

# Rencontre de la Communauté Pentaho

## Extension spatiale pour Pentaho Data Integration



**Puissance  
et agilité**  
pour votre  
système d'information

 <b>Communication digitale</b> Navigateurs et smartphones pour diffuser vos messages	 <b>Dématérialisation</b> Capitalisez sur vos processus GED, BPM Collaboratif Recherche	 <b>Business Intelligence</b> Prenez les bonnes décisions Big data Infocentre Géo-décisionnel	 <b>Système d'information géographique</b> Ajoutez une dimension géographique à vos applications	 <b>Applications Web et mobiles</b> Vos solutions métiers sur mesure
---	---	---	---	---

C L O U D   C O M P U T I N G   I N F O G É R A N C E

# Extension spatiale pour PDI

L'information géographique... une définition



« L'information géographique est la **représentation** d'un **objet** où d'un **phénomène réel** ou imaginaire, présent, passé ou futur, localisé dans **l'espace** à un moment donné et quelles qu'en soient la **dimension** et **l'échelle** de représentation »

*Source : Wikipedia*

# Extension spatiale pour PDI

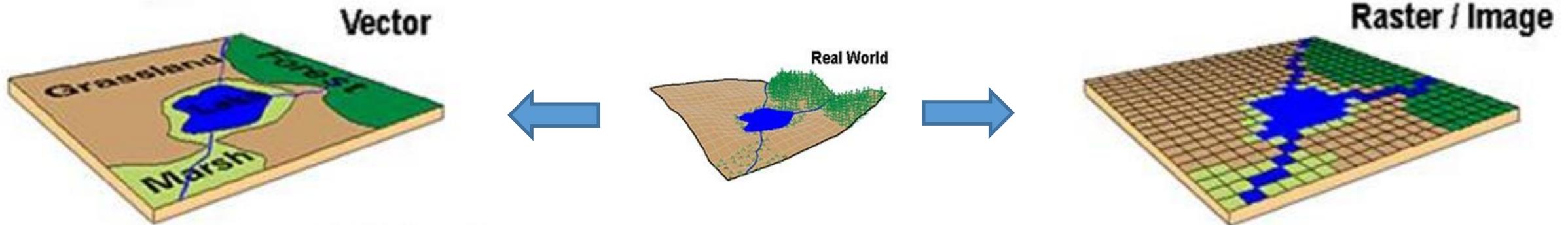
L'information géographique dans les systèmes informatiques

## L'information vectorielle

- Les objets ou phénomènes sont décrits par une liste de coordonnées géographiques

## L'information matricielle

- Les objets ou phénomènes sont décrits au sein d'une grille de valeurs



# Extension spatiale pour PDI

L'information géographique dans PDI... Pourquoi une extension ?

Faciliter l'exploitation des données entre  
2 familles d'outils

- Les systèmes d'informations géographiques

- Les outils décisionnels

Cela existe déjà ! ...

- GeoKettle / Spatialytics ETL

- FME : ETL propriétaire



- Spatial extension for Talenc

ECM, GED, RSE, BPM

WEB SIG



PORTAIL, CMS,  
E-COMMERCE

BUSINESS  
INTELLIGENCE

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Quel périmètre ?

- **La ligne de conduite... que nous somme fixée**
  - Une extension pour Pentaho Data Integration et non un fork
  - Ne pas chercher à concurrencer les outils spécifiques
  - Se limiter, dans un premier temps, à la gestion de l'information géographique vectorielle
  - Du full Java
- **Les besoins minimaux à satisfaire**
  - Lire et écrire les formats « courants »
  - Supporter les extensions spatiales des SGBD les + répandus
  - Prendre en compte les systèmes de projection
  - Disposer d'un minimum de fonctionnalités spécifiques à la manipulation de l'information géographique

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

Les impacts de l'extension

## Pour le cœur...

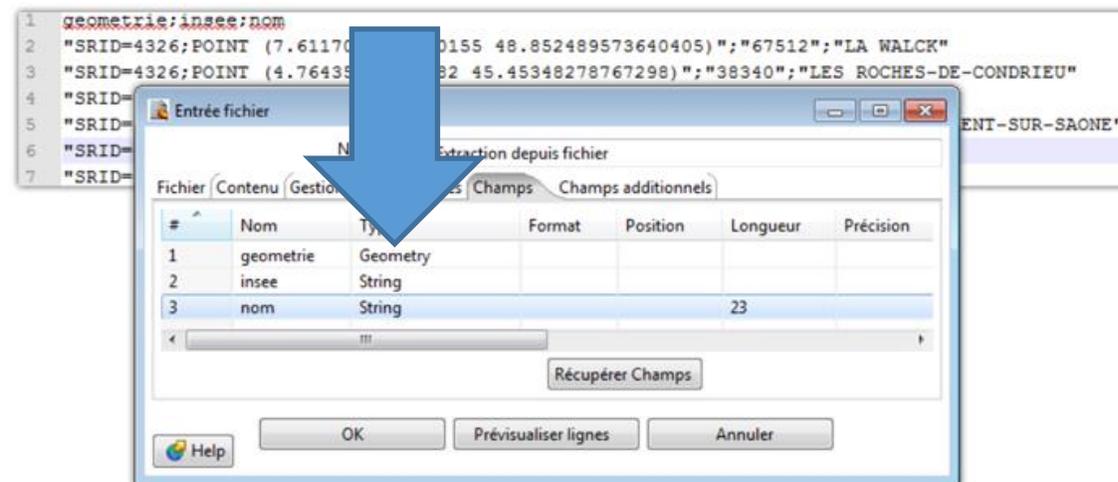
Ajout d'un type de donnée « Geometry » dédiée au stockage et à la manipulation de l'information géographique vectorielle



« ValueMetaGeometry » basée sur la librairie Java Topology Suite

## Pour les yeux...

Des briques spécifiques regroupées au sein d'une catégorie unique pour les transformations



-  Agrégation de géométries
-  Alimentation fichier SIG
-  Extraction depuis fichier SIG
-  Géotraitement
-  Information sur géométrie
-  Opération avec système de coordonnées
-  Relation spatiale et proximité

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

Les impacts de l'extension

- **Types de géométries reconnus**
  - Point et MultiPoint
  - LineString & MultiLineString
  - Polygon & MultiPolygon
- **En 2D ou en 3D... mais exploitation de la 2D uniquement pour :**
  - Les prédicats spatiaux (Intersecte, Est contenu dans, ...)
  - Les opérations sur les géométries (Union, Intersection, Difference, ...)



Point



Multipoint with 4 parts



Simple non-closed  
linestring



Simple multilinestring defined  
by 4 endpoints of 2 elements



Polygon defined  
by exterior ring



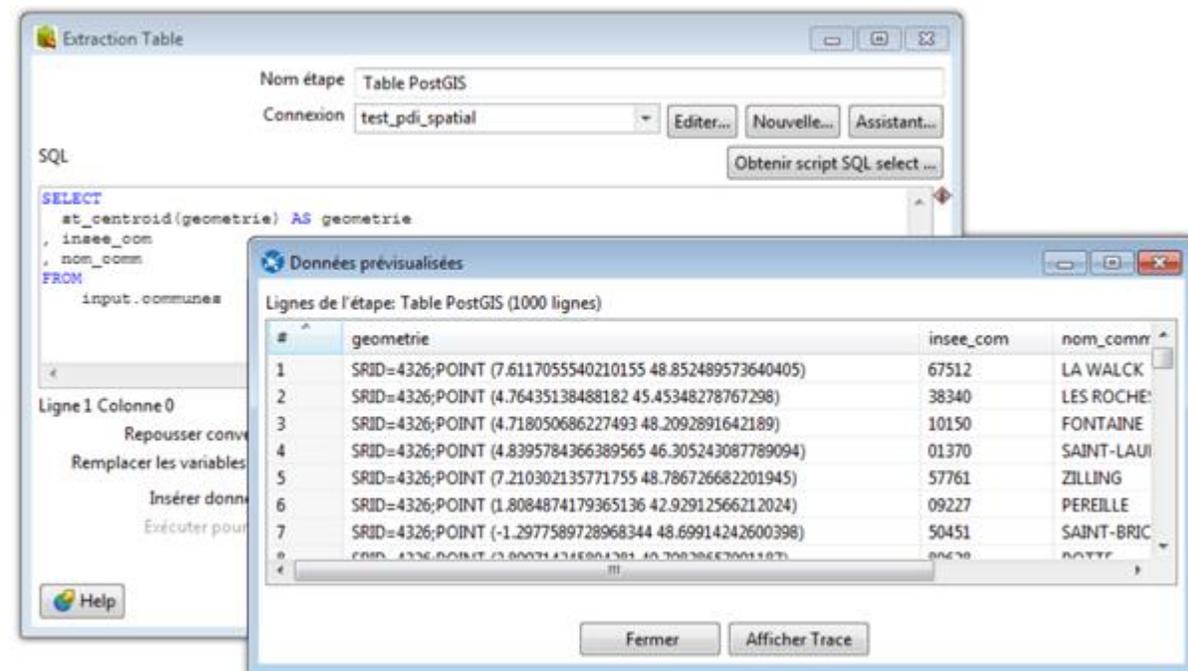
Multipolygon consisting  
of 2 elements defined  
by exterior rings and 3 interior rings

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

Les impacts de l'extension sur les **briques existantes**

- **Lecture et écriture depuis et vers les bases de données**
  - « Extraction depuis table »
  - « Insertion dans table »
  - « Insertion / Mise à jour table »
  - « Mise à jour table »
  - « Jointure Base de données »
  - « Recherche dans Base de données »
  - « Interrogation dynamique d'une BdD »
- **Brique statistique « dupliquée » et « enrichie »**



# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

Les briques **spécifiques**



## Lecture de fichiers SIG

Pouvoir lire quelques formats de fichiers couramment exploités par les SIG bureautiques

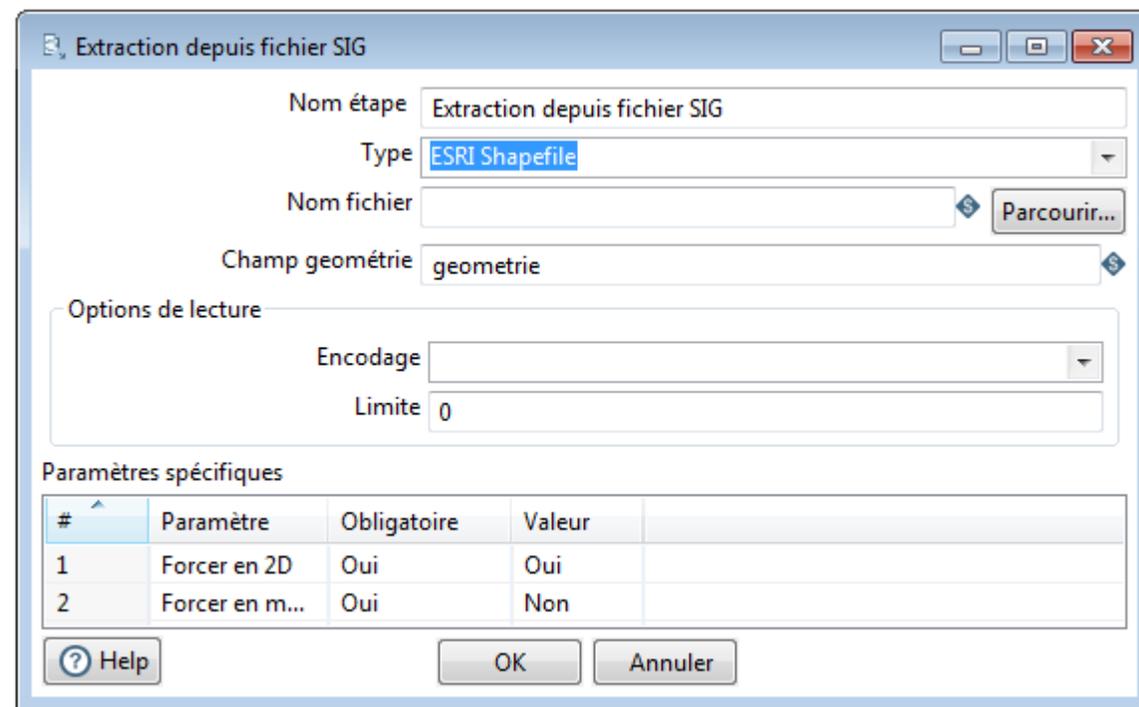
ou le « web carto »

**Actuellement**

Drawing eXchange Format  
**ESRI Shapefile**  
GeoJSON  
MapInfo Interchange Format  
Spatialite / SQLite

**Dans les tuyaux**

GPX,  
GeoPackage



The screenshot shows a dialog box titled 'Extraction depuis fichier SIG'. It contains the following fields and options:

- Nom étape: Extraction depuis fichier SIG
- Type: ESRI Shapefile
- Nom fichier: (empty field) with a 'Parcourir...' button.
- Champ géométrie: geometrie
- Options de lecture:
  - Encodage: (empty dropdown)
  - Limite: 0
- Paramètres spécifiques table:

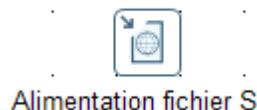
#	Paramètre	Obligatoire	Valeur
1	Forcer en 2D	Oui	Oui
2	Forcer en m...	Oui	Non

Buttons: Help, OK, Annuler

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Les briques spécifiques



### écriture de fichiers SIG

Pouvoir produire quelques formats de fichiers couramment exploités par les SIG bureautiques

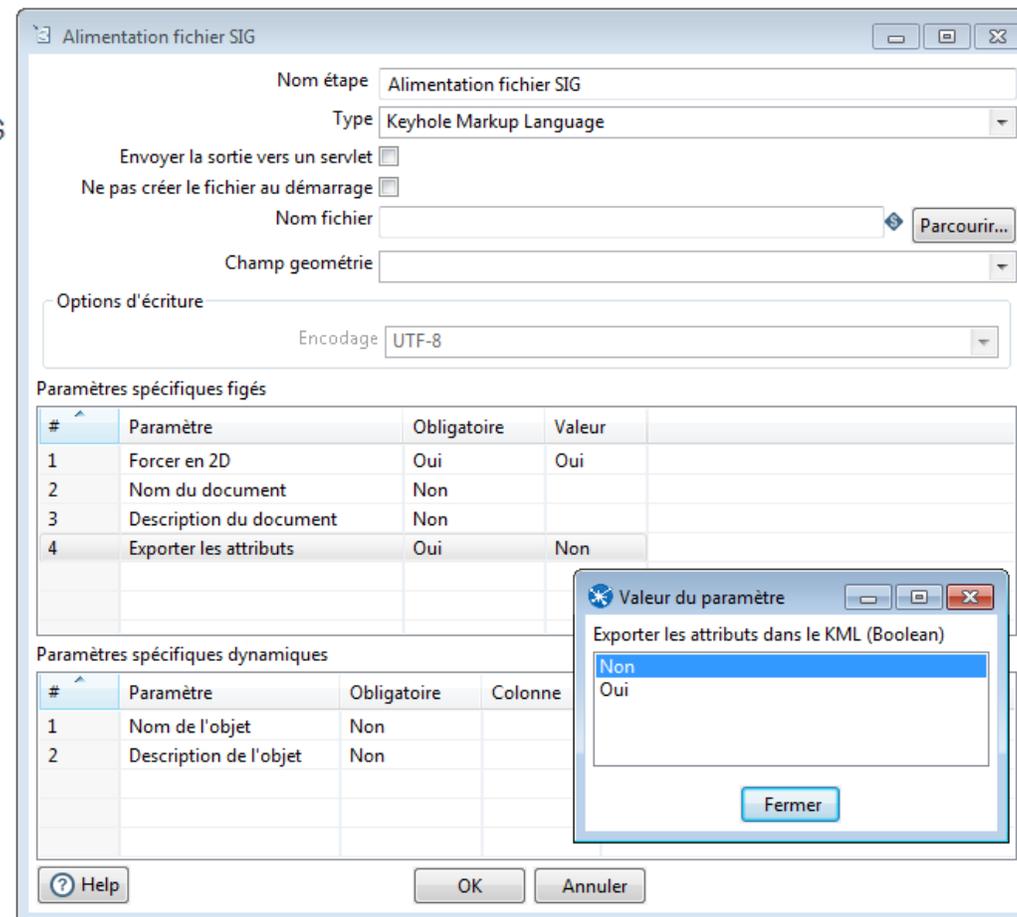
ou le « web carto »

**Actuellement**

Drawing eXchange Format  
ESRI Shapefile  
GeoJSON  
Keyhole Markup Language

**Dans les tuyaux**

GPX,  
GeoPackage  
Scalable Vector Graphics



# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Les briques **spécifiques**



Information sur géométrie

### Informations sur géométrie

Extraire certaines caractéristiques propres  
à une géométrie

*Attention : valeurs exprimées dans les unités du  
système de coordonnées*

(/!\ pour les géométries dont les  
coordonnées sont en DD ou DMS)

#	Information	Colonne
1	Dimension (2D/3D)	dimension
2	Longueur	perimetre_en_m
3	Surface	surface_en_m2
4	Type de géométrie	type_geometrie
5	Nombre de géométries	nombre_sous_geometries
6	Nombre de sommets	nombre_de_sommets
7	X minimal	x_min
8	Y minimal	y_min
9	X maximal	x_max
10	Y maximal	ymax

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Les briques spécifiques

### Prise en compte des systèmes de coordonnées

- *Assigner un système à une géométrie*
- *Transformer une géométrie d'un système vers un autre*

*Attention : Seuls les systèmes référencés dans la base « EPSG » peuvent être assignés*

*Prise en compte des « autorités » IGNF et ESRI pour la reprojection*

The image displays three screenshots of a software interface for handling coordinate systems. The top screenshot shows a list of coordinate systems to be assigned, with columns for year, name, and system. The middle screenshot shows the 'Opération avec système de coordonnées' dialog box with fields for 'Nom étape', 'Type d'opération', 'Champ géométrie', 'Géométrie', 'Système de coordonnées à assigner', and 'Système de coordonnées en sortie'. The bottom screenshot shows the same dialog box with specific values for 'Autorité' (ignf) and 'Code' (LAMB93).

Année	Nom	Système
2000	Anguilla 1957	British West Indies Grid
20004	Pulkovo 1995	Gauss-Kruger zone 4
20005	Pulkovo 1995	Gauss-Kruger zone 5
20006	Pulkovo 1995	Gauss-Kruger zone 6
20007	Pulkovo 1995	Gauss-Kruger zone 7
20008	Pulkovo 1995	Gauss-Kruger zone 8
20009	Pulkovo 1995	Gauss-Kruger zone 9
2001	Antigua 1943	British West Indies Grid
20010	Pulkovo 1995	Gauss-Kruger zone 10
20011	Pulkovo 1995	Gauss-Kruger zone 11
20012	Pulkovo 1995	Gauss-Kruger zone 12
20013	Pulkovo 1995	Gauss-Kruger zone 13
20014	Pulkovo 1995	Gauss-Kruger zone 14
20015	Pulkovo 1995	Gauss-Kruger zone 15
20016	Pulkovo 1995	Gauss-Kruger zone 16
20017	Pulkovo 1995	Gauss-Kruger zone 17
20018	Pulkovo 1995	Gauss-Kruger zone 18
20019	Pulkovo 1995	Gauss-Kruger zone 19
2002	Dominica 1945	British West Indies Grid
20020	Pulkovo 1995	Gauss-Kruger zone 20

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour  
Les briques **spécifiques**

## Les prédicats spatiaux

- *Connaître / vérifier les relations entre les géométries*

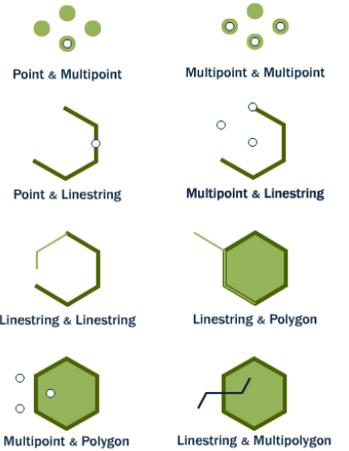
*Actuellement*

Nécessité de réaliser un produit cartésien

*Dans les tuyaux*

*Nouvelle brique « Jointure spatiale » basée sur le moteur SQLite permettant une indexation RTree (mode ELT)*

Intersects



Relation spatiale et proximité

Contient  
Couvre  
Distance maximale  
Distance minimale  
Est contenu dans  
Est couvert par  
Est disjoint de  
Est rigoureusement égal à  
Est à une distance inférieure à  
Est à une distance supérieure à  
Est égal à  
Intersecte  
Se superpose à  
Touche  
Traverse

Relation spatiale et proximité

Nom étape: Relation spatiale et proximité

Relation: Contient

La géométrie A contient la géométrie B

Paramètres en entrée

Géométrie A: [dropdown]

Géométrie B: [dropdown]

Distance dynamique?

Distance: [input] [input]

Paramètres du résultat

Filtrage de lignes: Retourner uniquement les lignes en succès

Résultat: Ne pas filtrer, Retourner uniquement les lignes en succès, Retourner uniquement les lignes en échec

Help

Jointure spatiale entre deux flux

Nom étape: Jointure spatiale

Type de jointure: Contient

La première géométrie contient la seconde géométrie

Première étape: [dropdown]

Première Géométrie: [dropdown]

Seconde étape: [dropdown]

Seconde Géométrie: [dropdown]

Help OK Annuler



Jointure spatiale

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Les briques spécifiques

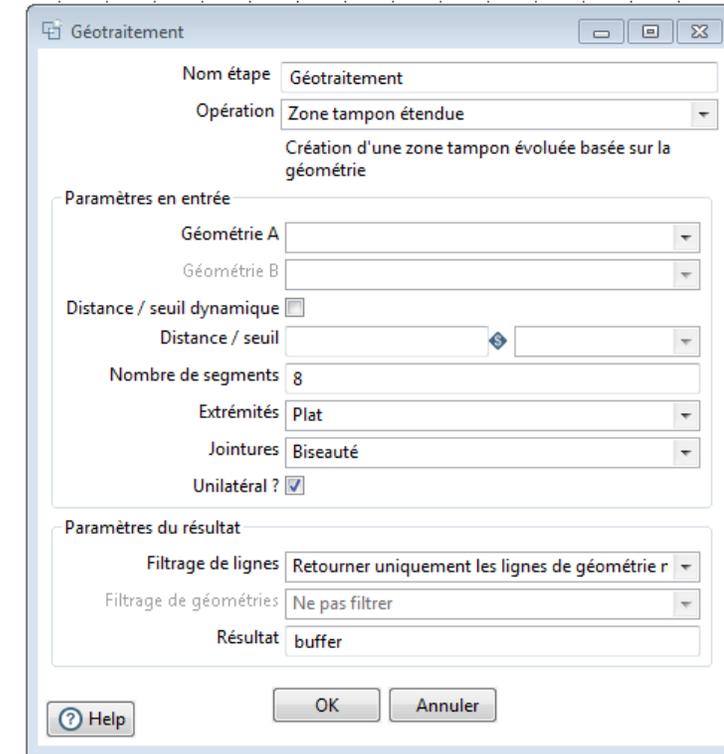
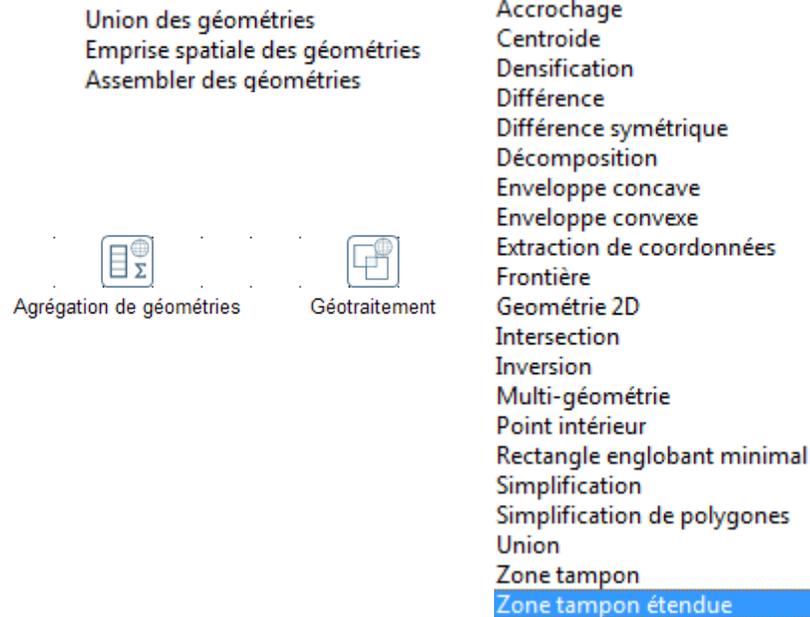
### Les géotraitements

- *Transformer et agréger les géométries*

#### Dans les tuyaux

- *Création de polygones à partir d'un réseau de lignes*
- *Dissolution de frontières*
- *Simplification par la méthode de Visvalingam-Whyatt*
- *Abscisse curviligne*

### Actuellement



# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

## Les limites

- Pas de possibilité d'associer des métadonnées au nouveau type « Geometry »
- *Gourmand en mémoire sur les géométries complexes (modifier les paramètres de Spoon)*
- *Pas de possibilité d'indexation*
- *Adaptations à réaliser pour version 6.1 (bibliothèques manquantes)*
- *Quelques bugs (ex : tri dans un prévisu)*

# Extension spatiale pour PDI

L'extension spatiale « Information géographique » pour PDI

Pour en savoir plus / télécharger

- Blog Atol CD :  
<https://blog.atolcd.com/index.php/2015/06/26/une-extension-gis-dans-pentaho-data-integration-5>
- *Github Atol CD* :  
<https://github.com/atolcd/pentaho-gis-plugins>

pentaho-gis-plugins	Fixes #2 Coordinate System Operation Step Opening break	7 months ago
shp	Première version publique des plugins	11 months ago
spatialite	Première version publique des plugins	11 months ago
.gitignore	Première version publique des plugins	11 months ago
README.md	Add contributing informations and links to blog and videos	11 months ago
pom.xml	Première version publique des plugins	11 months ago

**Pentaho Data Integrator GIS Plugins and modified dependencies**

The target project is pentaho-gis-plugins. All its dependencies are available in public maven repositories or in other sub projects of this one, which have been forked from opensource projects.

## Une extension « GIS » dans Pentaho Data Integration 5

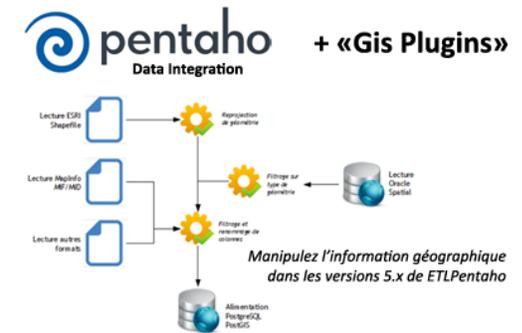
par CÉDRIC DARBON le 26 JUIN 2015 · 1 COMMENTAIRE

Pentaho Data Integration

a toujours autorisé la création de « plugins » permettant d'étendre les fonctionnalités de cet ETL en ajoutant notamment de nouvelles « JobEntries » pour les « Jobs » ou de nouveaux « Steps » pour les « Transformations ».

Cette possibilité offerte à

l'utilisateur permettait de couvrir nombre de besoins mais il fallait se limiter à la



Merci pour votre écoute