

## Fouille de données spatio-temporelles

Sandra Bringay et Maguelonne TEISSEIRE

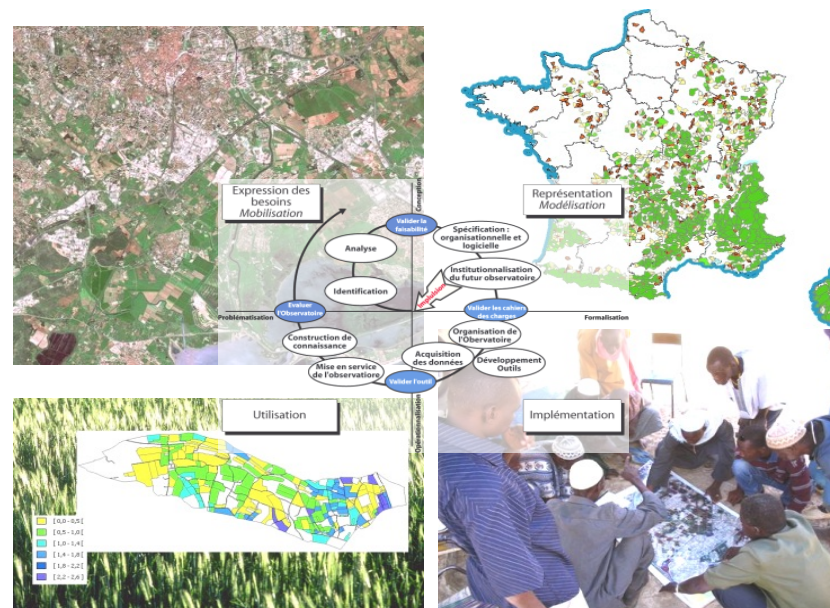
# + Plan

1. TETIS, SISO et la fouille de données
2. Les données spatio-temporelles
3. Deux méthodes de fouille de données spatio-temporelles
4. Conclusion et perspectives



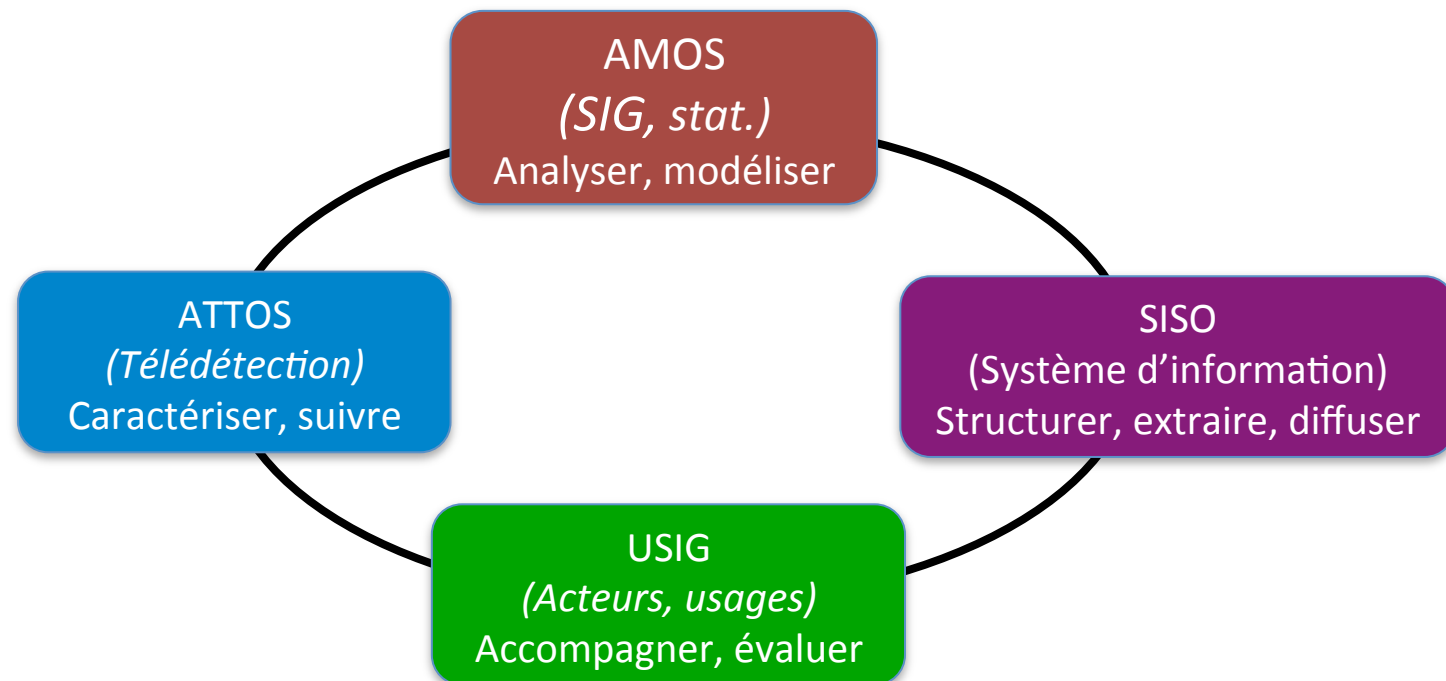
# UMR TETIS

## Territoire, Environnement, Télédétection et Information spatiale



# + Maîtrise de la « chaîne de l'information spatiale »

« Une information utile, utilisable, utilisée »



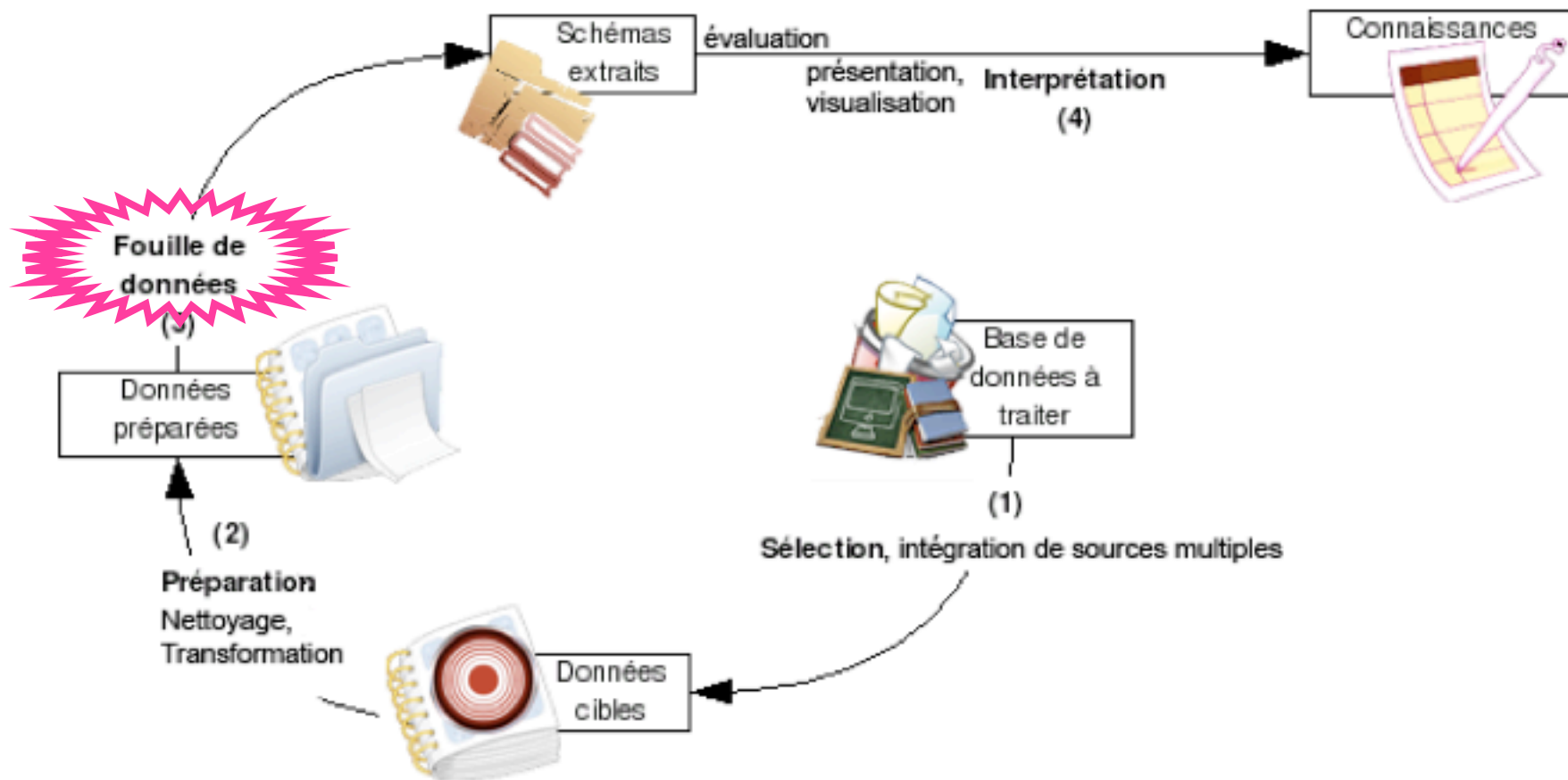
## Champs d'application

Agriculture, biodiversité, environnement, forêt, ressources en eau, risques, territoires

**Echelles intermédiaires** de la gestion : local à régional, instantané à pluri-décennal

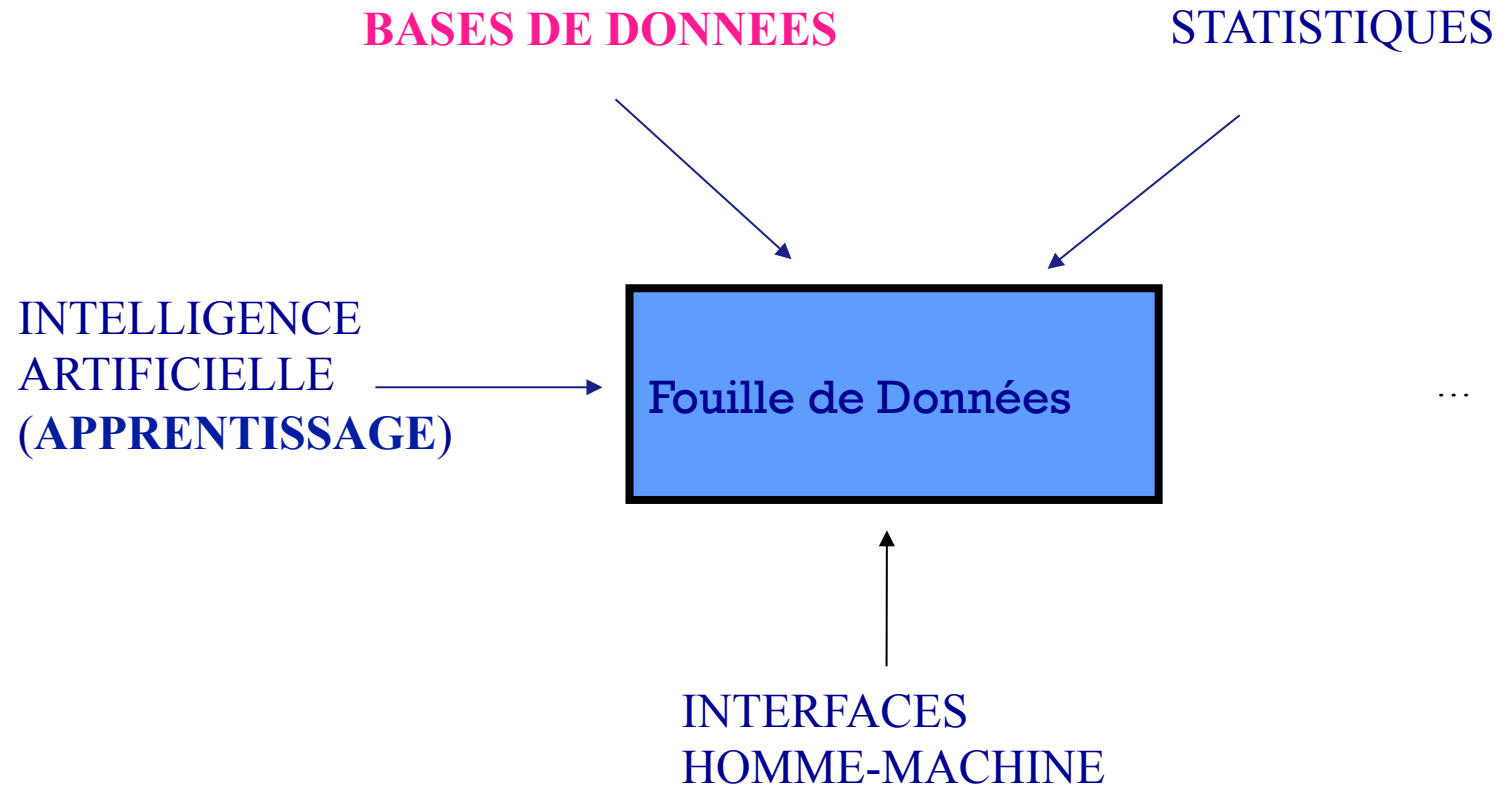


# Le Processus d'Extraction de Connaissances



Fouille de données spatio-temporelles : des données aux motifs

# + Fouille de données ?



# + Les tâches de la fouille de données

## ■ Fouille de données : de nombreuses tâches possibles (27)

### ■ **Classification**

Créer une fonction qui classe une donnée élémentaire parmi plusieurs classes prédéfinies existantes

### ■ **Régression**

Créer une fonction qui donne une donnée élémentaire à une variable de prévision avec des données réelles

### ■ **Groupement (clustering)**

Rechercher à identifier un ensemble fini de catégories ou groupe en vue de décrire les données

### ■ **Résumé**

Affiner une description compacte d'un sous-ensemble de données

### ■ **Modélisation des dépendances**

Trouver un modèle qui décrit des dépendances significatives entre les variables

### ■ **Détection de changement et déviation**

Découvrir les changements les plus significatifs dans les données

# + Quels algorithmes de fouille de données ?

- Non pas 1 mais  $n$  approches ... avec  $m$  techniques ...
- 3 approches principales (*R. Agrawal*) **vision Base de Données**

*Classification*

*Règles d'association*

*Motifs séquentiels*



# + Règles d'Association vs Motifs Séquentiels

## ■ **Corrélation entre les produits (RA)**

*Les personnes qui achètent des couches achètent de la bière*

## ■ **Comportement des clients au cours du temps (MS)**

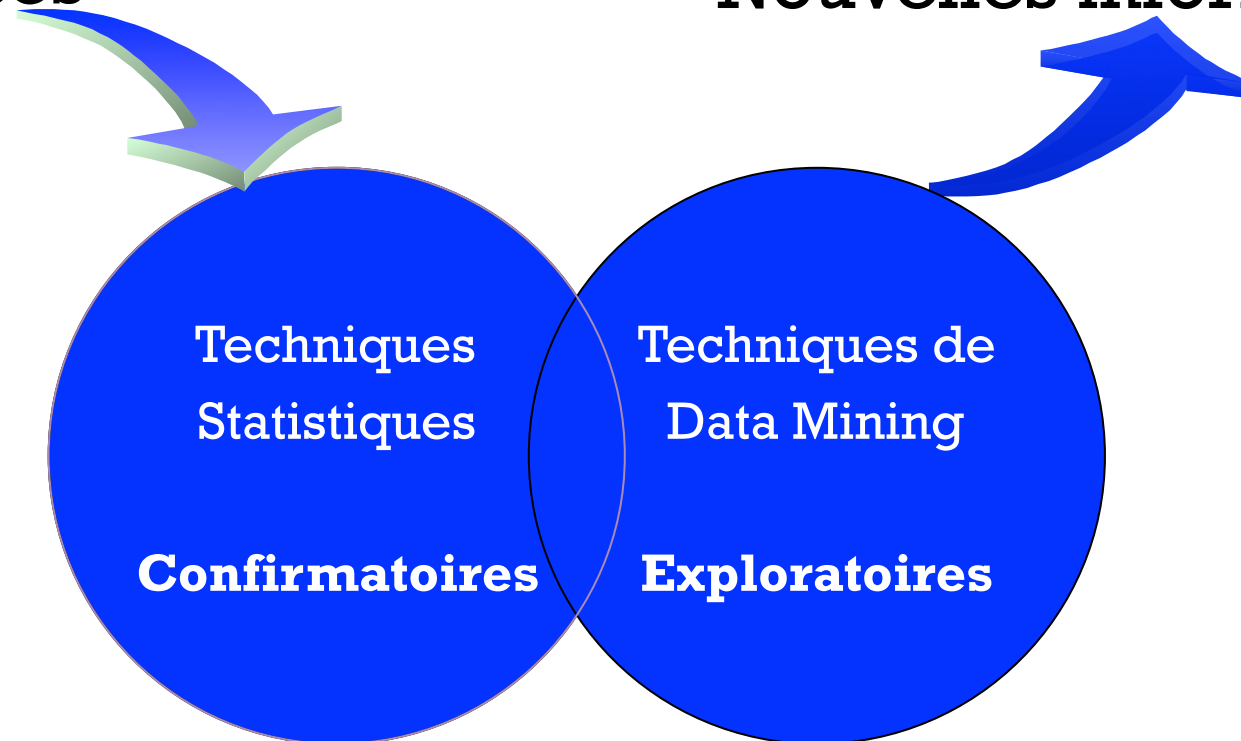
*Les personnes qui achètent des couches achètent **trois jours après** de la bière*

**Motifs fréquents, motifs rares,  
Tendances, anomalies**

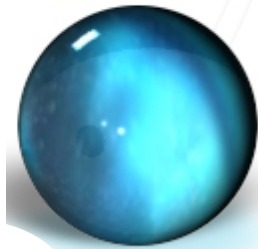
# + Fouille de données vs Statistiques

Hypothèses

Nouvelles informations



# + Machine Learning vs Data Mining



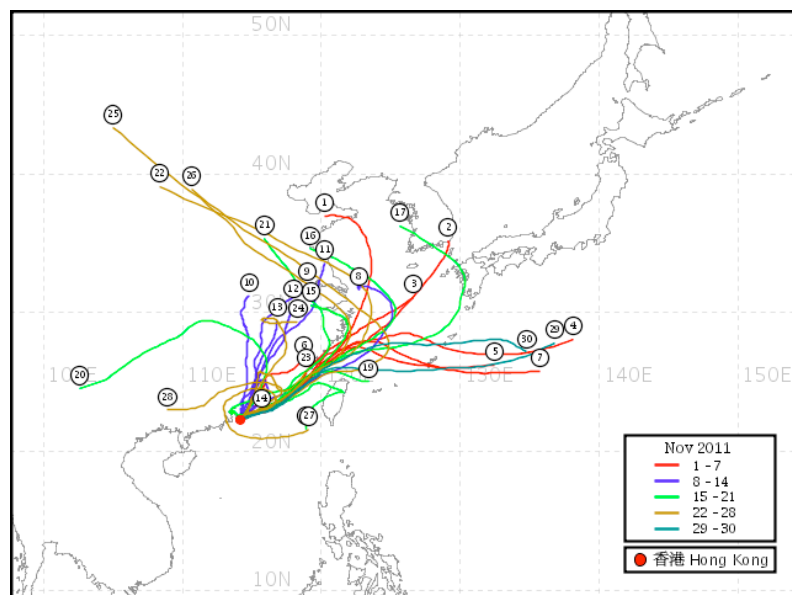
## Passage à l'échelle



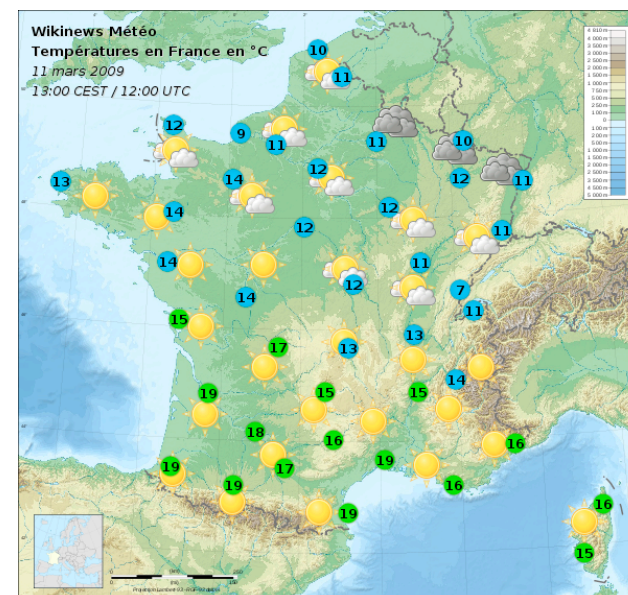
+ les données spatio-temporelles  
Deux perspectives

# + Deux perspectives

## ■ Les objets mobiles



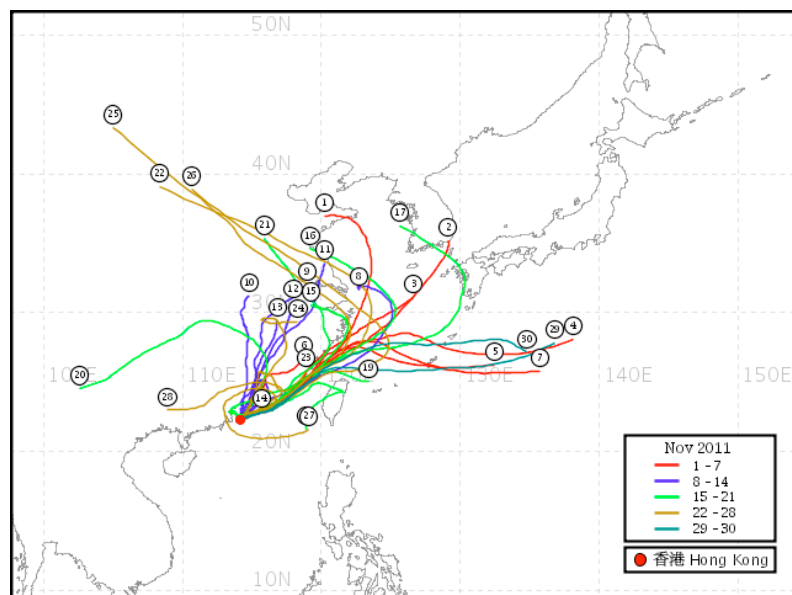
## ■ Les événements dans des zones fixées



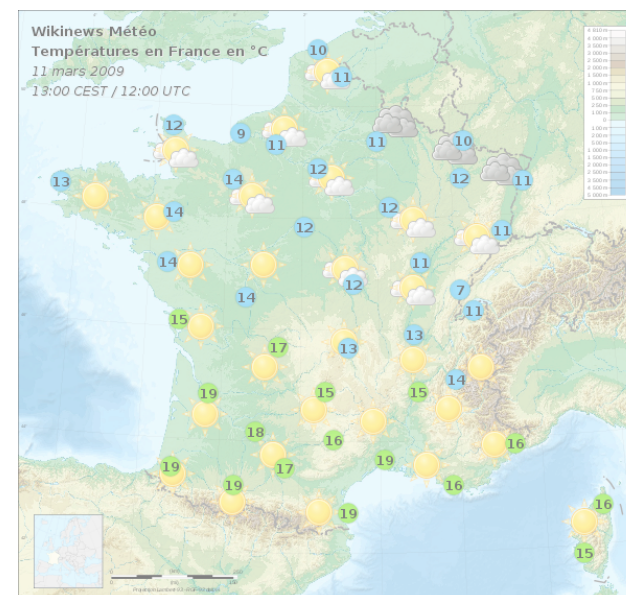
Fouille de données spatio-temporelles : des données aux motifs

# + Deux perspectives

## ■ Les objets mobiles



## ■ Les événements dans des zones fixées

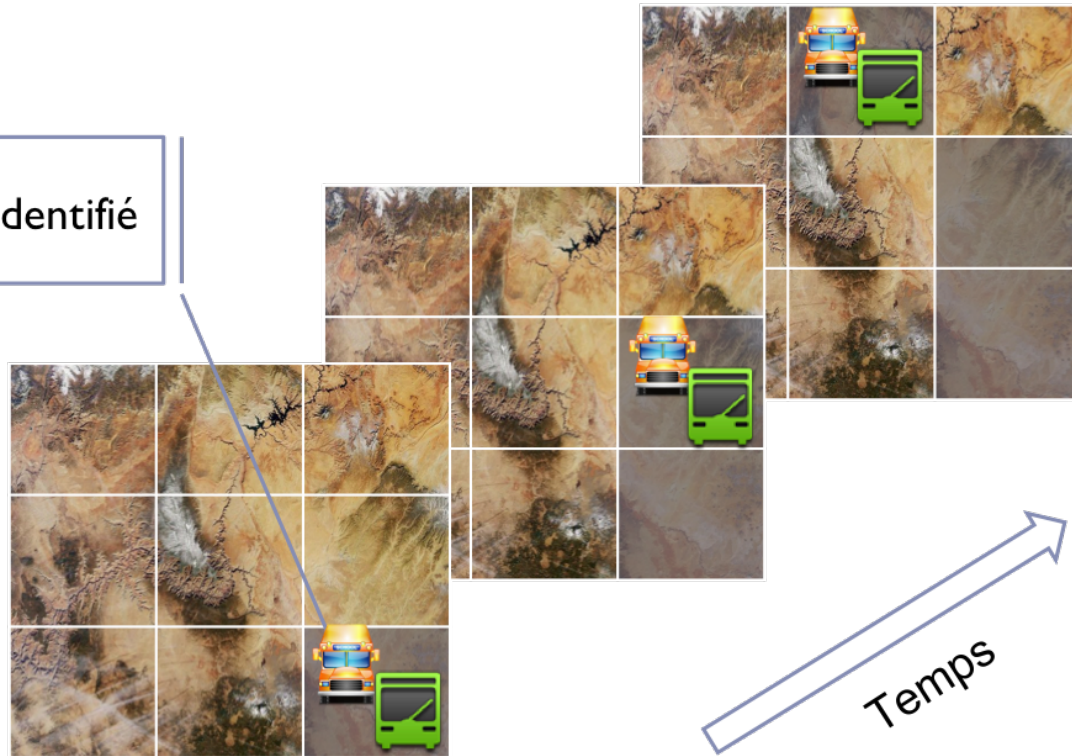


Fouille de données spatio-temporelles : des données aux motifs

# + Trajectoires d'objets mobiles


- Points localisés dans l'espace et le temps (**time-stamped, coordinates**)
- $T=(t_1, x_1, y_1), \dots, (t_n, x_n, y_n) \Rightarrow$  position de l'objet  $(x_i, y_i)$  au temps  $t_i$

Un seul objet identifié



# + Tableau de données

| Objet   | Date | $X_i$ | $Y_i$ |
|---------|------|-------|-------|
| Objet_1 | 1    | $X_1$ | $Y_1$ |
| Objet_1 | 2    | $x_2$ | $Y_2$ |
| Objet_2 | 1    | $x_3$ | $Y_3$ |
| Objet_2 | 2    | $x_4$ | $Y_4$ |
| Objet_3 | 1    | $X_5$ | $Y_5$ |
| ...     | ...  | ...   | ...   |



Dimension temporelle                      Dimension spatiale



# + Composante temporelle des trajectoires

- **time-based recording:** les positions des objets sont enregistrées à intervalle de temps réguliers
  - e.g. toutes les 5 minutes
- **change-based recording:** un enregistrement de la position se fait à chaque fois que l'objet change de position
  - e.g. à chaque fois que l'utilisateur d'un smartphone se connecte sur facebook dans un lieu « connu »
- **location-based recording:** : un enregistrement de la position se fait quand l'objet se rapproche d'une position
  - e.g. passage près de capteurs
- **event-based recording:** la localisation et le temps sont enregistrés pendant des événements
  - e.g. appel depuis un téléphone mobile
- **Durée dans le temps, périodicité**

# + Composante spatiale des trajectoires

■ **Géolocalisation** : positionnement d'un objet sur un plan ou une carte à l'aide de ses coordonnées géographiques.

- Latitude / Longitude
- Coordonnées polaires
- .... Extrait de textes



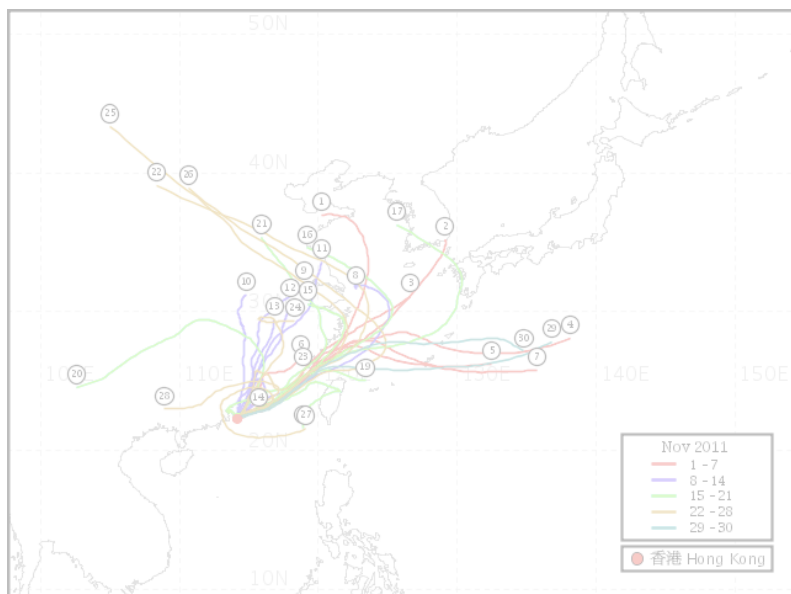
■ **Direction**

- Point cardinal
- Déplacement en ligne droite, curviligne, circulaire...
- Points d'inflexions dans le temps, l'espace, angle...
- Direction initiale, finale, vitesse

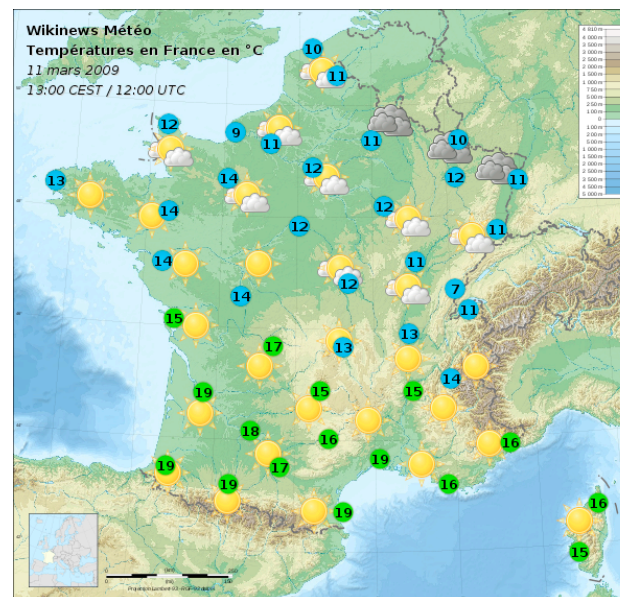


# + Deux perspectives

## ■ Les objets mobiles



## ■ Les évènements dans des zones fixées



# + Etude d'évènements localisés

- Zones localisées dans l'espace dans lesquelles se déroulent des événements dont on connaît l'estampille temporelle (**zone, time-stamped, event**)
- $T=(t_1, z_1, e_i), \dots, (t_n, z_n, e_n) \Rightarrow$  zone  $z_i$  dans laquelle se déroulent les événements  $e_i$  au temps  $t_i$



# + Tableau de données

| Zone   | Date | Température | Pluie | Vent |
|--------|------|-------------|-------|------|
| Zone_1 | 1    | bas         | bas   | 35   |
| Zone_1 | 2    | bas         | haut  | -    |
| Zone_2 | 1    | bas         | moyen | 12   |
| Zone_2 | 2    | haut        | moyen | -    |
| Zone_3 | 1    | bas         | -     | -    |
| Zone_3 | 2    | bas         | bas   | 23   |
| ...    | ...  | ...         | ...   | ...  |

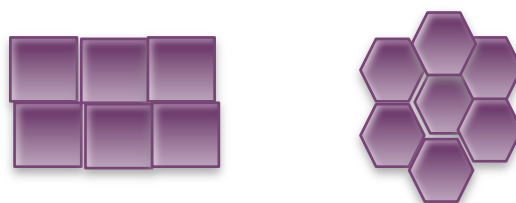


## + Composante temporelle des événements localisés

- **time-based recording:** tous les événements sont enregistrés à intervalle de temps réguliers dans une zone
  - e.g. toutes les 5 minutes
- **alert-based recording:** un enregistrement se fait à chaque fois qu'une dimension d'analyse particulière dépasse un seuil
  - e.g. à chaque fois que la température dépasse une valeur
- **identifiant-based recording:** l'enregistrement se fait par identification du matériel à un temps déterminé.
  - e.g. l'unité P0221 doit enregistrer le 04/02/2012 de 13h00 à 14h10

# + Composante spatiale des événements localisés

- **Différents zonages** : différents découpage
  - administratif: région, département, agglomération, ville...
  - démographique: bassin de vie, zone urbaine, ...
  - ....
- **Différentes représentation de l'espace** : Différentes grilles




- **Différentes proximités spatiales**
  - À côté de, près de, loin de
  - Amont / Aval ... un réseau hydrographique, routier



# Données spatiales **ET** temporelles





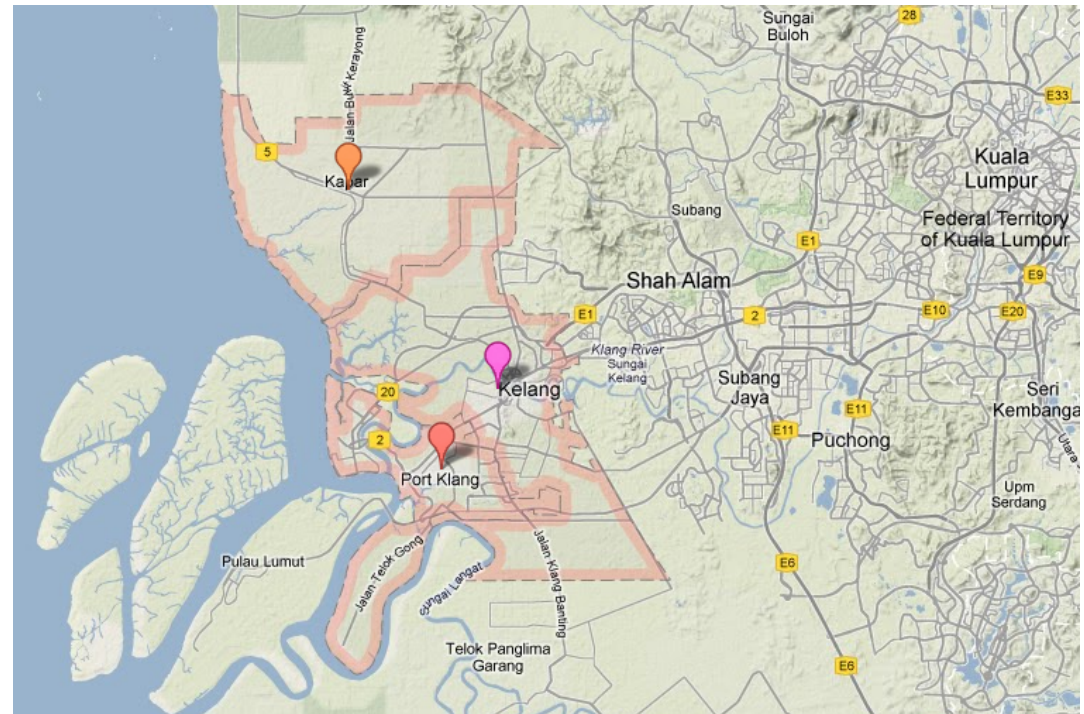
+ Illustration d'une méthode de  
recherche de séquences spatio-  
temporelles

# + Objectifs

- **Capturer des comportements collectifs**
  - Extraire des motifs d'événements spatiaux, fréquents au cours du temps
  - **Séquences spatio-temporelles**
- Développer des **approches efficaces** malgré un espace de recherche très important

# + Zonage

| area 1               | area 2          |
|----------------------|-----------------|
| Port Klang<br>Kelang | Kelang<br>Kapar |



Fouille de données spatio-temporelles : des données aux motifs

## + Bases de séquences spatio-temporelles

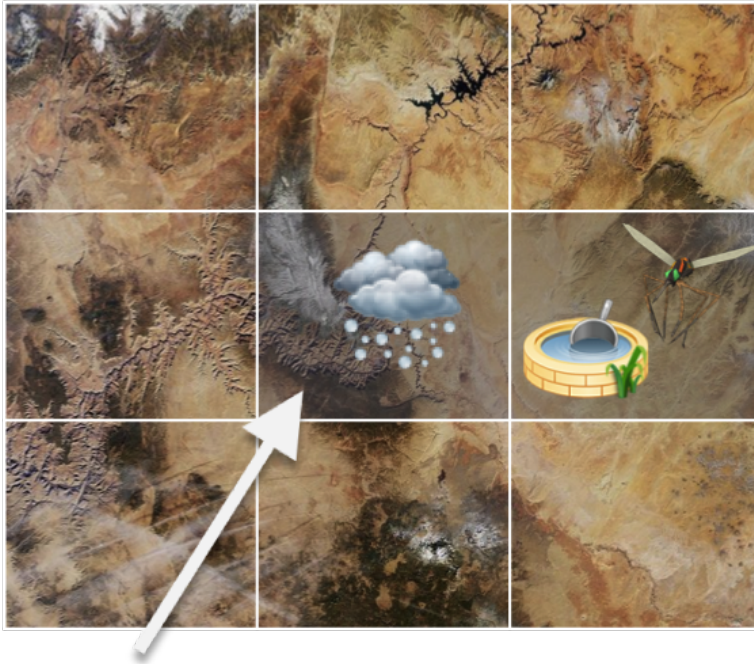
- On groupe les dimensions d'analyse par date et zone pour obtenir une **base de séquences sDB**

| area       | sequence of itemsets  |
|------------|---|
| Port Klang | (temp_low rain_low wind_35) (temp_low rain_high wind_-)     |
| Kelang     | (temp_low rain_mean wind_12) (temp_high rain_mean wind_-)   |
| Kapar      | (temp_low rain_- wind_-) <u>(temp_low rain_low wind_23)</u> |

*itemset*

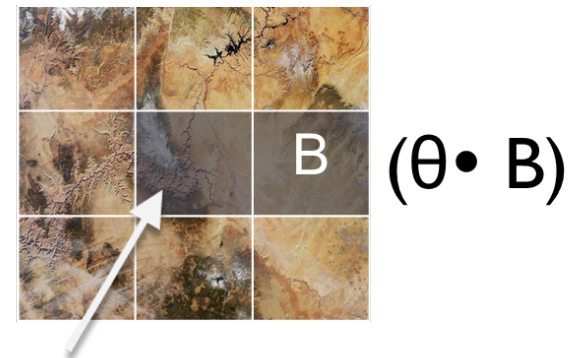
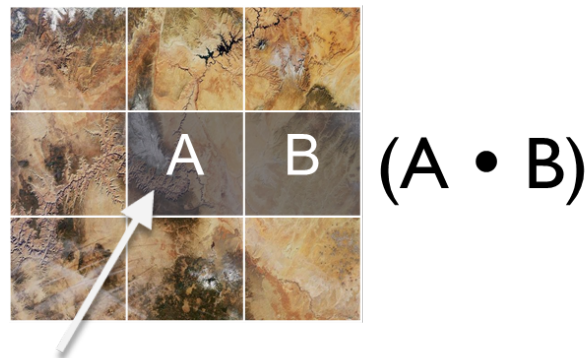
# + Itemset Spatial

- 2 items spatialement proches



# + Dynamique spatiale

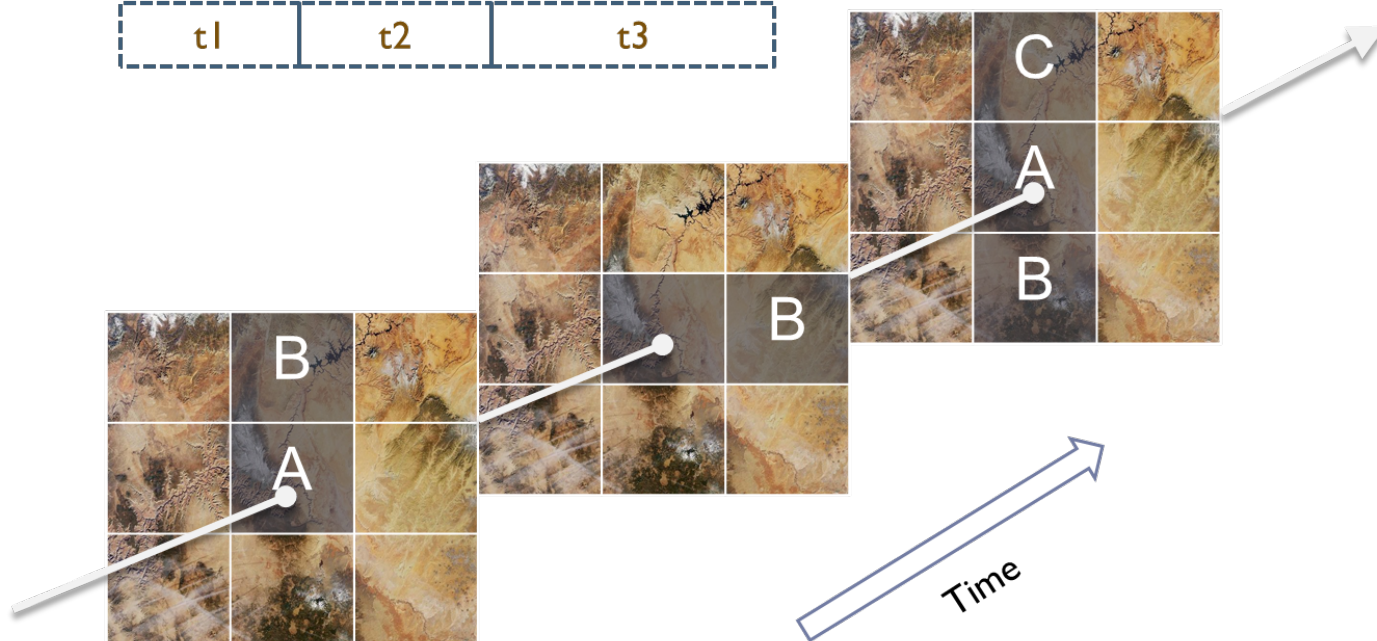
- Opérateur spatial  $\bullet$  (près de)
- Opérateur de groupage [ ]
- Symbole représentant l'absence d'itemsets  $\theta$



# + Séquences Spatiales 2S

## ■ Séquence d'itemsets spatiaux

$\langle (A \bullet B) (\theta \bullet B) (A \bullet [B ; C]) \rangle$



# + Problématique

## Fouille de séquences spatio-temporelles (2S)

- Extraire un ensemble de 2S fréquentes dans la base de séquences sDB :

$$\text{support}(2S) \geq \alpha$$

avec  $\alpha$  le support minimum

- Une 2S fréquente est également appelée **motif spatio-temporel ou S2P**



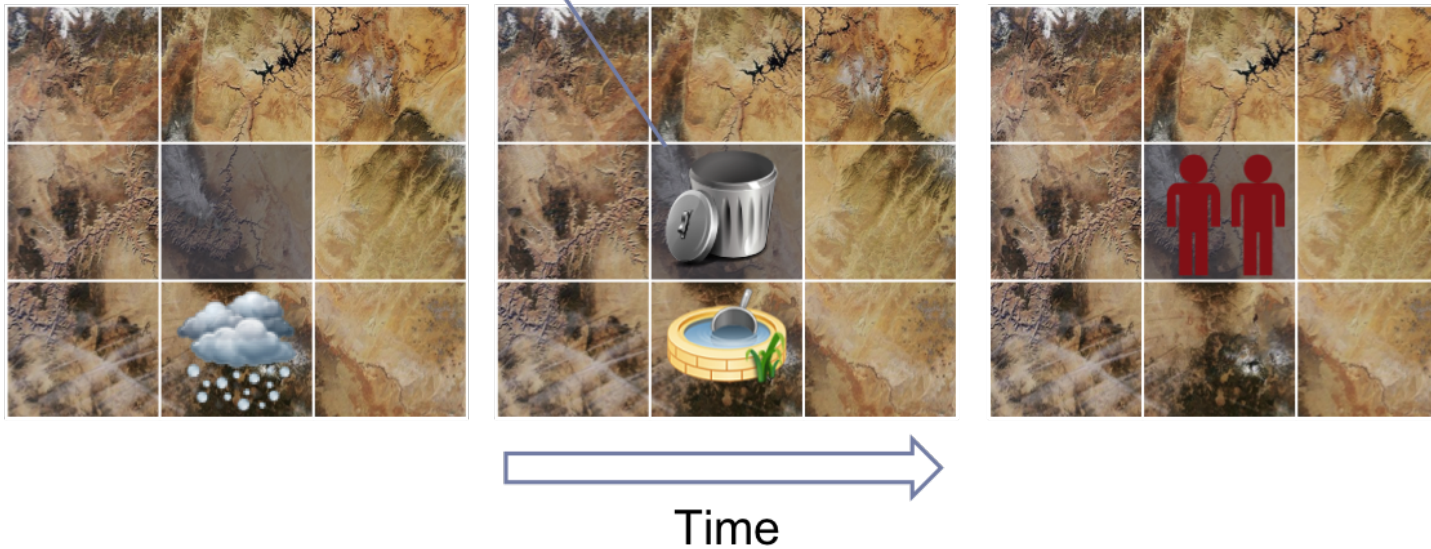
## + Résultats expérimentaux : application à la dengue en NC

- Données de la **population** : e.g., nombre d'habitants, le nombre de maisons, etc. (compris entre 1996 et 2003 par quartier)
- Données **entomologiques** : e.g., indice Breteau, l'indice d'haute risque entomologique, etc. (par mois et par quartier)
- Données **météo** : e.g., des précipitations, la température, l'humidité, etc. (journalier et pour chaque quartier)
- Données **urbaines** : e.g., nombre of piscines, nombre de bassin, nombre de fontaines, etc. (par quartier)
- Données **médicales** : e.g., nombre de cas de dengue (journalier et par quartier)

# + Exemple de motif

$\langle (\theta \cdot \text{precip} \leq 0.10) (\text{waste container} \leq 39.00 \cdot \text{indoor deposit} : (2126.00; 2692.50]) (\text{nb cas dengue} \leq 6.00) \rangle$

A set of characteristics



## + Contraintes appliquées aux motifs extraits

- Contrainte d'écart (GAP): spécifie la **durée maximale** (ou minimale) entre deux ensembles d'évènements spatiaux

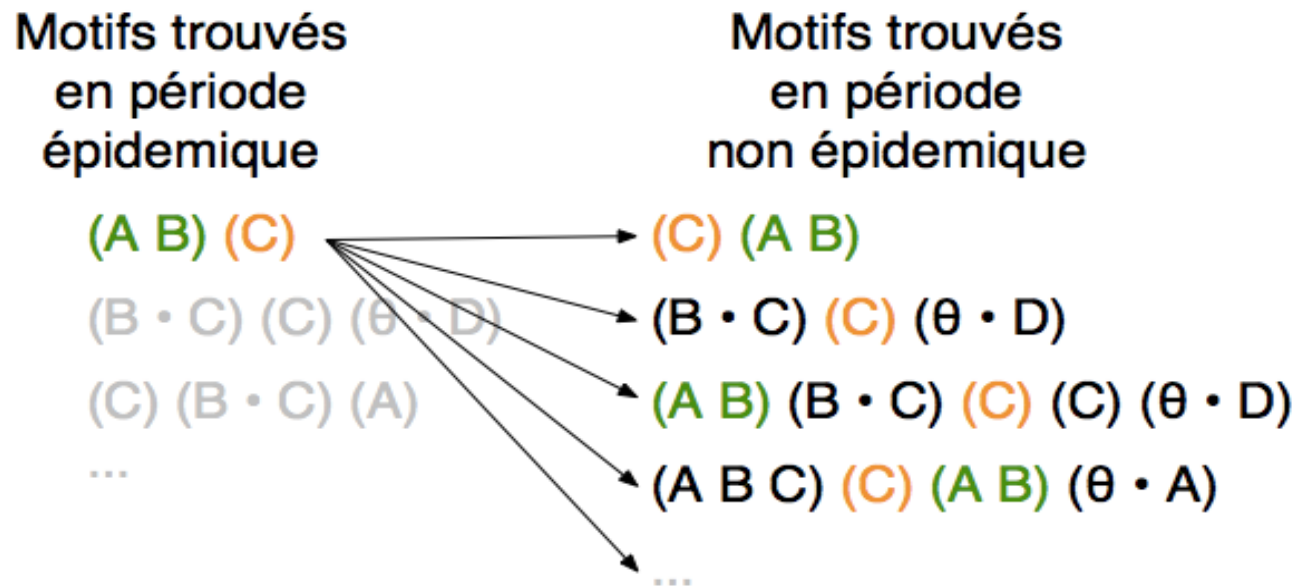
$$\underbrace{(AB)(B \bullet C)(C)(C)(\theta \bullet D)}_{\text{GAP}=2}$$

- Séquence maximale: la **plus grande** par rapport à l'inclusion

$$\begin{aligned} &(B \bullet C)(\theta \bullet D) \\ &(AB)(B \bullet C)(C)(C)(\theta \bullet D) \end{aligned}$$

# + Mesures de filtrage

- Recherche de séquences fréquentes qui sont **peu contredites** par les données (**moins contradiction temporelle**)
- Mesure **discriminante** (e.g., dengue/pas dengue)



# + Visualisation de motifs

**S2P Visualization**  
The Pattern Next Door: Towards Spatio-Sequential Pattern Discovery

Nouméa Map | Statistical data | Sequential patterns

Image © 2012 DigitalGlobe

**Legend**

- Simple itemset
- Itemset containing interesting event
- Time connection
- - - Neighbor connection

**Spatio Sequential Patterns**  
S2P for Ducos with a minimal support of 0.6 and a den as a interesting event

```
602; precip:<=0.10; $waste_container:<=39.00 $nb_cas_dengue:<=6.00 $graveyard:<=2.00; 0.6;
603; precip:<=0.10; $community_gather:<=20.00 $nb_cas_dengue:<=6.00; 0.7;
604; precip:<=0.10; $community_gather:<=20.00 $nb_cas_dengue:<=6.00 $outdoor_deposit:<=122.50; 0.6;
605; precip:<=0.10; $community_gather:<=20.00 $nb_cas_dengue:<=6.00 $outdoor_deposit:<=122.50 $graveyard:<=2.00; 0.6;
606; precip:<=0.10; $community_gather:<=20.00 $nb_cas_dengue:<=6.00 $graveyard:<=2.00; 0.6;
607; precip:<=0.10 community_gather:<=20.00 nb_cas_dengue:<=6.00; 0.7;
608; precip:<=0.10 community_gather:<=20.00; $waste_container:<=39.00 $community_gather:<=20.00 $nb_cas_dengue:<=6.00; 0.6;
609; precip:<=0.10 community_gather:<=20.00; $waste_container:<=39.00 $nb_cas_dengue:<=6.00; 0.6;
610; precip:<=0.10 community_gather:<=20.00; $community_gather:<=20.00 $nb_cas_dengue:<=6.00; 0.6;
611; precip:<=0.10 community_gather:<=20.00; $nb_cas_dengue:<=6.00; 0.6;
612; precip:<=0.10; $nb_cas_dengue:<=6.00; 0.7;
```

**Pattern visualisation**

Sequence: 609; precip:<=0.10 community\_gather:<=20.00; \$waste\_container:<=39.00 \$nb\_cas\_dengue:<=6.00; 0.6;


<http://datamining.univ-nc.nc/>

## + Bilan

- Analyser les **changements dans une zone** (quartier, rivière, pixel, etc.) par rapport au **temps**, tout en prenant en compte son **entourage proche**
- Deux **algorithmes génériques** ont été proposés: DFS-S2PMiner (depth- first stratégie) et BFS-S2PMiner (level- wise approche) ainsi que des **mesures d'intérêt** (spatio-temporal participation index, the least spatio-temporal contradiction)
- Outil **accessible** pour les experts

## + Perspectives

- Etendre la notion de voisinage à des **relations spatiales plus complexes**
- Prendre en compte les **évolution des zones géographiques** au cours de temps, e.g., si deux zones se fusionnent



+ Illustration d'une méthode  
de recherche de trajectoires



# + Objectifs

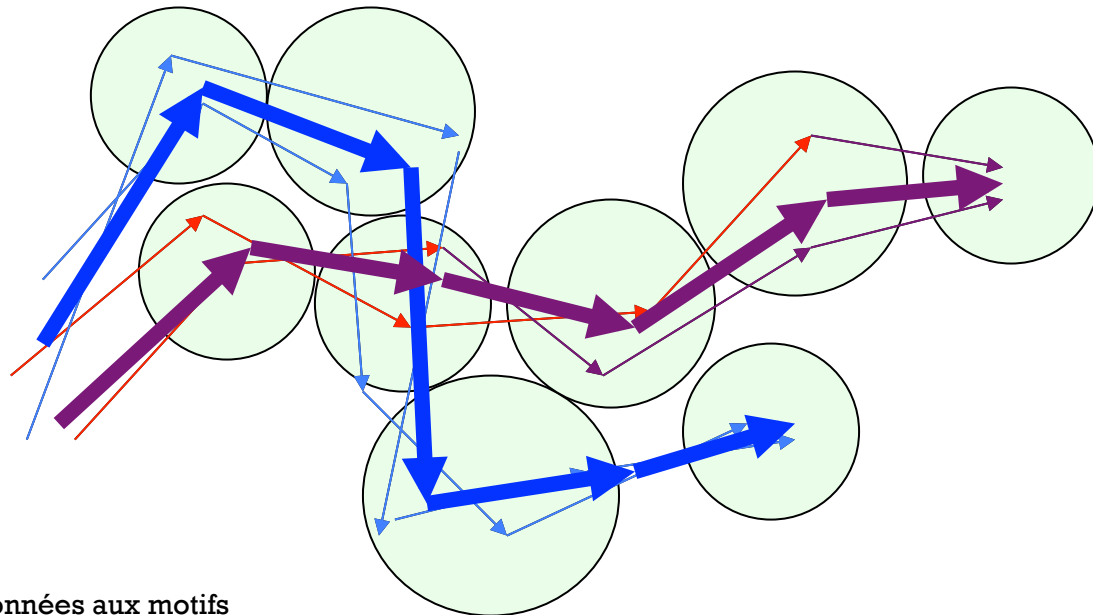
- **Capturer des comportements collectifs**
  - Extraire des motifs qui capturent les comportements de groupe d'objets au cours du temps
- **Trajectoires**
- Développement d'**approches efficaces** pour identifier des groupes d'objets mobiles pour lesquels il existe une relation forte ou une interaction
  - dans une région spatiale définie
  - pendant une certaine durée

# + Fouille de motifs Spatio-Temporels depuis des données de trajectoire

## ■ Clustering

- Grouper ensemble des trajectoires similaires
- Pour chaque groupe, produire un résumé : *Flock*, *convoy*, *moving cluster*, *swarm*, *closed swarm*, *Star*, *k-Star*

Exemple de clustering



# + Problématique

- De nombreux types de motifs existent (*Flock, convoy, moving cluster, swarm, closed swarm, Star, k-Star*) avec leur propre méthode d'extraction
- Des approches très efficaces ont été définies pour extraire des **règles d'association**
  - *Apriori, Closed, Derivable, ...*
- **Intuition ...**

Est il possible de les adapter pour extraire des trajectoires?

# + Cluster Matrix

- **Motifs:** Evolution de cluster
- **Objects:** transactions
- **Clusters:** items

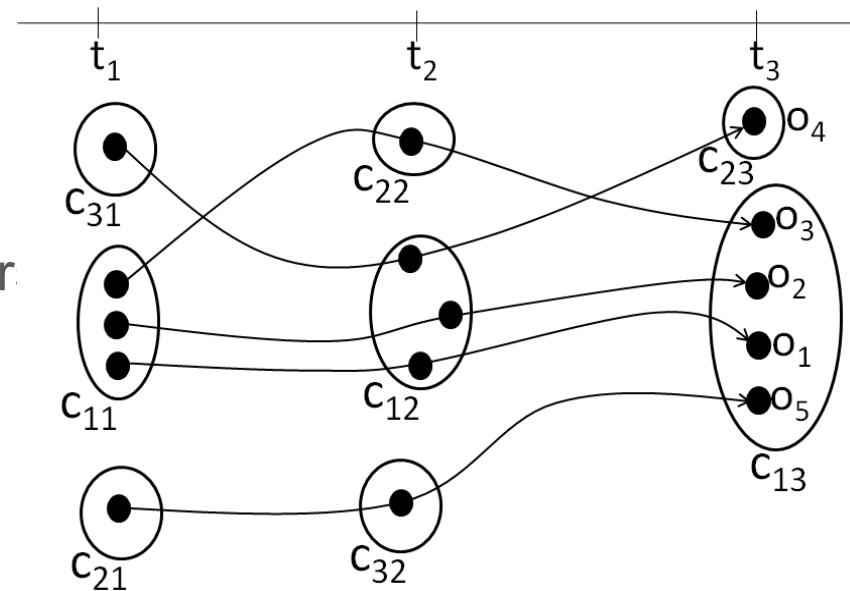


Table 2: Cluster Matrix

| $T_{DB}$          |       | $t_1$    |          |          | $t_2$    |          |          | $t_3$    |          |
|-------------------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| Clusters $C_{DB}$ |       | $c_{11}$ | $c_{21}$ | $c_{31}$ | $c_{12}$ | $c_{22}$ | $c_{32}$ | $c_{13}$ | $c_{23}$ |
| $O_{DB}$          | $o_1$ | 1        |          |          | 1        |          |          | 1        |          |
|                   | $o_2$ | 1        |          |          | 1        |          |          | 1        |          |
|                   | $o_3$ | 1        |          |          |          | 1        |          | 1        |          |
|                   | $o_4$ |          |          | 1        | 1        |          |          |          | 1        |
|                   | $o_5$ |          | 1        |          |          |          |          | 1        | 1        |



Database: Buffalo  
min\_o: 10 min\_t: 10  
[About the parameters](#)  
Search Clean Map

(a) One of discovered closed swarms.

Running time: 0.00 seconds  
No. Closed Swarms: 808 No. Convoys: 10 No. Group Patterns: 10

**Pattern: Closed swarm 811; Animal name: Leo**

**Closed Swarm: 811**

- Objects: Aphrodite, Rose, Nkonya, Helena, Santa, Lorraine, Mbirhi, Antigone, Cillie, Green, Leo, Nebula,
- Timespan: 10/11/2003 - 9/6/2004

**Closed Swarm: 812**

**Convoy: 0**

- Objects: Nkombo, Nkonya, Viola, Xitekwa, Kaye, Xirhumbana, Lontshwa, Nharhu, Tihlo, Nhungu, Tatana, Nkulukumba, Tinkohe, Xidlodlo, Xigangu, Wansati, Ntsumi, Khume,
- Timespan: 7/9/2001 - 7/23/2001

**Convoy: 1**

**Group Pattern: 0**

- Object: Nkombo, Nkonya, Viola, Xitekwa, Kaye, Xirhumbana, Lontshwa, Nharhu, Tihlo, Nhungu,

# + Bilan

- Une approche **unifiée** et efficace pour extraire différents types de trajectoires
- Possibilité de **minimiser** le nombre de motifs extraits (top-k)
  - Trouver seulement les trajectoires les plus pertinentes par rapport aux données
- Deux démonstrateurs

<http://www.lirmm.fr/~phan/index.jsp>

<http://www.lirmm.fr/~phan/realgp.jsp>



+

Conclusion

# + Challenges pour la fouille de données spatio-temporelles

## ■ Méthodes bien connues

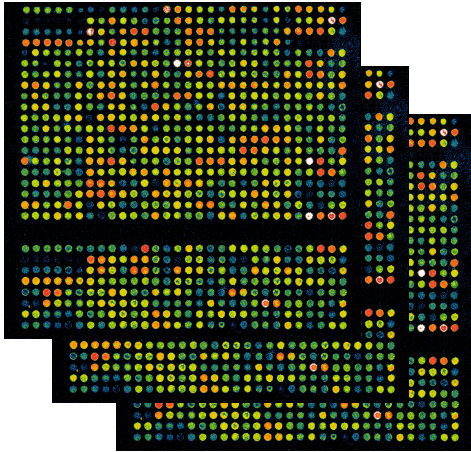
- Clustering
- Classification
- Motifs
  - RA spatiales
  - Co-location
  - Trajectoires
- ....

## ■ Questions classiques en Fouille

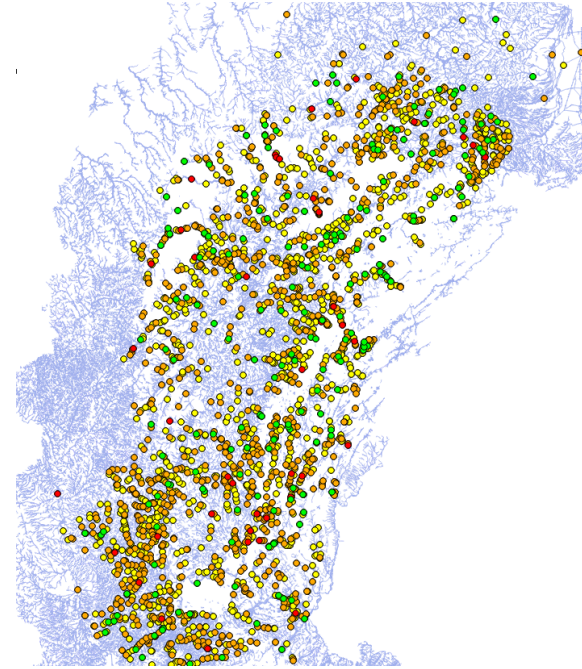
- Passage à l'échelle
- Qualité des motifs extraits
- Intégration des connaissances du domaine
  - Metadonnées
  - Ontologies
  - ...



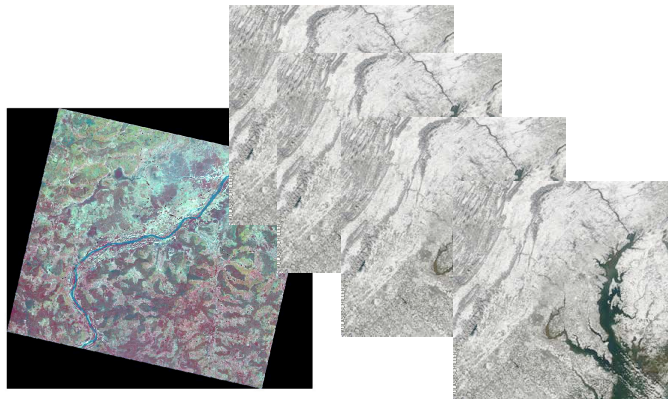
# + Fouille de données spatiales ?



Puces ADN



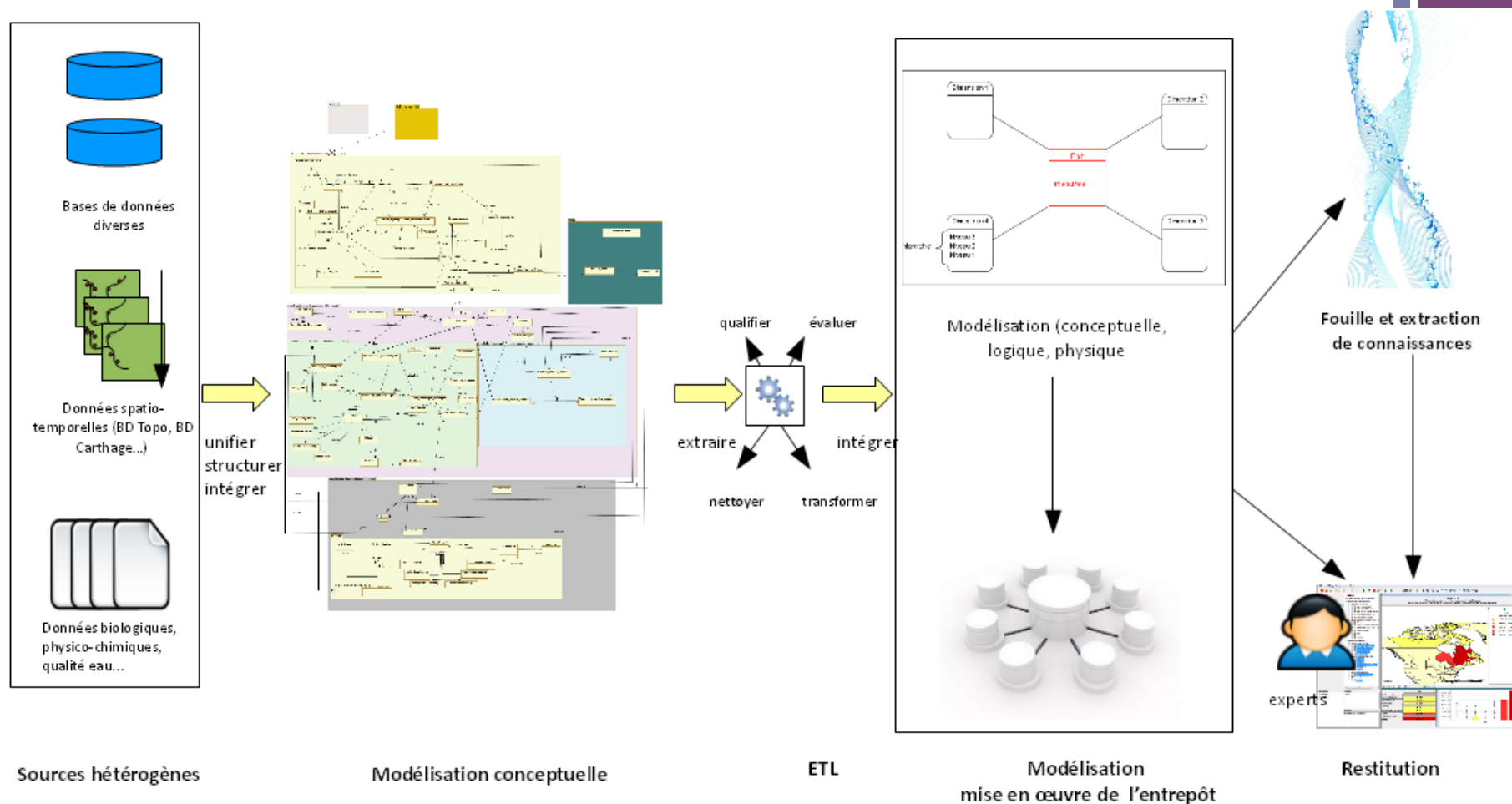
Stations du bassin de la Saône



Images satellites (Spot - Modis)

Fouille de données spatio-temporelles : des données aux motifs

# + Projet Fresqueau (ANR MN – MD)



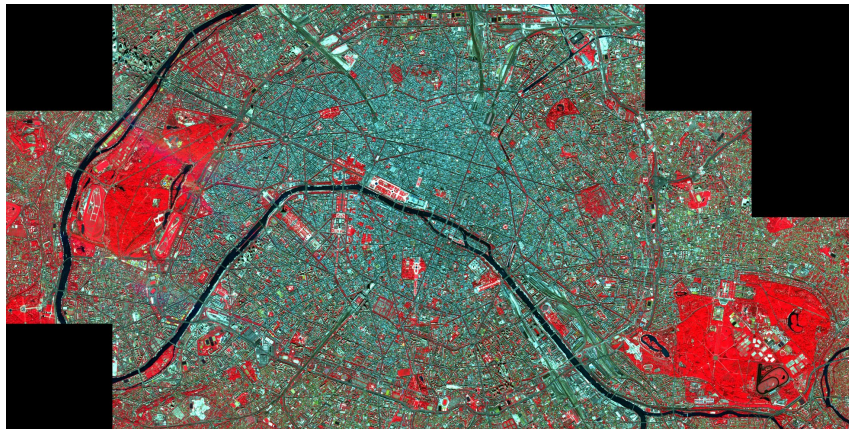
Fouille de données spatio-temporelles : des données aux motifs

## + A court terme

- Prendre en compte **la sémantique** pour identifier des motifs "plus" **intéressants**
  - Événements: inattendus, moindre contradiction temporelle...
  - Trajectoires: les maximales, top-k...
- **Comparer** les motifs
  - Événements: inclusion spatiale, temporelle...
  - Trajectoires: similarité de forme, de direction, de proximité...
  - Besoin d'une similarité sémantique

## + Et puis ?

- Fouille d'images satellites (Remote Sensing Image Mining) avec une perspective temporelle et spatiale encore peu explorée
- Beaucoup de données,... mais peu de méthode de fouille de données ! **GEOSUD**



Fouille de données spatio-temporelles : des données aux motifs



# + Des images et ... des textes

## Notre corpus



Images Pléiades (résolution spatiale 2x2 m – 4 bandes spectrales) couvrant l'ensemble de la région Thau Agglomération

(juillet et septembre 2012, EQUIPEX GEOSUD)



Villeveyrac

Correspondant

6 avril 2013

Midi Libre

Le rond-point de la route de Montagnac sera achevé fin avril

Depuis quelques semaines, les emprises des ronds-points sont bien délimitées. Un travail sur demi-chaussée est entrepris pour traiter les bretelles de Montagnac (RD 5), avec une circulation réglementée par feux tricolores. Le giratoire, décalé au niveau de l'usine 3S pour englober la RD 5 via le barreau en direction de la RD 2, ainsi que le petit chemin des agriculteurs, est presque terminé avec les bordures en ciment mises en place. D'ici fin avril, le rond-point de la route de Montagnac pourra être utilisé totalement, avec bien entendu une végétalisation prévue par la suite.

Corpus textuel : 94000 articles de  
presse du Midi Libre

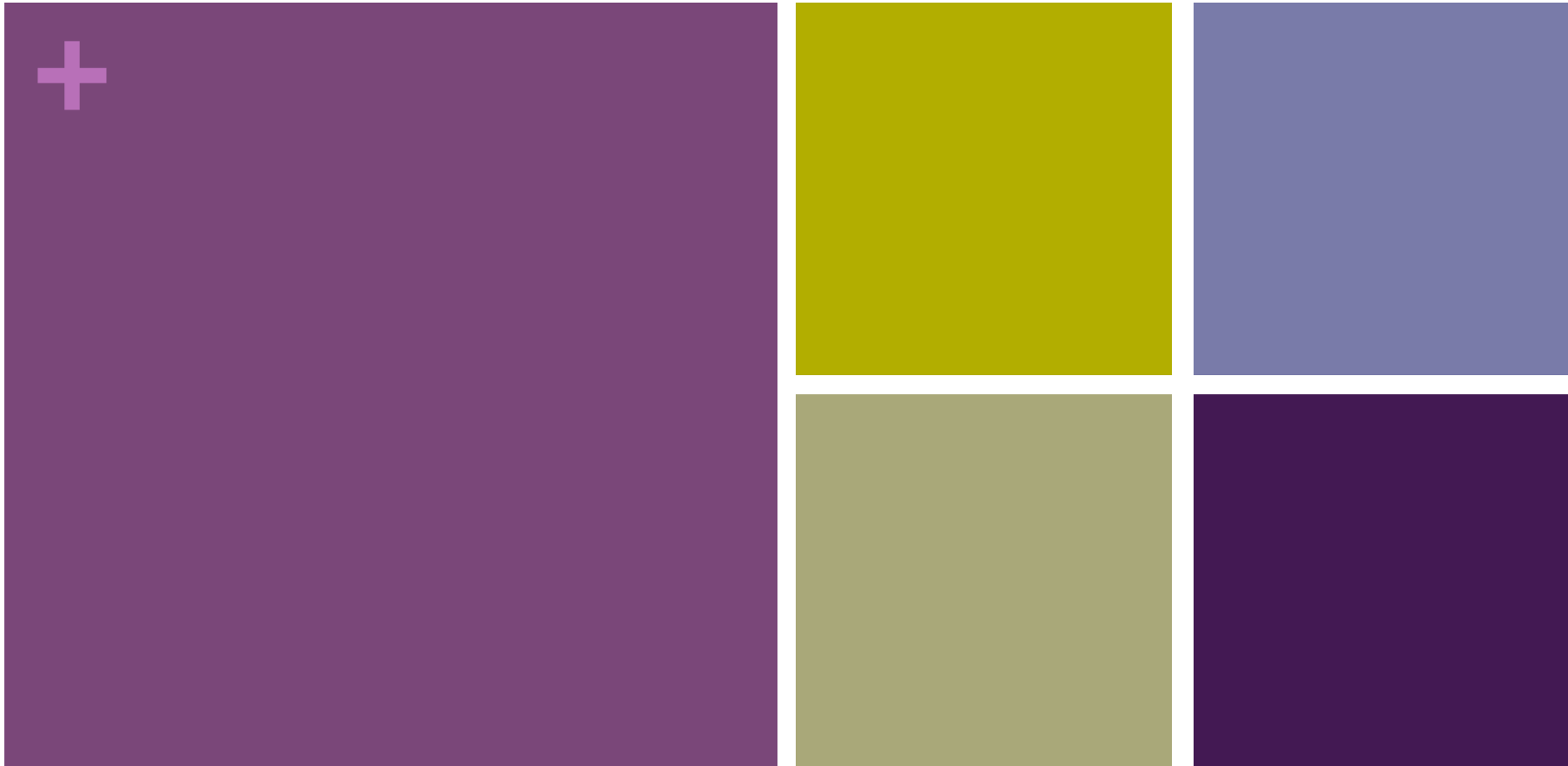
(2001 – 2014)

# + Références

- I. Tsoukatos, D. Gunopulos, «Efficient mining of spatiotemporal patterns», In Advances in Spatial and Temporal Databases, p. 425-442, 2001
- H. Cao, N. Mamoulis, D. Cheung, «Mining frequent spatio-temporal sequential patterns», In ICDM 2005
- Y. Huang, L. Zhang, and P. Zhang. «A framework for mining sequential patterns from spatio-temporal event data sets». In TKDE, p. 433-448, 2008
- P. Mohan, S. Shekhar, J. A. Shine, and J. P. Rogers. Cascading spatio-temporal pattern discovery: A summary of results». In SDM, p. 327-338, 2010.
- LVQ Anh, M Gertz, «Mining Spatio-temporal Patterns in the Presence of Concept Hierarchies», In ICDMW, 2012
- N. Mamoulis, H. Cao, G. Kollios, M. Hadjieleftheriou, Y. Tao, and D. W. Cheung. Mining, indexing, and querying historical spatio-temporal data. In KDD '04, pages 236–245, 2004.
- P. Kalnis, N. Mamoulis, and S. Bakiras. On discovering moving clusters in spatio-temporal data. In SSTD'05, pages 364–381, 2005.

# + Références

- Y. Wang, E.-P. Lim, and S.-Y. Hwang. Efficient mining of group patterns from user movement data. *DKE*, 57(3):240–282, June 2006.
- C.S. Jensen, D. Lin, and B. C. Ooi. Continuous clustering of moving objects. *TKDE*, 19(9):1161–1174, sept. 2007.
- H. Jeung, M. L. Yiu, X. Zhou, C. S. Jensen, and H. T. Shen. Discovery of convoys in trajectory databases. In *PVLDB*, 1(1):1068–1080, August 2008.
- M. R. Vieira, P. Bakalov, and V. J. Tsotras. On-line discovery of flock patterns in spatio-temporal data. In *GIS '09*, pages 286–295, 2009.
- H. H. Aung, K. -L. Tan: Discovery of Evolving Convoys. In *SSDBM*, pages 196-213, 2010.
- Z. Li, B. Ding, J. Han, and R. Kays. Swarm: mining relaxed temporal moving object clusters. In *PVLDB*, 3(1-2):723–734, September 2010.
- P. N. Hai, D. Ienco, P. Poncelet, and M. Teisseire. Mining time relaxed gradual moving object clusters. In *GIS '12*, pages 478-481, 2012.
- P. N. Hai, P. Poncelet, and M. Teisseire. Get\_Move: An efficient and unifying spatio-temporal pattern mining algorithm for moving objects. In *IDA*, pages 276–288, 2012.
- K. Zheng, Y. Zheng, J. Yuan, and S. Shang. On discovery of gathering patterns from trajectories. In *ICDE*, 2013.



## Fouille de données spatio-temporelles

Sandra BRINGAY et Maguelonne TEISSEIRE