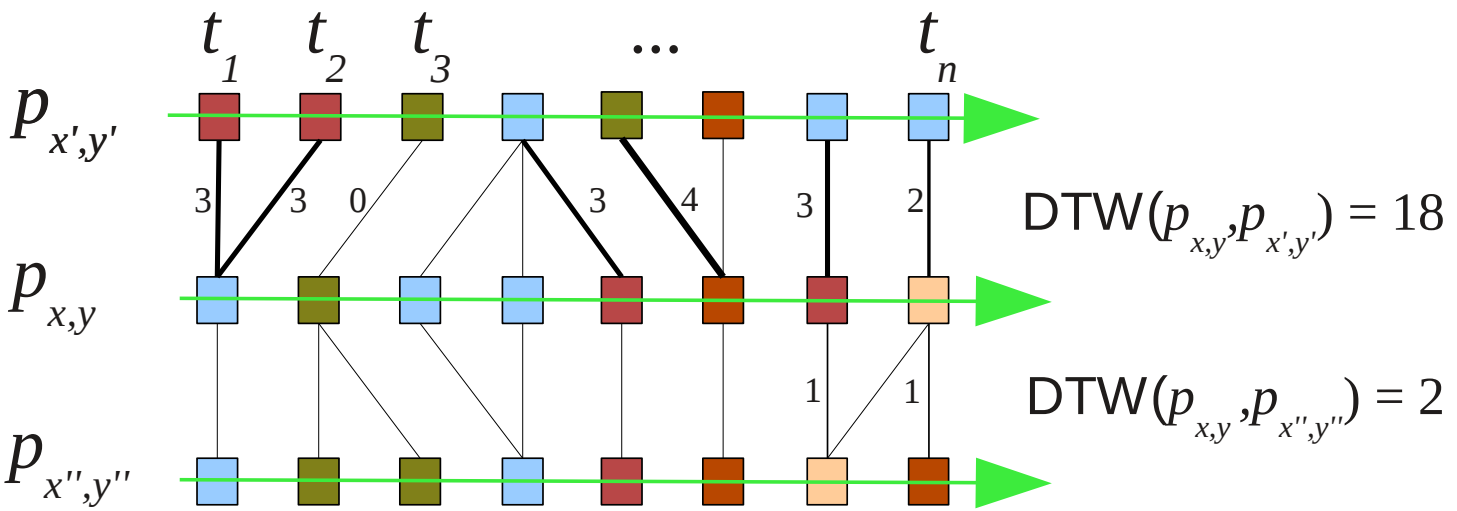


- On construit une séquence par pixel

Utilisation de **DTW**

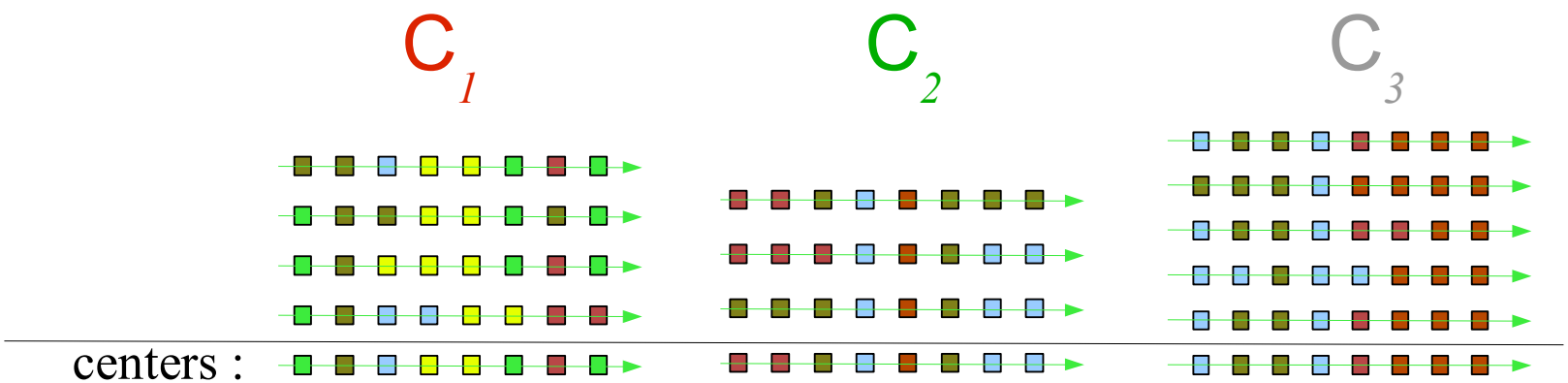
- On suppose disposer d'une mesure de similarité entre deux états de deux séquences différentes indépendante du temps

→ Utilisation de DTW (Dynamic time wrapping) possible

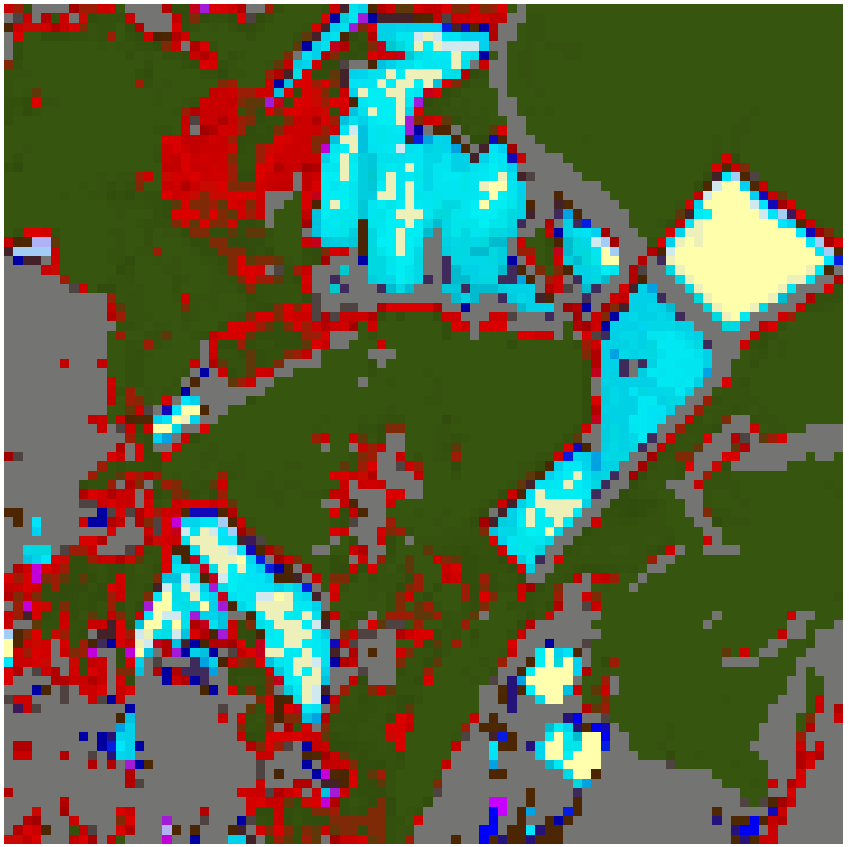


Application d'un algorithme de clustering (exemple Kmeans) :

- On donne une mesure de similarité entre deux états : différence de radiométrie, similarité entre concepts . . .
- Utilisation de DTW couplée à la moyenne **DBA**



Classification de séquences → Une carte des clusters d'évolution des pixels



cluster #1

cluster #2

cluster #3

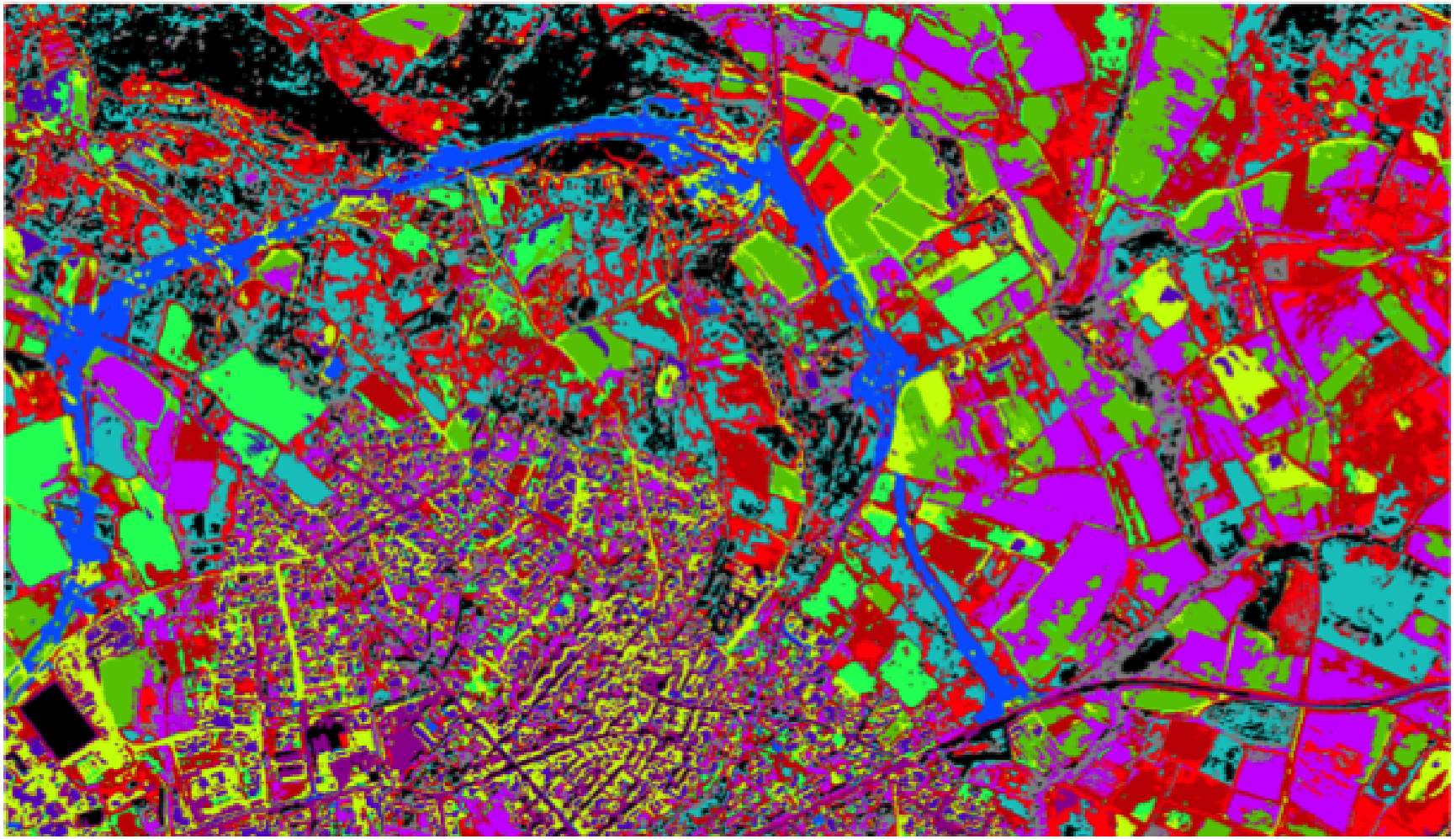
cluster #4

cluster #5

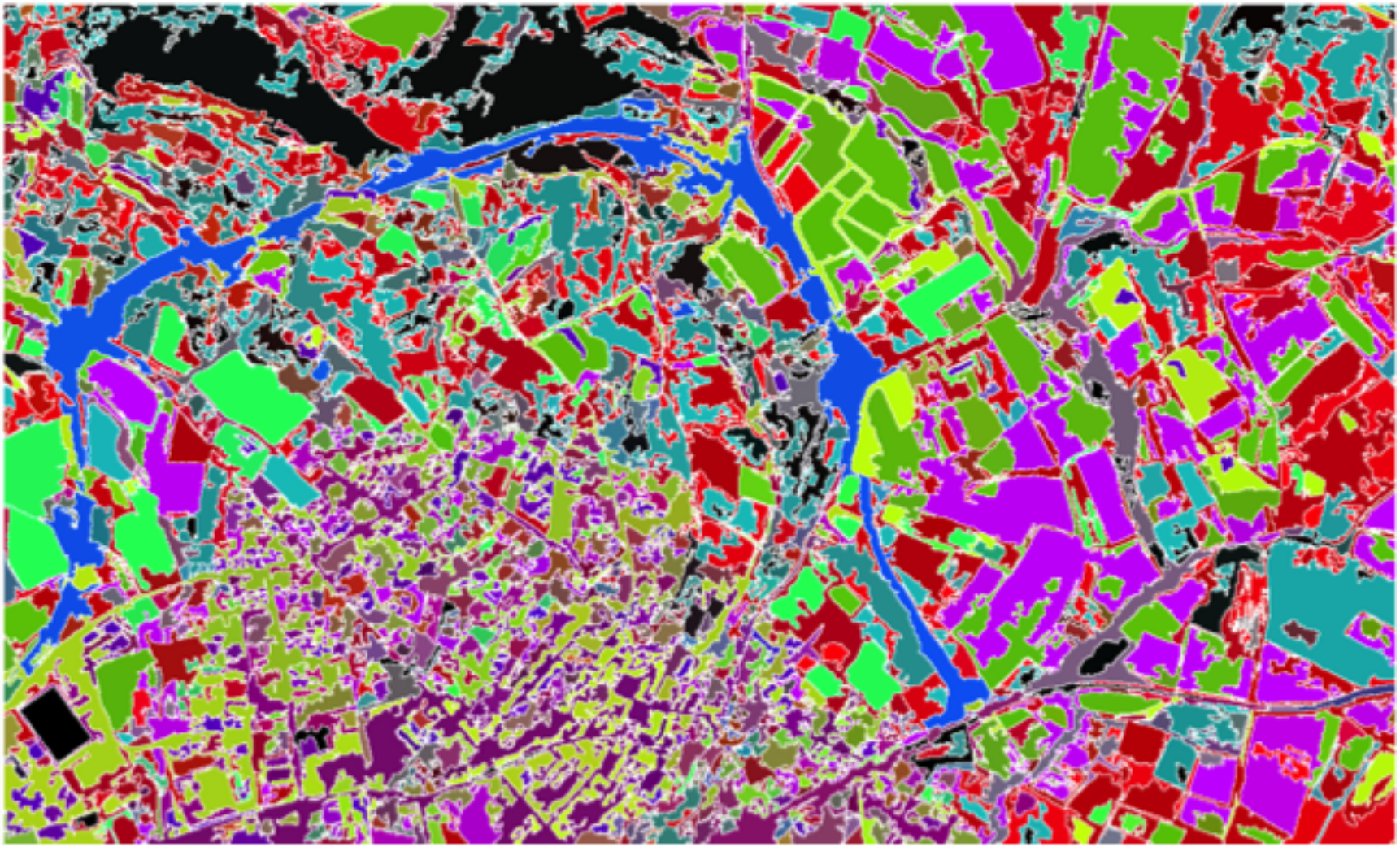
cluster #6

Classification de séquences : Exemple

Etape 1 Classification de séquences



Etape 2 Segmentation et classification de la carte de séquences



Etape 3 Extraction des clusters d'intérêt

Comment associer ces objets à la rocade de Villeveyrac ?

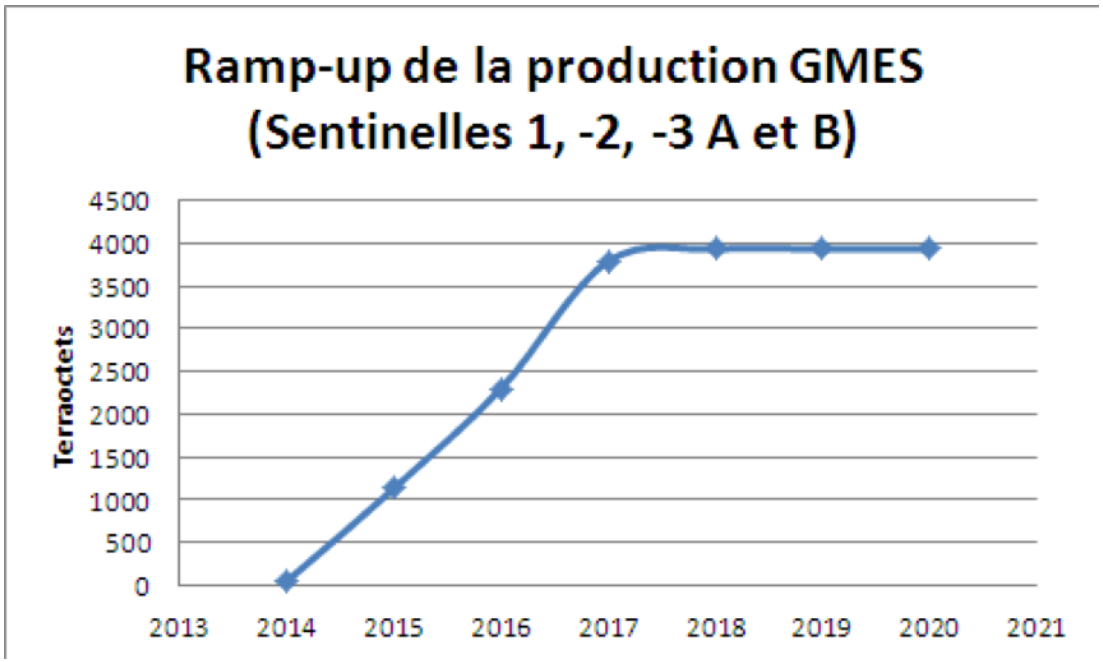
→ **Bilan 2014 :**

- Corpus et ressources acquises,
- Chaîne de traitement définie,
- Premiers modules de la chaîne expérimentés et verrous identifiés.

→ **Perspectives pour 2015 :**

- Prendre en compte les aspects temporels (images et textes),
- Identification fines des thématiques abordées dans les textes et faire le lien avec les images,

Et à plus long terme : ANIMITEX utilisation du flux **Sentinel-2**
$\rightarrow$ une image tous les 5 jours d'ici trois ans !
(*aujourd'hui une dizaine d'images par an (satellites SPOT, Landsat, etc.)*)



Sentinel : #1 - 1,6 To/day, #2 - 1,6 To/day, #3 - 2,2 To/day

Des projets en cours . . .

- Dépôt d'un **projet ANR**
- Organisation de **workshops et sessions spéciales**
  - Sessions spéciales à **DSAA'2014** (Shanghai) et **DSAA'2015** (Paris)
  - Ateliers à **EGC'2013** (Rennes) et **EGC'2014** (Luxembourg)
- Co-organisation d'une **école thématique** → **Focolise**

Publications

- Gaio M., Nguyen V.T., Sallaberry C. Typage de noms toponymiques a des fins d ?indexation geographique, Revue Traitement Automatique des Langues, Vol. 53, n° 2, p. 143-176, 2012
- Petitjean F., Inglada J., Gançarski P. Satellite Image Time Series Analysis under Time Warping IEEE Transactions on Geoscience and Remote Sensing, pages 3081–3095, Vol. 50, No 8, 2012
- Kergosien E., Maurel P., Roche M., Teisseire M. SENTERRITOIRE pour la detection d'opinions liees a l'amenagement d'un territoire. Revue Internationale de Géomatique, version étendue de SAGEO'13 (Spatial Analysis and GEOmatics), 2014
- Sallaberry C., Geographical Information Retrieval in Textual Corpora. *FOCUS Series in GIS*, 2013

http ://www.lirmm.fr/∼mroche/ANIMITEX/

Mail aux membres du projet :
mastodons_animitex@lirmm.fr