



Ecole Nationale Polytechnique d'Oran – Maurice Audin (Polytech'Oran)
Département de Génie des Systèmes Informatiques
Filière: Systèmes d'Information & Télécom
Spécialité: Ingénierie et Management des Systèmes d'Information (IMSI) & RTélécom

1

Introduction aux SIG

Systèmes d'Information Géographique



Complément du Master (IMSI) & Ingénieur RT
2022/2023

DR. BRAHAMI MENAOUER (HDR)
E-MAIL: MENAOUER.BRAHAMI@ENP-ORAN.DZ

Présent sur:



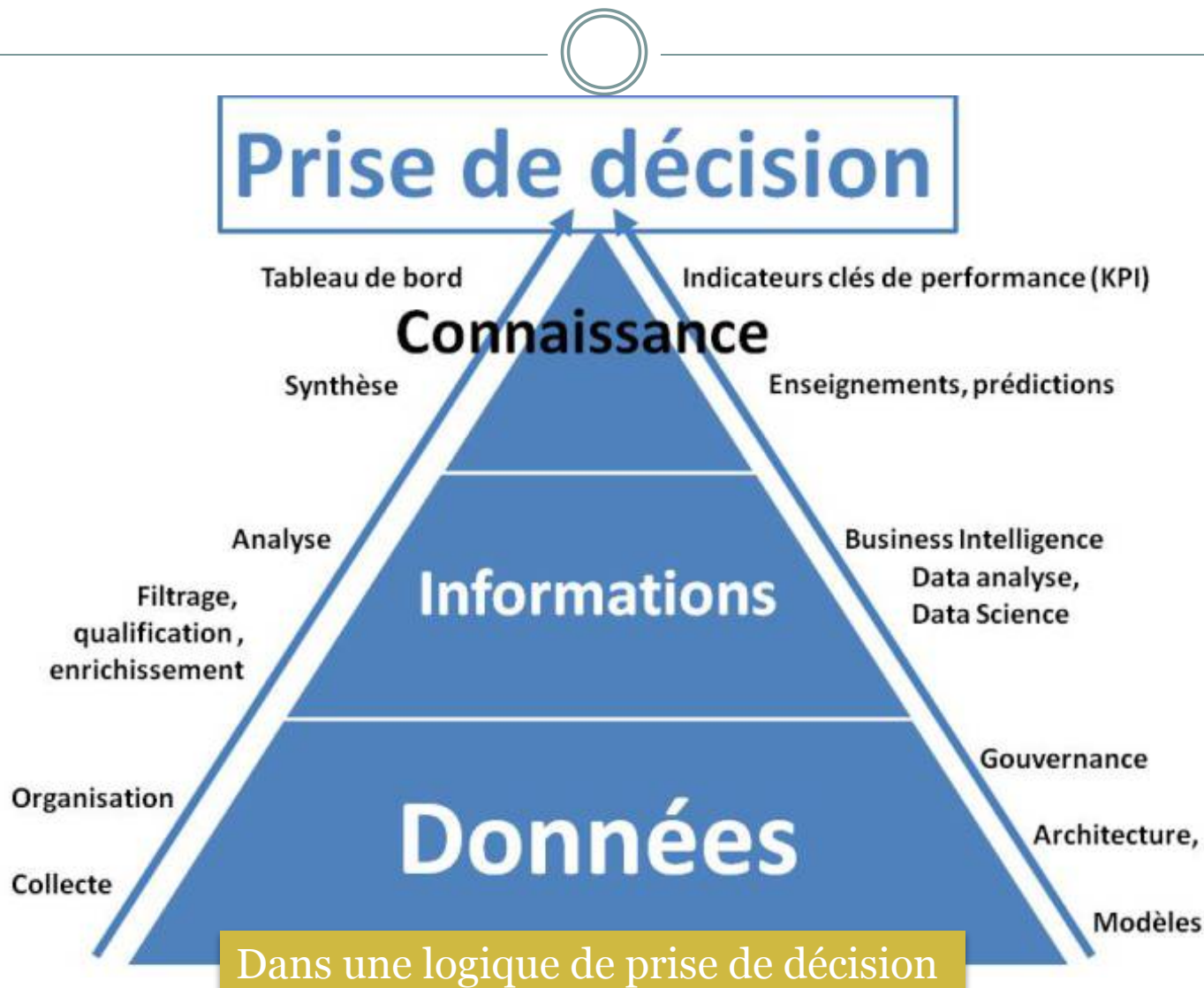
Partie4

Spatial Data Mining

2

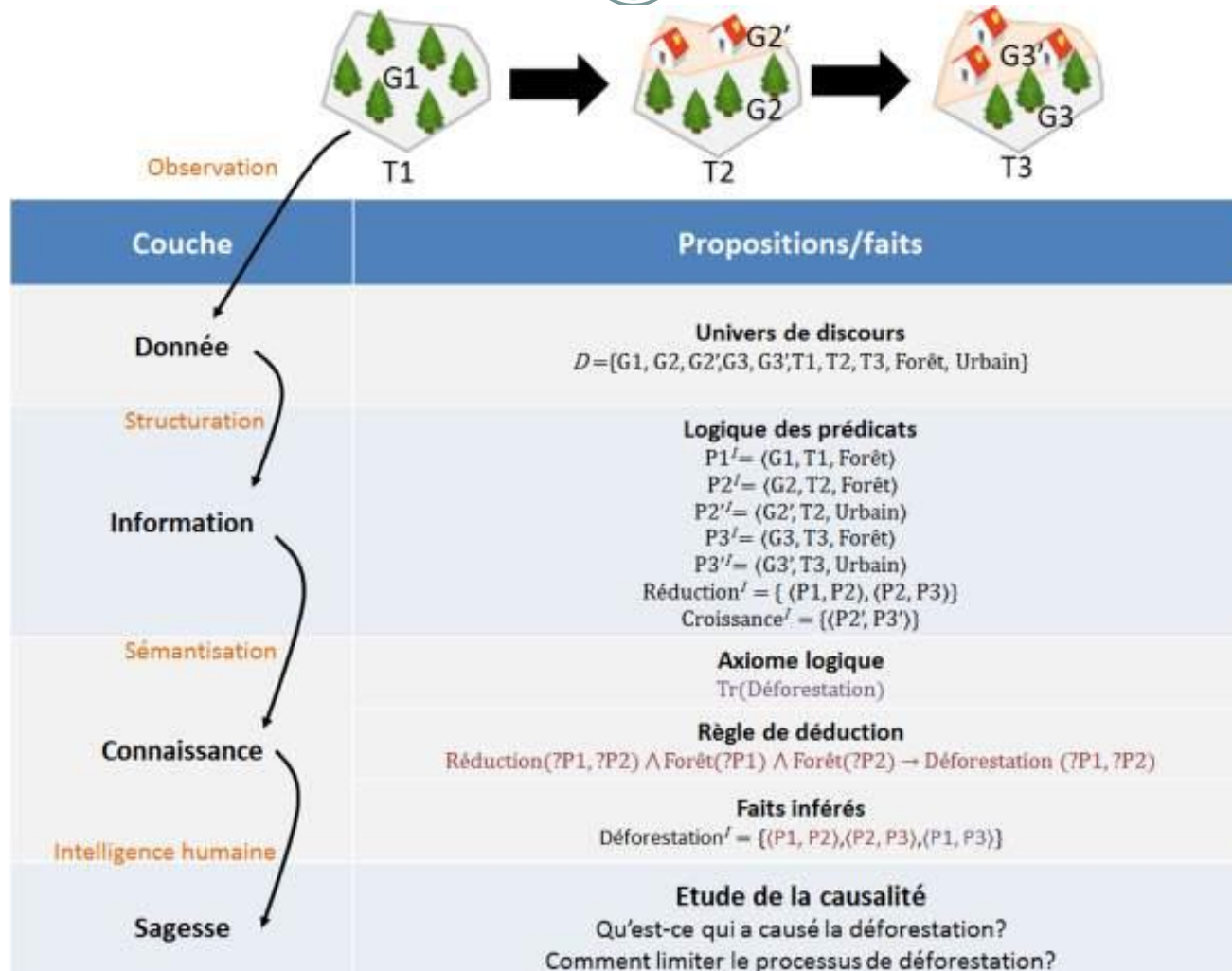
**IMPORTATION & INTERROGATION DES BASES DE
DONNÉES SPATIALES**

Données aux Connaissances



Exemple de déforestation (D, I & K)

4



Introduction: Différents Types d'analyse



Analyse
Descriptive



Que se
passe-t-il ?

Analyse
Diagnosticque



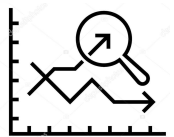
Pourquoi est-
ce arrivé ?

Que devrais-
je faire ?



Analyse
Prescriptive

Que
pourrait-il
arriver ?



Analyse
Prédictive

Introduction: Motivation

- ❑ Croissance de la capacité de stockage ;
- ❑ Croissance de la puissance de calcul;
- ❑ Croissance de la bande passante ;
- ❑ Amélioration des algorithmes et de la recherche;
- ❑ Besoins accrus d'avoir une vue globale synthétique en temps réel ;
- ❑ Demandes des entreprises, des gouvernements et bientôt aussi des individus.



Les données Spatiales

7

- Les **données spatiales** sont de plus en plus nombreuses grâce à l'évolution:
 1. des outils d'acquisition de données :exp. GPS, images satellite, photo aériennes, etc.);
 2. des méthodes de structuration: exp. raster, vecteur;
 3. des représentations: exp. représentations 2D, 3D.
- Une **donnée spatiale** est généralement composée de trois parties:
 - ✓ une partie géométrique (la forme par exemple **ligne, point, polygone**),
 - ✓ une partie descriptive (exp. **le nom, la population d'une ville**),
 - ✓ une partie métrique calculée (exp. **la superficie, le périmètre**)

Les données spatiotemporelles

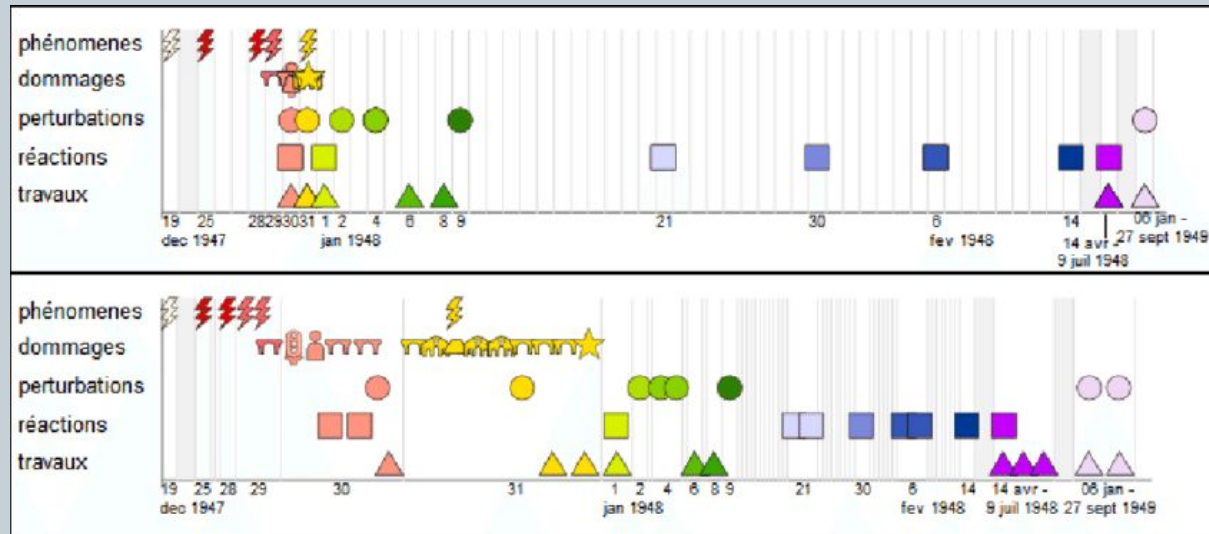
8

- Une donnée spatio-temporelle peut être définie comme **une donnée spatiale dont la forme et/ou la position varie(nt) au cours du temps.**
- Modélisation logique de l'information géographique à référence spatial et temporel :
 - modéliser les évolutions du monde réel et de traiter des données spatio-temporelles.
- Compréhension de la dynamique des **phénomènes du monde réel.**
- Développement des **technologies SIG,**
- Développement des **dispositifs GPS**

Les données spatiotemporelles -2

9

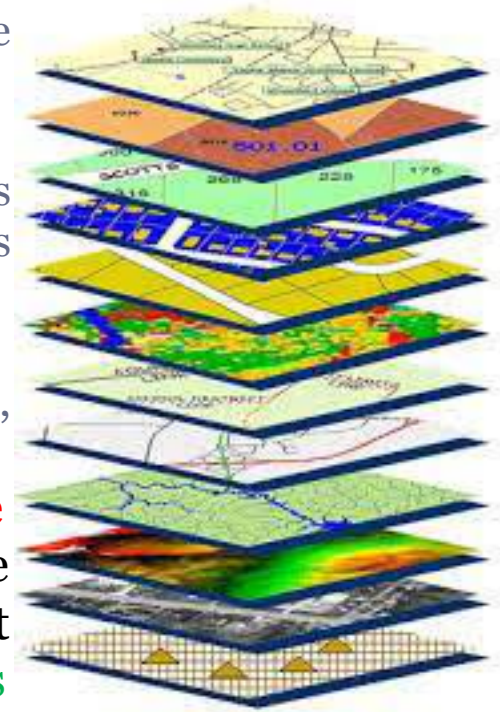
- Quelle que soit la perception que l'on a du **temps**, on peut résumer les différents concepts temporels nécessaires à sa modélisation :
 - **Événement** : Représente (un instant bien précis) sur la ligne du temps
 - Enregistrement GPS
 - **Intervalle temporel** : de modélisation/ représentation
 - Toutes les seconde, les minutes, les heures, les jours.
 - **Période**: Représente un ensemble d'intervalles
 - Sur une journée, une semaine, un mois.



Définition d'une BD spatiale

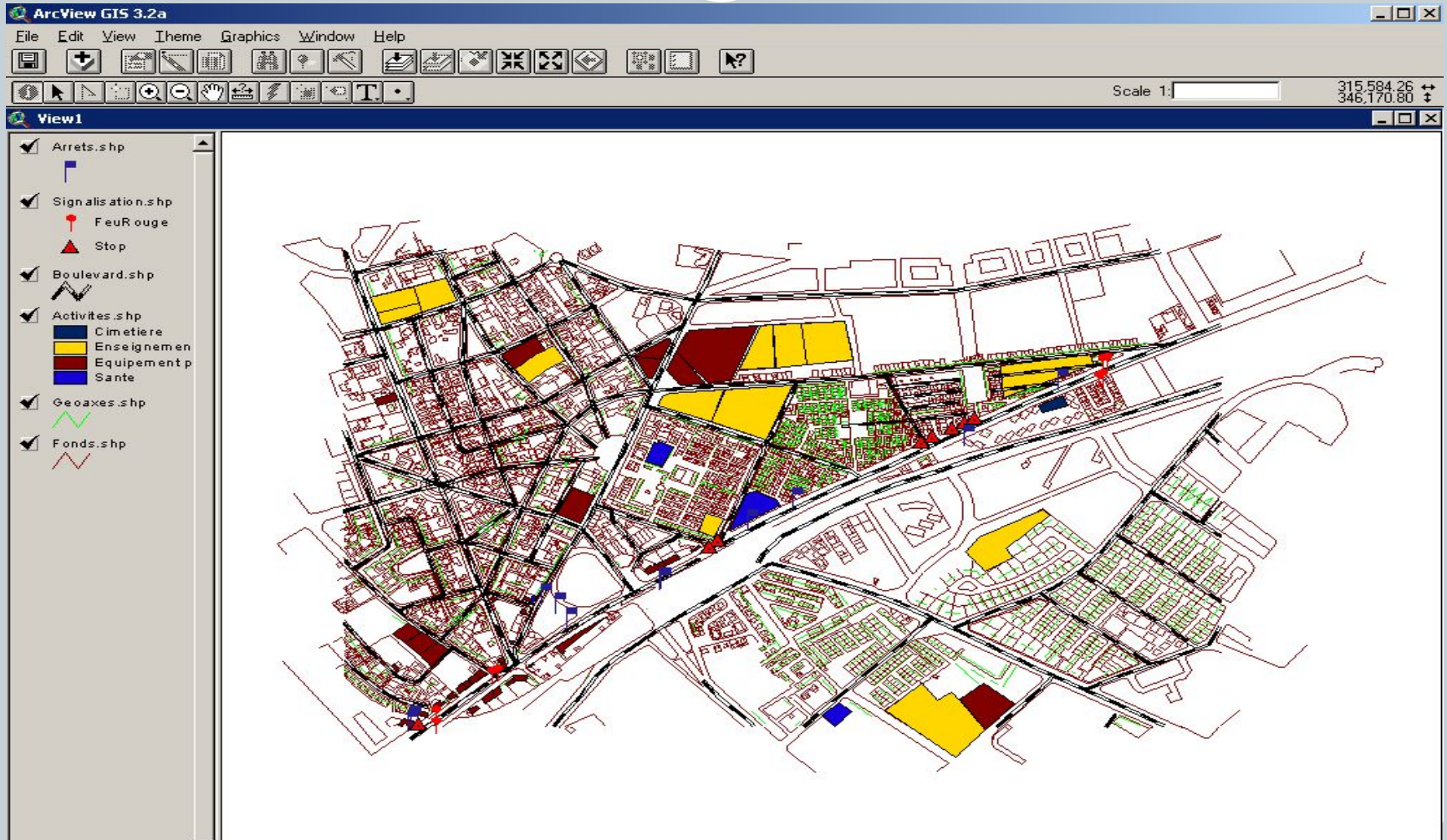
10

- Ensemble organisé d'objets géographiques :
 - Chaque objet est une association d'une description qualitative ou quantitative et d'une localisation spatiale ;
- Gérée au sein d'un SIG: Point, Ligne, Polygone
 - Tout type de données comportant un ou plusieurs champs concernant l'emplacement, la forme, la zone et des attributs similaires;
- Organisée en couches thématiques:
 - Ex : découpage administratif, Réseaux routier, Cadastre, POS, Topographie (courbes de niveau)...
- La fouille de données spatiales (FDS) est la **technique d'exploration de données géographiques** dont le but est de trouver des **motifs** intéressants dans les données contenant à la fois du **texte**, **des données temporelles**, **des données géométriques** .
- La FDS est définie comme **l'extraction de connaissances implicites** de **relations spatiales** ou **d'autres propriétés non explicitement** stockées dans les **bases de données spatiales**.



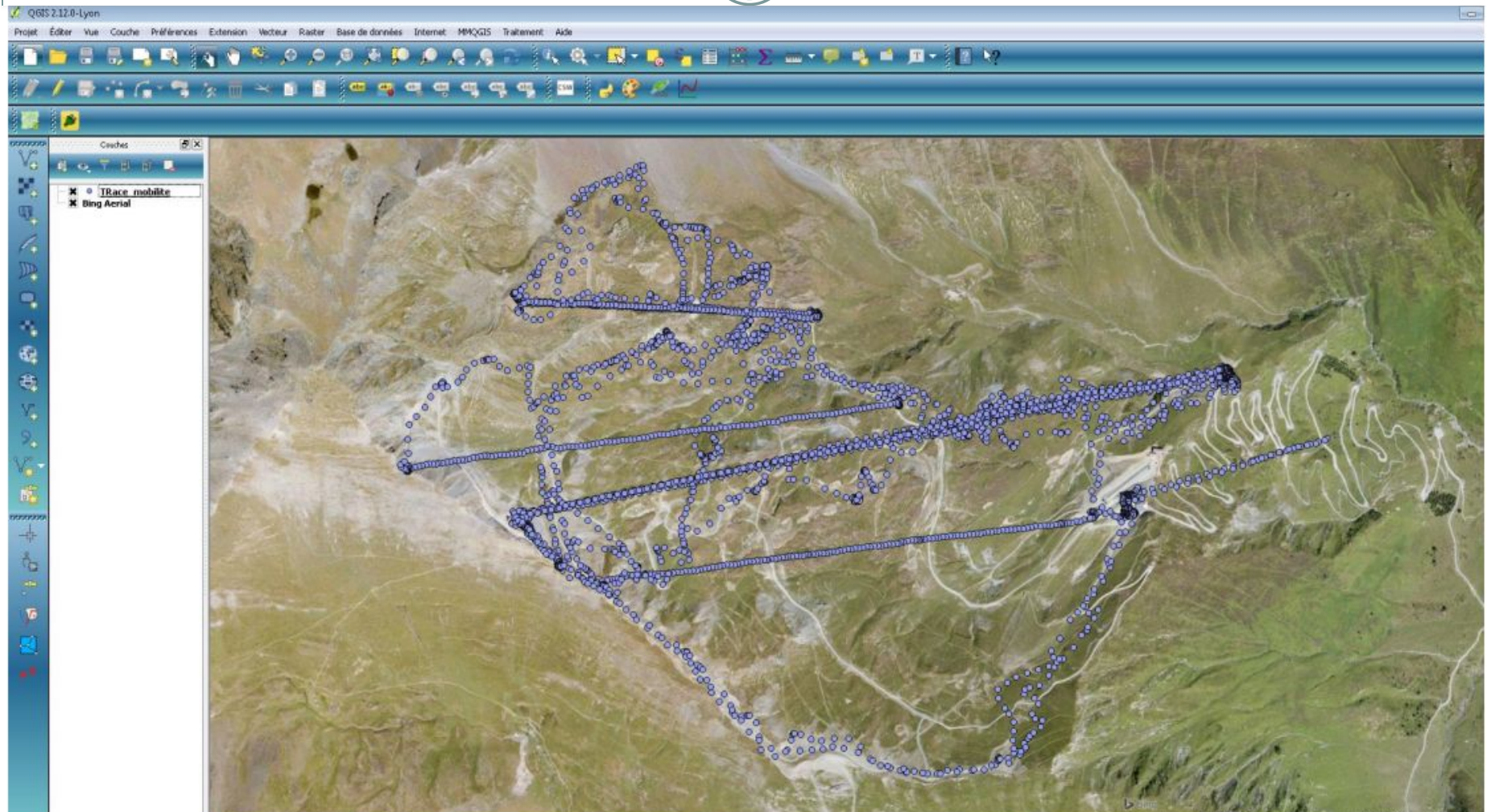
Visualisation des points des arrêts, Boulevards,...

11



Visualisation des points de GPS

12



Spécificités des données géographiques

13

- Positionnées / localisées par leur **emprise** spatiale : données spatiales en **2D ou 3D** par rapport à un référentiel :
 - **Géométrie de position** :
 - ▢ Discrètes : ponctuelles, linéaires, zonales; simples, composées ou complexes.
 - ▢ Continues : champs, surfaces (volumes).
 - **Topologie** :
 - ▢ Nœuds, arêtes (arcs), orientation, superposition...
- Qualifiées par leurs **attributs** : données descriptives alphanumériques.
 - **Qualitatives** : binaires, nominales (n-aires), ordinales.
 - **Quantitatives** : comptages, mesures...
 - **Éventuellement multimédia** : images, sons, etc. ou documentaire : rapports, plans, etc.
- Structurées selon une approche **vectorielle ou maillée** (image ou raster).

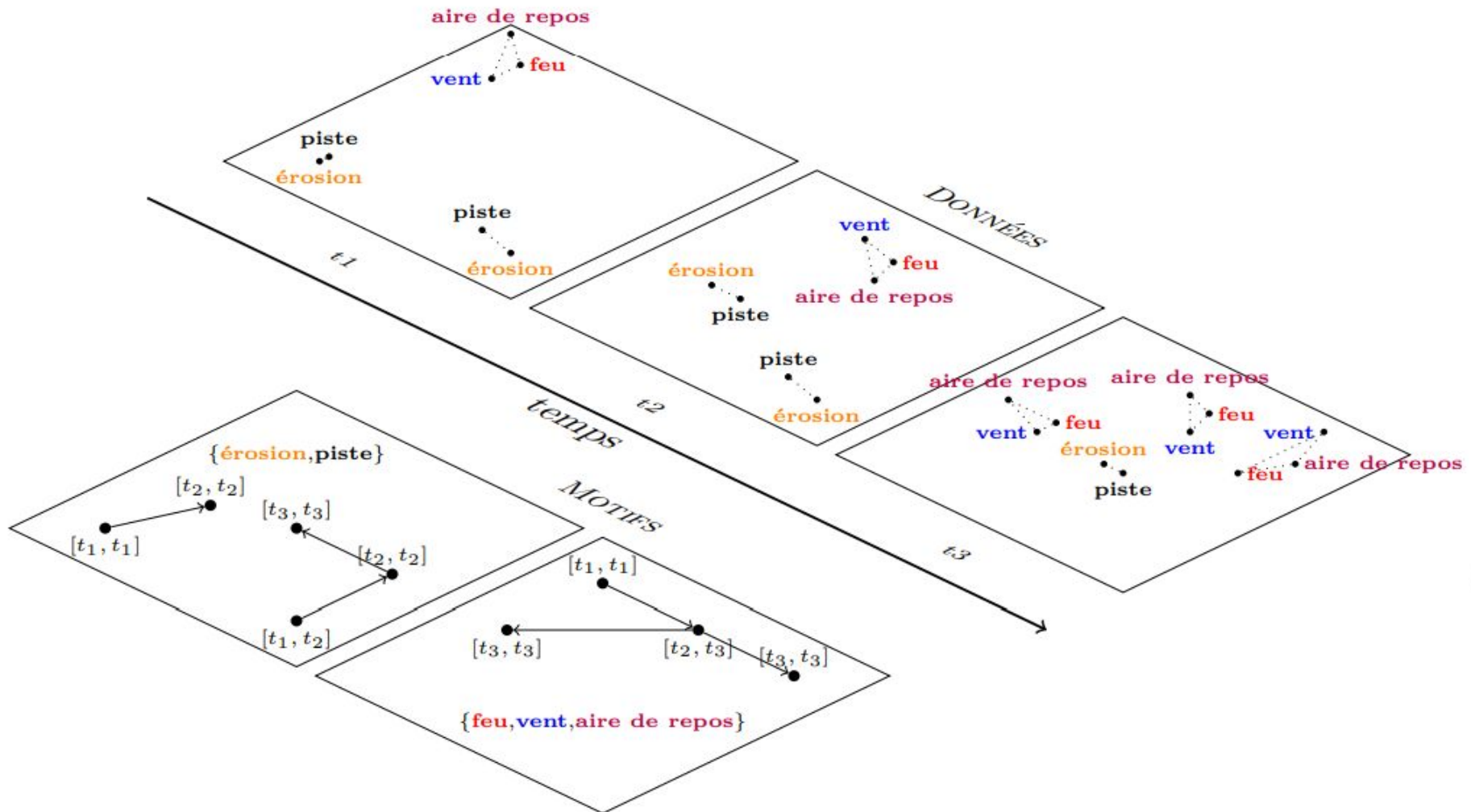
Représentation des données spatio-temporelles

14

- Avec la diffusion des outils de localisation (**tel que le GPS**) et l'évolution de la société, la représentation des données demande l'inclusion d'une **nouvelle dimension: le temps**. Ceci pour répondre à une multitude des questions comme par exemple :
 - la **vitesse d'éloignement** (ou de rapprochement) entre les objets,
 - l'historique des rencontres,
 - la **localisation des objets** à un moment précis, les données attributaires (qui évoluent, qui changent)...
- Ces données sont **difficiles à manipuler** dans les systèmes classiques de gestion de données, qui ne sont pas outillés pour les données de dimension supérieure à 1.
- Des travaux se font pour intégrer la **dimension « temps »** dans une base de données mais la représentation graphique reste timide.

Exemple de co-localisations spatio-temporelles

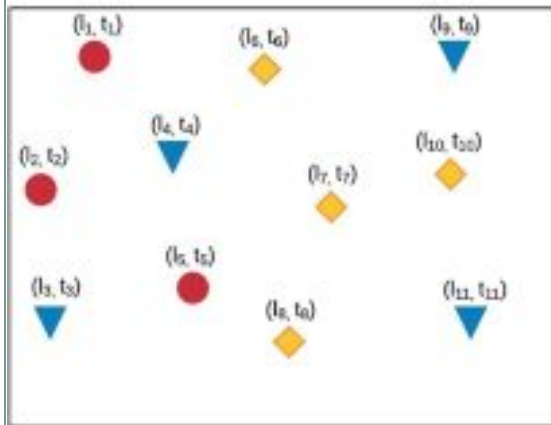
15



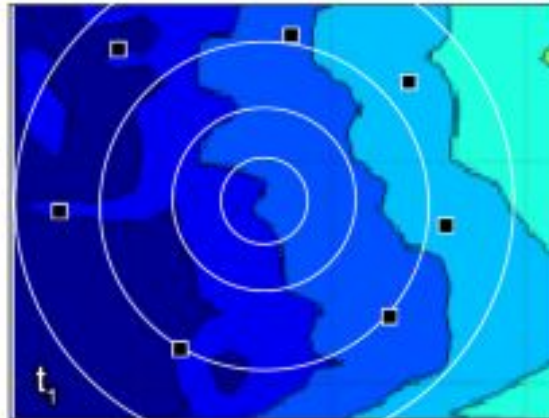
Types de données spatio-temporelles

16

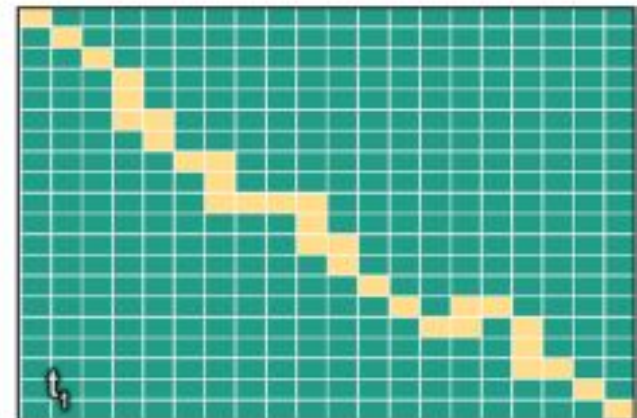
a



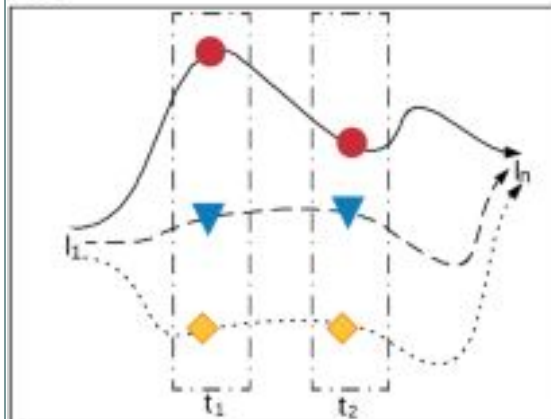
b



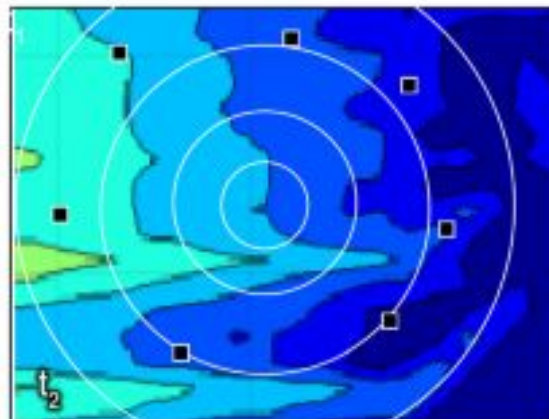
c



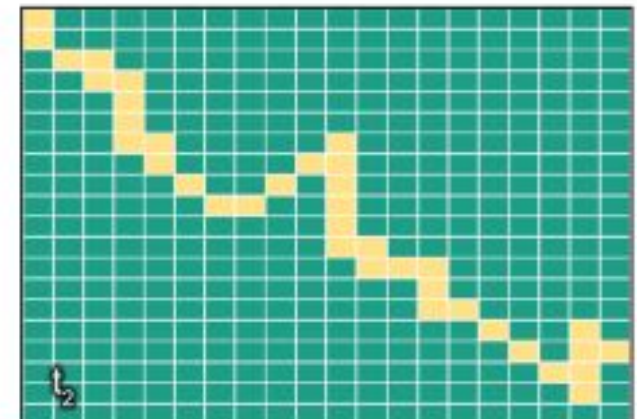
d



e



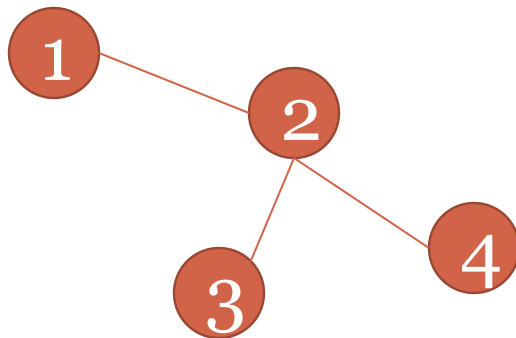
f



Matrice et graphe de voisinage

17

- Les **relations spatiales** sont communément formalisées par la notion de **graphe de voisinage**, et peuvent être **représentées** sous forme d'une **matrice de voisinage**.
- La **matrice de voisinage** est une **matrice binaire** M où $M[i,j]=1$ si l'objet i est **voisin** de l'objet j et $M[i,j]=0$ dans le cas **inverse**.



$$M = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

- La notion de **voisinage** est générale et peut aussi bien représenter une contiguïté entre **formes zonales** ou une **proximité sur des points**. Elle peut être pondérée en qualifiant la **proximité par une distance**.

Parallèle relationnel – Spatial

18

- Cette comparaison montre que :
 - les SGBD Géographiques sont spécifiques,
 - mais ils peuvent être vus comme une extension des SGBD relationnels

	RELATIONNEL	SPATIAL
Données	Entier, Réel, Texte, ...	Plus complexes: Point, Ligne, Région
Prédicats et calculs	Tests : =, >, ... Calculs : +, /, ... et fonctions simples	Prédicats et calculs géom. et topologiques: Tests : intersecte, adjacent à, ... Fonctions géom. : intersection, surface...
Manipulation	Opérateurs de l'algèbre : Sélection, Projection, Jointure... Agréats: Count, Sum, Avg...	Manipulation par thème ou inter-thèmes Sélection et jointure sur critère spatial Agréats : fusion d'objets adjacents
Liens entre objets	Par clés de jointures	Liens spatiaux (souvent) implicites
Méthodes d'accès	Index B-tree, hachage	Index R-tree, quad-tree, etc.

Introduction: Pourquoi le Data Mining?



❑ L'explosion des données

Les outils de collecte automatique des données et les bases de données conduisent à d'énormes masses de données stockées dans des entrepôts:

- Entrepôts du Web : ex. Google,
- Réseaux sociaux et hébergement de documents : ex. Facebook, gmail...
- e-commerce Achats dans les supermarchés,
- Transactions de cartes bancaires

❑ Les données sont collectées et stockées rapidement (GB/heures):

- Capteurs : RFID, supervision de procédé,
- Télescopes,
- Puces à ADN générant des expressions de gènes,
- Simulations générant de téra-octets de données.

❑ Submergés par les données, manque de connaissance !

Définition: KDD vs DM



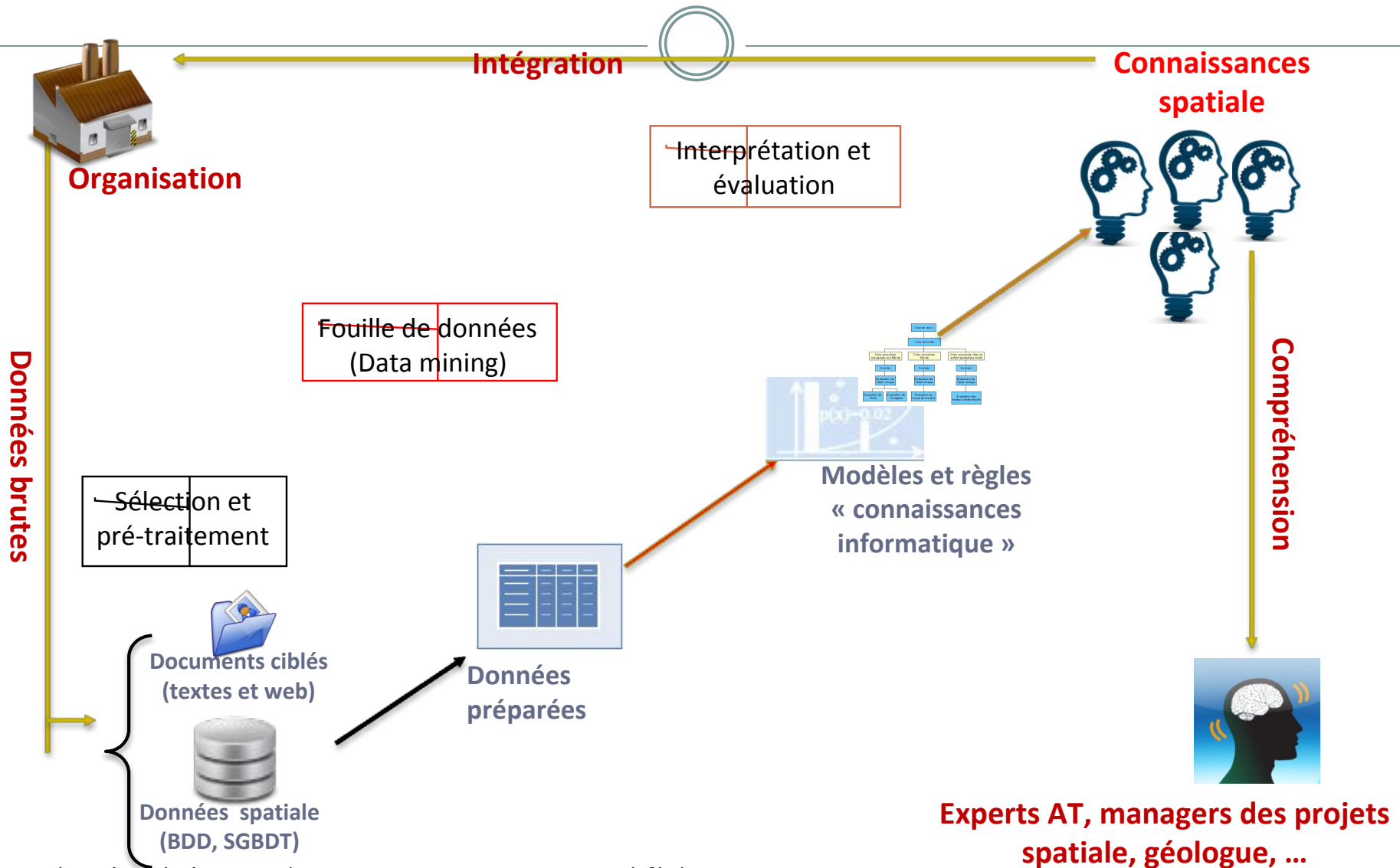
- **Data Mining**: ensemble des techniques et méthodes destinées à l'exploration et l'analyse de grandes bases de données informatiques en vue de **détecter** dans ces données des **règles**, des **associations**, des **structures** pour en extraire l'essentiel de l'information utile dans l'objectif est **l'aide à la décision**.
- Le principe du **Data Mining** consiste à rechercher des **structures** reliant ces données. Cette recherche peut se faire de **manière automatique**, en mettant en œuvre, par exemple, des **algorithmes** dont le but est de trouver des **règles d'association**.

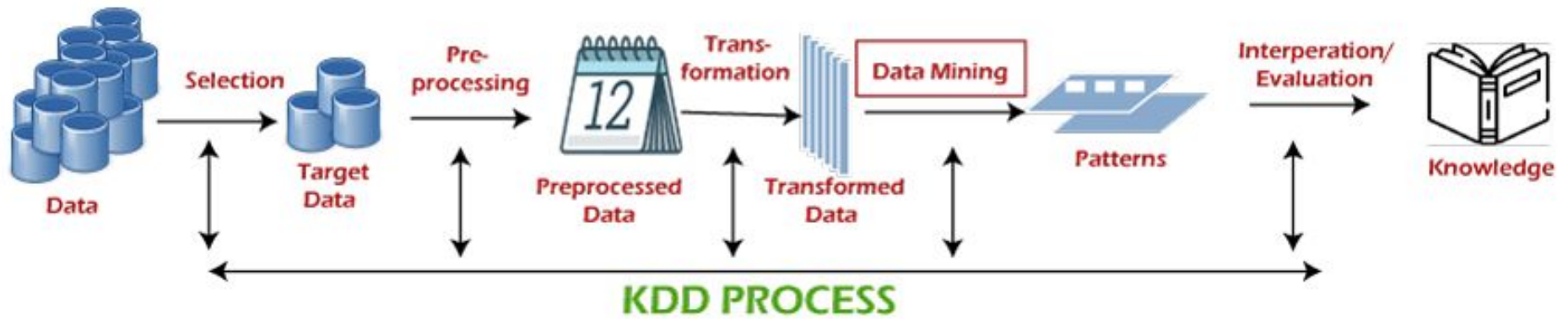
En résumé : c'est quoi le Data Mining ??

- **Découverte d'informations intéressantes dans un paquet de données.**
 - ❖ qu'est-ce qu'un paquet de **données** ?
 - ❖ qu'est-ce qu'une **information intéressante** ?
 - ❖ qu'entend-on par **découvrir** ?

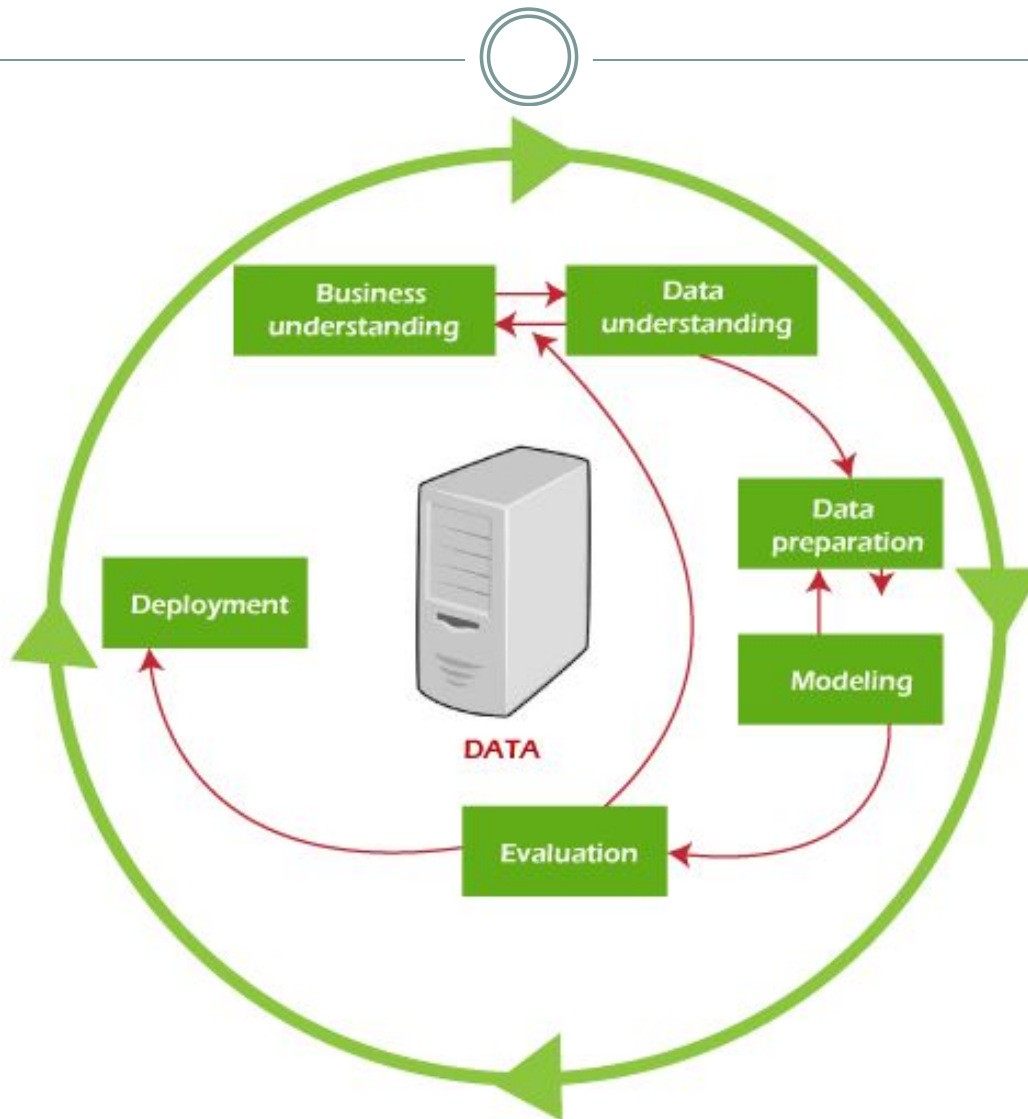
En anglais : **Data Mining**: Fortement lié à l'apprentissage automatique (**machine learning**).

Le processus de Data Mining



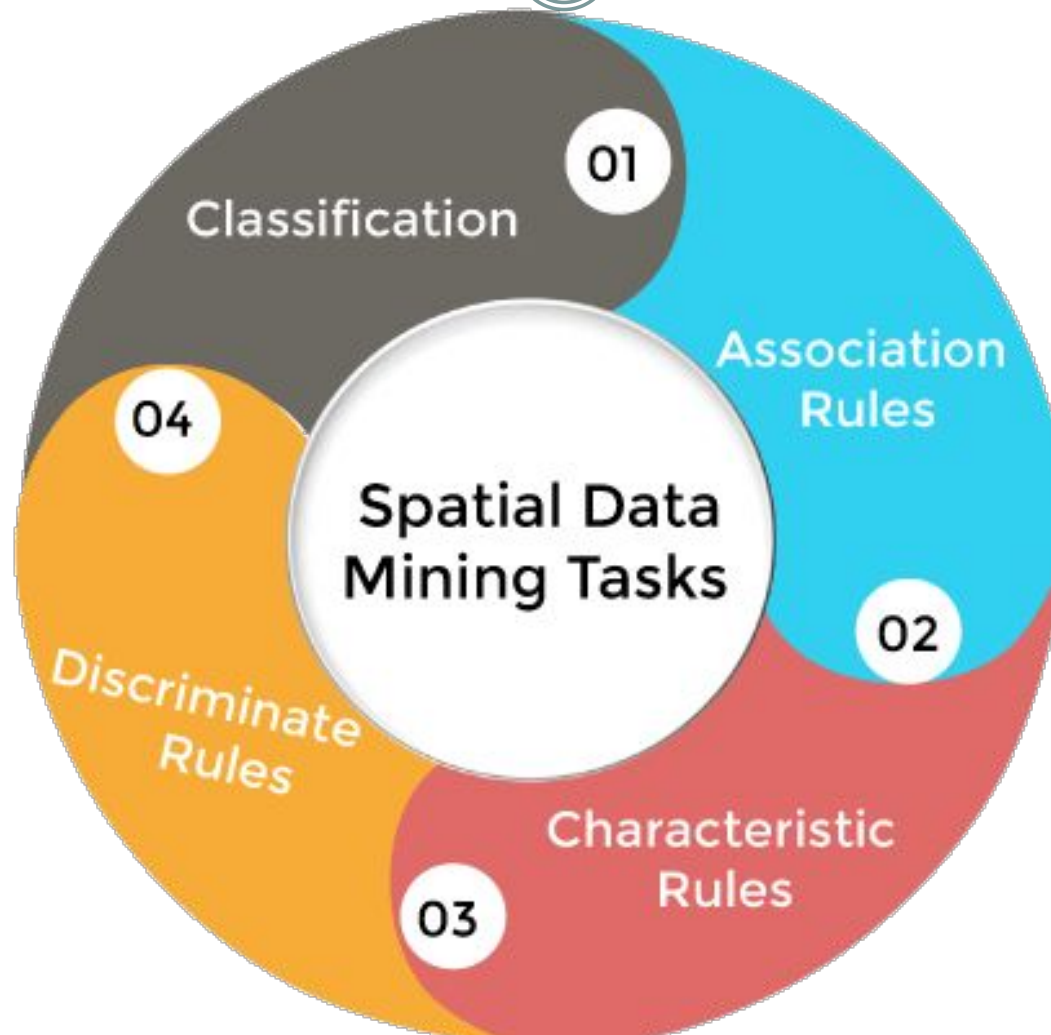


Les étapes de la méthode CRISP-DM



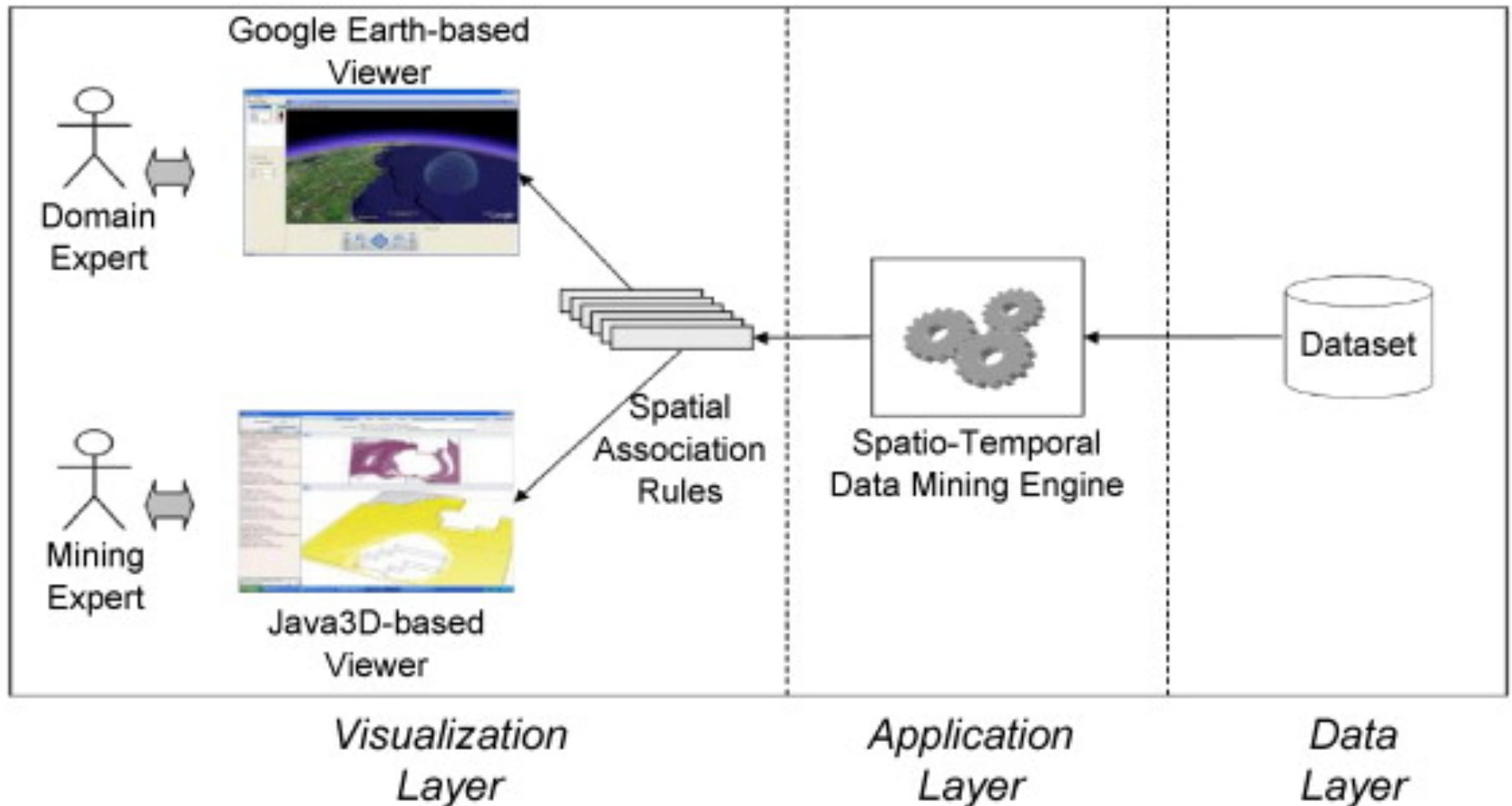
Les tâches du Spatial data mining

24



Exemple de Data Mining Spatial

25



Les algorithmes d'apprentissage

26

- Les algorithmes d'apprentissage peuvent se catégoriser selon le type d'apprentissage qu'ils emploient.

Algorithmes permettant de classer, regrouper des profils homogènes à partir de données non étiquetées.
Ex. détection d'anomalies

Apprentissage non supervisé

Algorithmes permettant d'adapter sa stratégie en temps réel en fonction d'un signal positif ou négatif.
Ex. jeux, robots

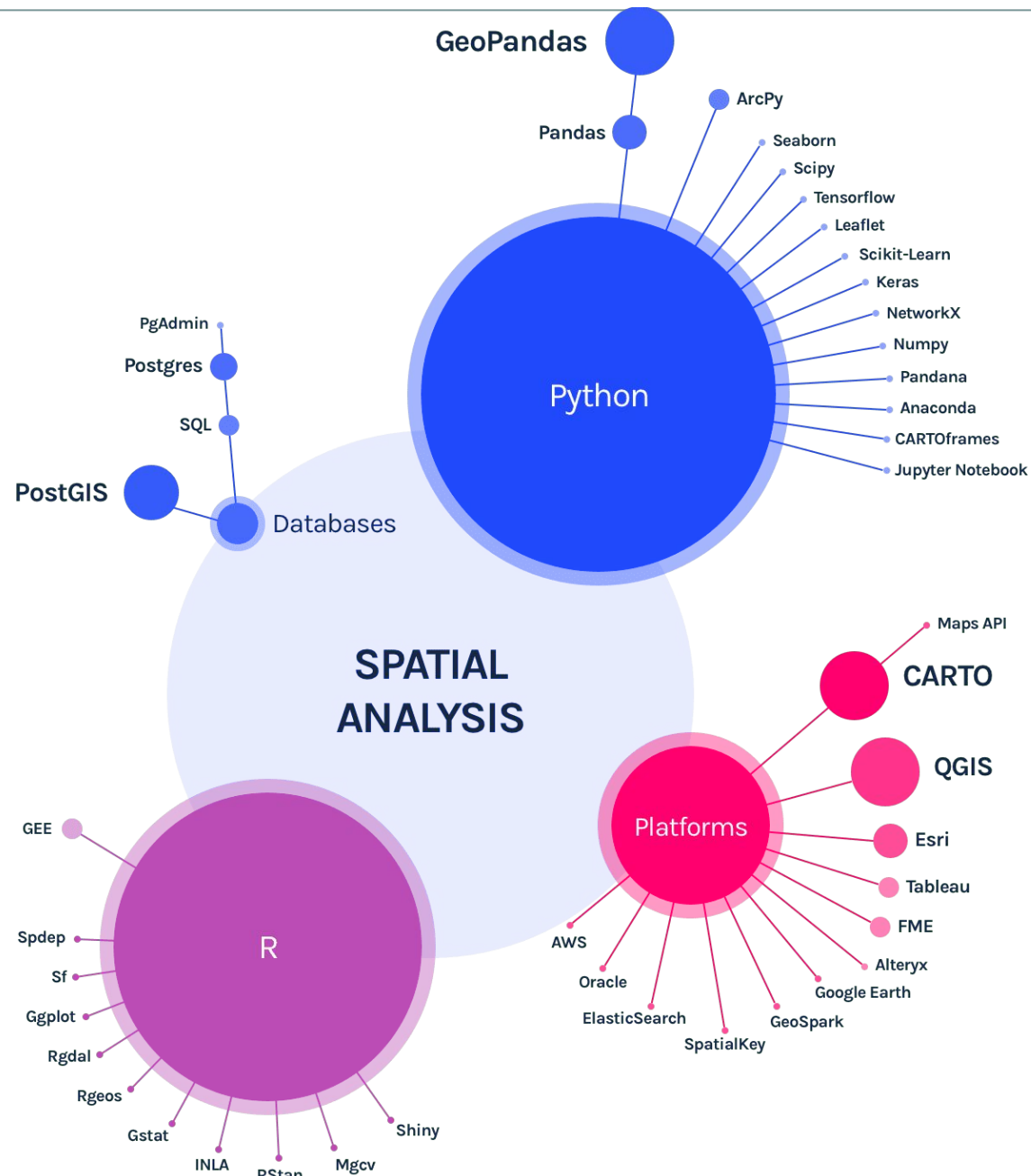
Apprentissage par renforcement

Machine Learning

Algorithmes permettant de prévoir ou classer à partir de données historiques et étiquetées.
Ex. la prévision de l'attrition ou d'incident

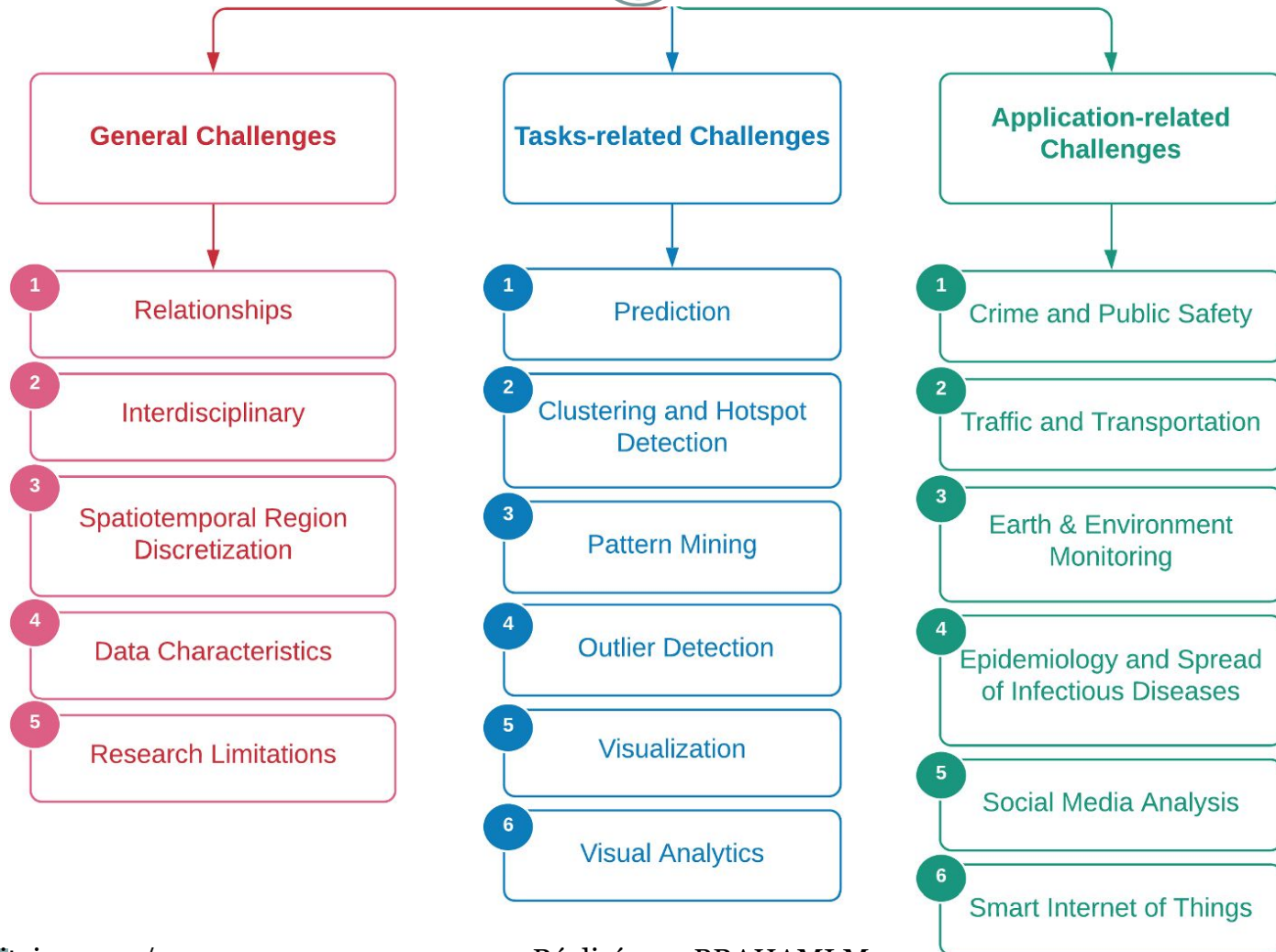
Apprentissage Supervisé

Data Science tools for spatial analysis



Une taxonomie de la structure de défis STDM

28



Logiciels de fouilles de données spatiales

29

- Geominer
- Descartes
- Fuzzy Spatial OQL for Fuzzy KDD
- GWiM
- GeoKD
- S+SpatialStats
- Modules spécialisés dans R
- Spatial Statistics Toolbox for Matlab/Fortran :
- NEM

**Merci de votre attention
Thanks!**

30

**JE VOUS SOUHAITE
BONNE
CONTINUATION...**