

# **Inventaire automatique de données et cartes**

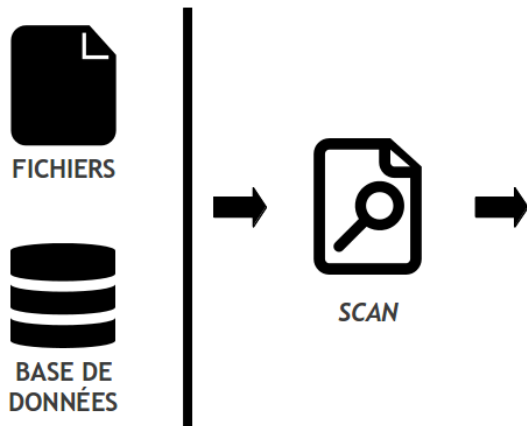
## Présentateurs d'aujourd'hui

- François PRUNAYRE (titellus),
- Etienne TAFFOUREAU (BRGM),
- Yves JACOLIN (Camptocamp)

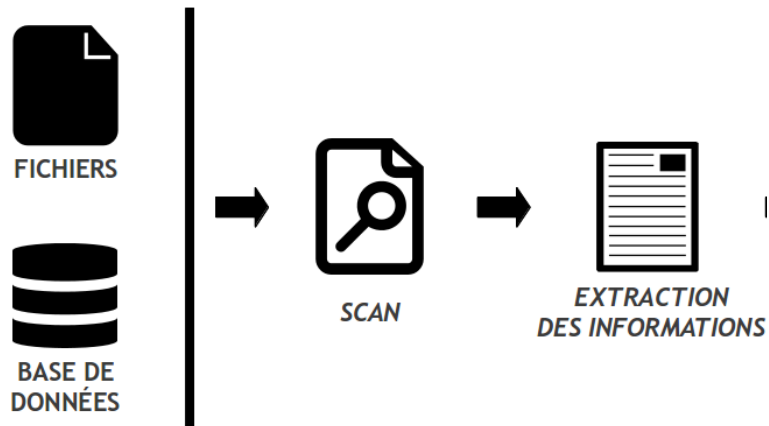
## et contributeurs d'autrefois

- Mathieu Coudert (Astrium GEO-Information Services)
- Bruno Lafage ([Parc Nationaux de France](#))
- Paul Hasenohr (EEA)

Déployer un “agent” capable de  
rechercher dans les sources d’  
informations.



# Extraire les métadonnées





FICHIERS



BASE DE  
DONNÉES



SCAN



EXTRACTION  
DES INFORMATIONS



RÉFÉRENCEMENT



PUBLICATION

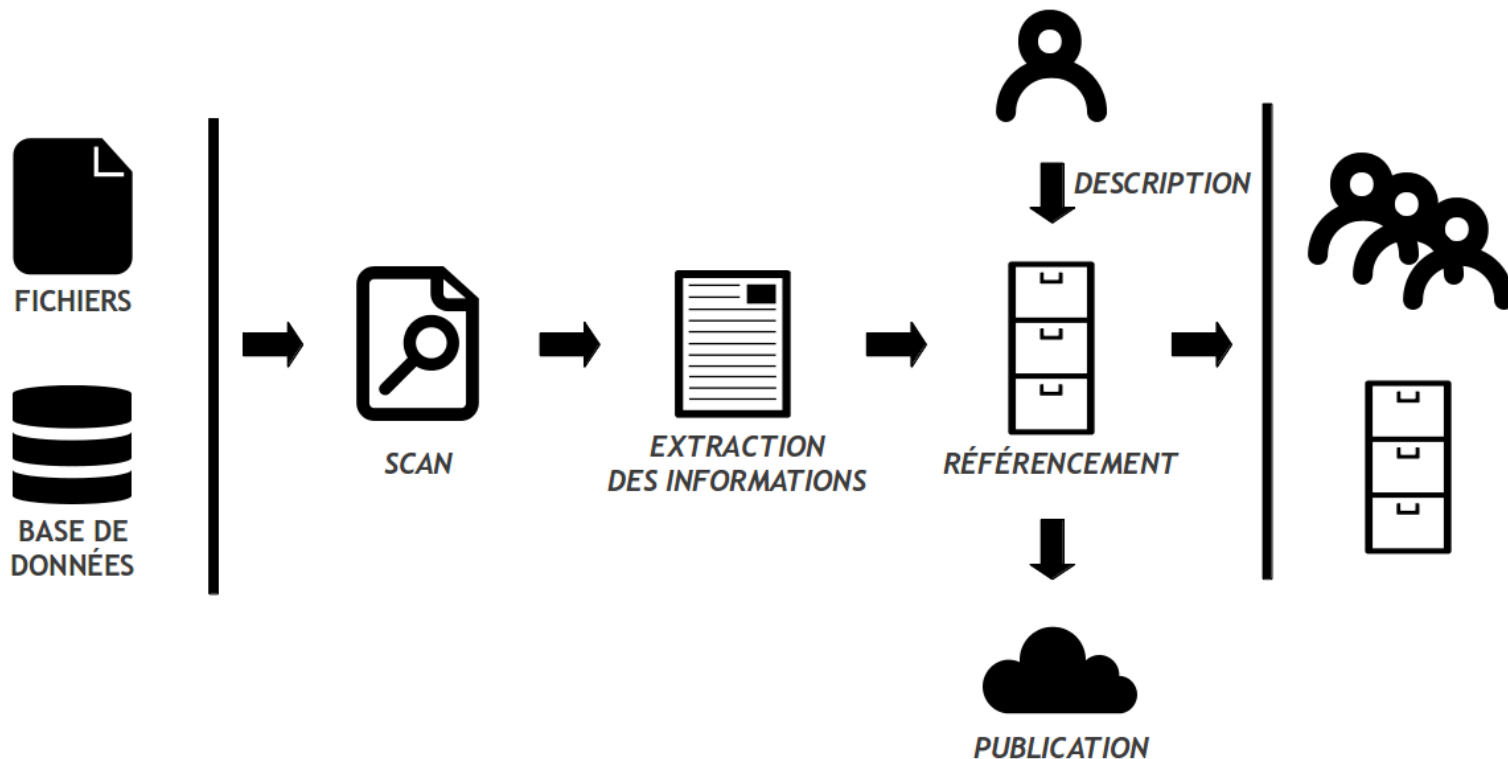


Alimenter un  
catalogue



DESCRIPTION

# Le principe



# Origines

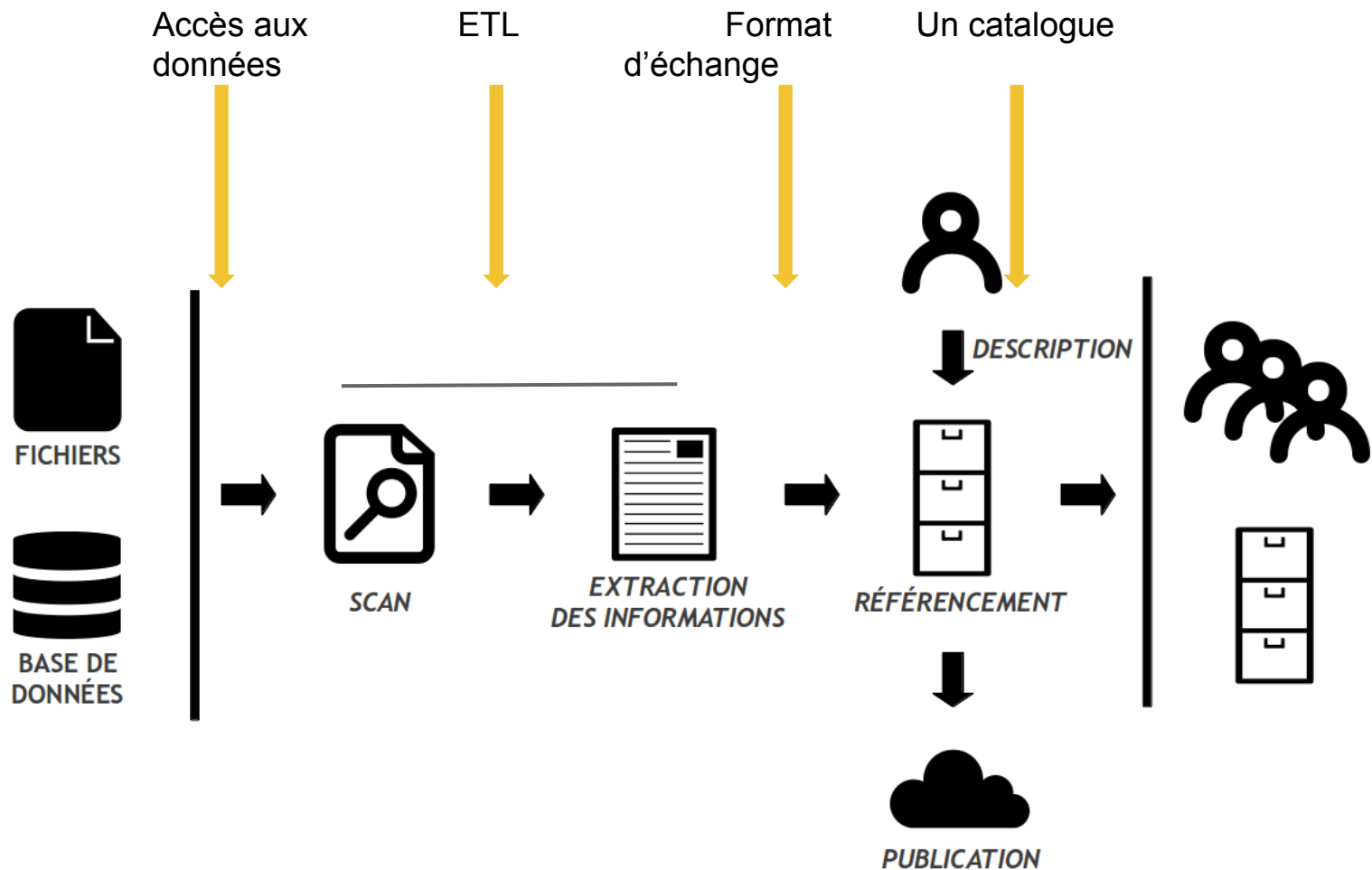
UNHCR : Catalogues de cartes

OneGeology : Moissonnage de services OGC

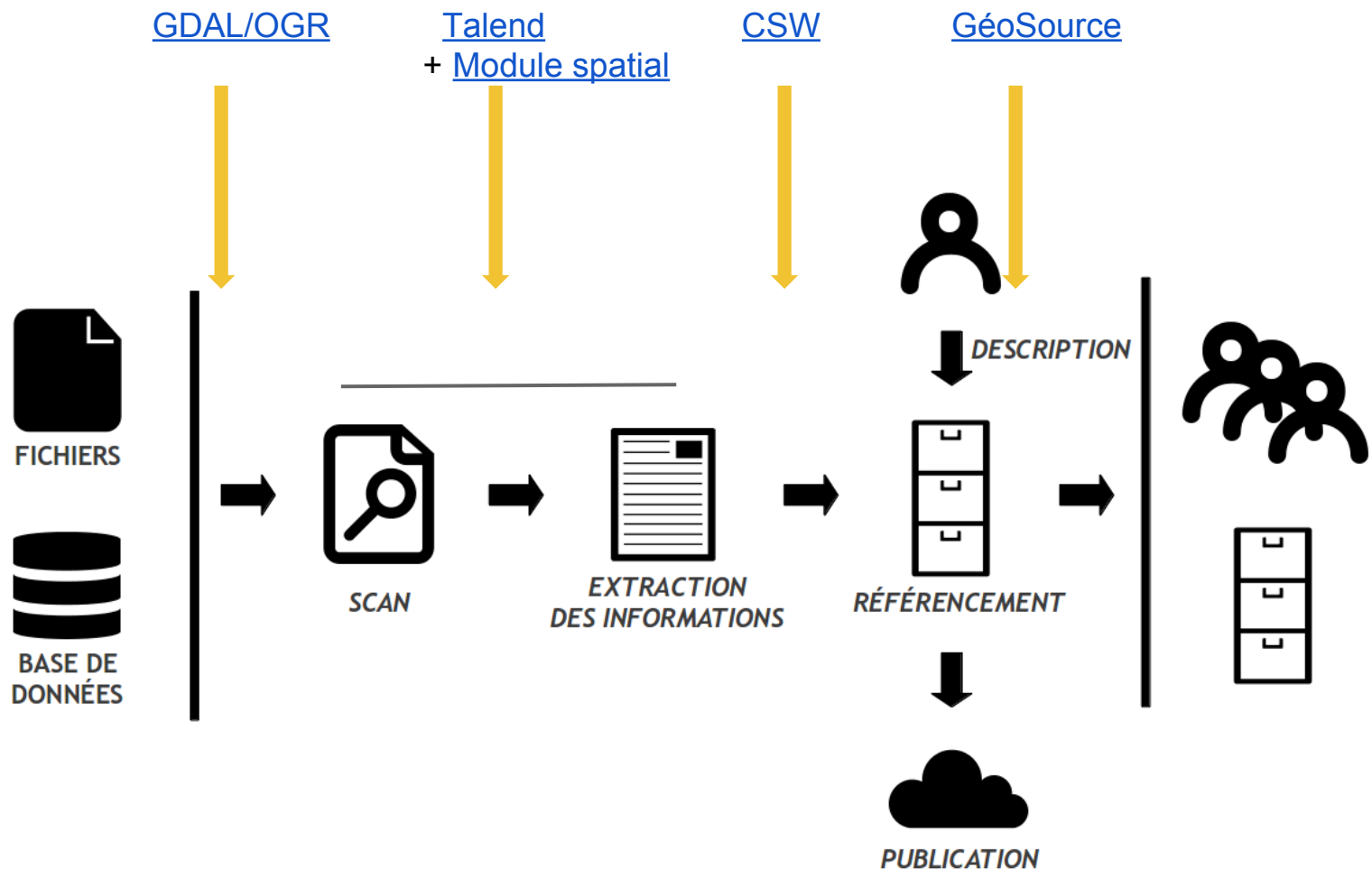
EEA : Analyse de données & rapports qualité

PNF : Initialisation des fiches de métadonnées

EEA : Suivi de production des cartes







# Les formats

La librairie [GDAL/OGR](#) est utilisée pour les données vecteurs et RASTER

A word cloud of GIS data formats and standards. The text is arranged in a circular pattern, with each word or phrase rotated to follow the curve. The colors of the text vary, including shades of blue, green, yellow, orange, red, and purple. The words are of varying sizes, with some being larger and more prominent than others. The formats include:

- ESRI Shapefile
- GeoJSON
- KML
- GeoPackage
- OpenDocument Spreadsheet
- SQLite Spatialite
- PostgreSQL PostGIS
- OpenAir
- Geospatial PDF
- Selafin Seraphin format
- Walk
- OpenDocument Spreadsheet
- UK NTF
- PCI Geomatics Database File
- ESRI ArcObjects
- INGRES
- Oracle Spatial
- PDS
- FME Objects Gateway
- AutoCAD DXF
- GPX
- OGDI Vectors VPF VMAP DCW
- Aeronav FAA files
- GML
- Access MDB PGeo and Geomedia c
- Google Fusion Tables
- AutoCAD DWG
- ESRI ArcSDE
- GeoRSS
- NAS ALKIS
- Czech Cadastral Exchange Data
- GeorSS
- DODSOPeNDAP
- Géoconcept Export
- GMT
- EDIGEO
- US Census TIGERLine
- OGC WFS Web Feature Service
- GPStabel
- SEGY
- Informix DataBlade
- ODBC
- CouchDB
- GeoCouch
- OpenStreetMap XML and PBF
- SVG
- GRASS Vector Format
- Comma Separated Value csv
- SUA
- ESRI Personal GeoDatabase
- MS Excel format
- INTERLIS
- EPInfo REC
- MySQL
- ArcInfo Binary Coverage
- Mapinfo File
- Atlas BNA
- Hydrographic Transfer Format
- ESRI Generate
- ArcInfo E00 ASCII Coverage
- WASP map format
- MS Office Open XML spreadsheet
- VRT Virtual Datasource
- S57 ENC
- Storage and eXchange Format
- SDTS
- LIBKML
- SEGP1 UKOOA P190
- Idrisi Vector VCT
- Microstation DGN
- Google Maps Engine
- ElasticSearch
- Geomedia mdb
- Norwegian SOSI Standard
- XPlaneFlightgear aeronautical
- MS SQL Spatial

[illegible][illegible]

# Les formats testés

- Vecteur
  - ESRI Shapefile
  - Mapinfo Tab
  - Table PostGIS
  - GeoDatabase
  - GPX
  - KML
  - DXF
  - WFS
- RASTER
  - TIFF
  - ASC
  - GéoPDF
  - ECW
  - AIG
- Cartes
  - QGS project

# L'ETL

Talend (TDI) et son module spatial

[illegible]

# **L'atelier en 3 heures**

1. Installation (vidéo)
2. Mon premier Job
3. Configuration et présentation du fonctionnement
4. Les résultats dans GéoSource
5. Travaux à venir



# Les données aujourd'hui

- [Données vecteur Urban Atlas \(\(c\) Directorate-General Enterprise and Industry \(DG-ENTR\), Directorate-General for Regional Policy\) au format ESRI Shapefile](#)
- [Données vecteur OSM \(\(c\) OpenStreetMap contributors\) au format ESRI Shapefile](#)
- [Données RASTER Corine Land Cover \(\(c\) European Environment Agency \(EEA\)\) au format TIFF](#)

# Une clé USB bootable

Kubuntu

Talend Data Integration 5.4.1


GéoSource 2.11.0b0

Espace de travail Talend

Les données

Documentation en français et en anglais

Cette présentation



# Spatial extension for Talend

OPEN SOURCE GEO SPATIAL ETL

- Download
- FAQ
- Contact & Support

## What it does?

Transform and Integrate Data Between Geographic Information Systems

GeoSourceConnector Talend is a generic ETL framework [\[1\]](#) designed to [\[2\]](#) **enable** integrating geospatial data, processing, storing and analyzing information.

GeoSourceConnector Talend is a generic ETL framework designed to enable integrating geospatial data, processing, storing and analyzing information. It is designed to be used in a variety of environments, including on-premise, cloud, and hybrid environments. It is designed to be used in a variety of environments, including on-premise, cloud, and hybrid environments. It is designed to be used in a variety of environments, including on-premise, cloud, and hybrid environments.

## How it works?

Bridging robust Open Source geospatial libraries for efficient data management

The main connector uses [\[3\]](#) **GeoTools**, [\[4\]](#) **GeoServer**, [\[5\]](#) **GeoFence** for loading and writing GIS data and for processing geospatial data. The connector is designed to be used in a variety of environments, including on-premise, cloud, and hybrid environments.

## What are the supported formats?

GeoSourceConnector Talend supports the following formats: [\[6\]](#) **GeoJSON**, [\[7\]](#) **GeoXML**, [\[8\]](#) **GeoRSS**, [\[9\]](#) **GeoML**, [\[10\]](#) **GeoML2**, [\[11\]](#) **GeoML3**, [\[12\]](#) **GeoML4**, [\[13\]](#) **GeoML5**, [\[14\]](#) **GeoML6**, [\[15\]](#) **GeoML7**, [\[16\]](#) **GeoML8**, [\[17\]](#) **GeoML9**, [\[18\]](#) **GeoML10**, [\[19\]](#) **GeoML11**, [\[20\]](#) **GeoML12**, [\[21\]](#) **GeoML13**, [\[22\]](#) **GeoML14**, [\[23\]](#) **GeoML15**, [\[24\]](#) **GeoML16**, [\[25\]](#) **GeoML17**, [\[26\]](#) **GeoML18**, [\[27\]](#) **GeoML19**, [\[28\]](#) **GeoML20**, [\[29\]](#) **GeoML21**, [\[30\]](#) **GeoML22**, [\[31\]](#) **GeoML23**, [\[32\]](#) **GeoML24**, [\[33\]](#) **GeoML25**, [\[34\]](#) **GeoML26**, [\[35\]](#) **GeoML27**, [\[36\]](#) **GeoML28**, [\[37\]](#) **GeoML29**, [\[38\]](#) **GeoML30**, [\[39\]](#) **GeoML31**, [\[40\]](#) **GeoML32**, [\[41\]](#) **GeoML33**, [\[42\]](#) **GeoML34**, [\[43\]](#) **GeoML35**, [\[44\]](#) **GeoML36**, [\[45\]](#) **GeoML37**, [\[46\]](#) **GeoML38**, [\[47\]](#) **GeoML39**, [\[48\]](#) **GeoML40**, [\[49\]](#) **GeoML41**, [\[50\]](#) **GeoML42**, [\[51\]](#) **GeoML43**, [\[52\]](#) **GeoML44**, [\[53\]](#) **GeoML45**, [\[54\]](#) **GeoML46**, [\[55\]](#) **GeoML47**, [\[56\]](#) **GeoML48**, [\[57\]](#) **GeoML49**, [\[58\]](#) **GeoML50**, [\[59\]](#) **GeoML51**, [\[60\]](#) **GeoML52**, [\[61\]](#) **GeoML53**, [\[62\]](#) **GeoML54**, [\[63\]](#) **GeoML55**, [\[64\]](#) **GeoML56**, [\[65\]](#) **GeoML57**, [\[66\]](#) **GeoML58**, [\[67\]](#) **GeoML59**, [\[68\]](#) **GeoML60**, [\[69\]](#) **GeoML61**, [\[70\]](#) **GeoML62**, [\[71\]](#) **GeoML63**, [\[72\]](#) **GeoML64**, [\[73\]](#) **GeoML65**, [\[74\]](#) **GeoML66**, [\[75\]](#) **GeoML67**, [\[76\]](#) **GeoML68**, [\[77\]](#) **GeoML69**, [\[78\]](#) **GeoML70**, [\[79\]](#) **GeoML71**, [\[80\]](#) **GeoML72**, [\[81\]](#) **GeoML73**, [\[82\]](#) **GeoML74**, [\[83\]](#) **GeoML75**, [\[84\]](#) **GeoML76**, [\[85\]](#) **GeoML77**, [\[86\]](#) **GeoML78**, [\[87\]](#) **GeoML79**, [\[88\]](#) **GeoML80**, [\[89\]](#) **GeoML81**, [\[90\]](#) **GeoML82**, [\[91\]](#) **GeoML83**, [\[92\]](#) **GeoML84**, [\[93\]](#) **GeoML85**, [\[94\]](#) **GeoML86**, [\[95\]](#) **GeoML87**, [\[96\]](#) **GeoML88**, [\[97\]](#) **GeoML89**, [\[98\]](#) **GeoML90**, [\[99\]](#) **GeoML91**, [\[100\]](#) **GeoML92**, [\[101\]](#) **GeoML93**, [\[102\]](#) **GeoML94**, [\[103\]](#) **GeoML95**, [\[104\]](#) **GeoML96**, [\[105\]](#) **GeoML97**, [\[106\]](#) **GeoML98**, [\[107\]](#) **GeoML99**, [\[108\]](#) **GeoML100**, [\[109\]](#) **GeoML101**, [\[110\]](#) **GeoML102**, [\[111\]](#) **GeoML103**, [\[112\]](#) **GeoML104**, [\[113\]](#) **GeoML105**, [\[114\]](#)

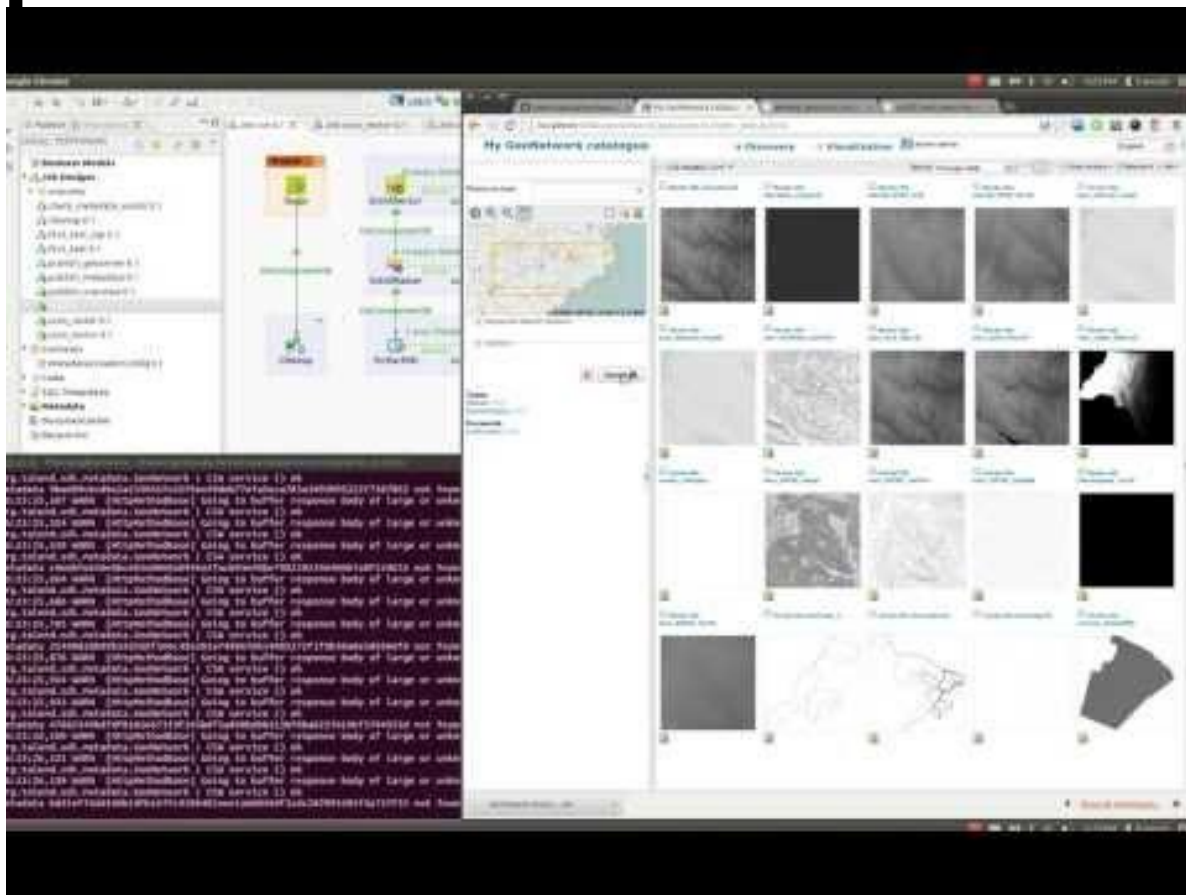
# **Mon premier Job**

# L'inventaire de données

<https://docs.google.com/document/d/10GyQLL9XORawurlLI6AUCpP7PDfcQ3D3nkBO1Le3UI8/edit?usp=sharing>

# GéoSource

# Aller plus loin





# Travaux à venir & idées

Support des cartes ArcGIS

Améliorer la mise à jour des fiches via CSW

Coupler l'agent d'analyse avec le catalogue en mode webservice





Automatic geospatial data inventory with Talend Spatial — Edit

17 commits

1 branch

0 releases

2 contributors



branch: master

[workspace-metadata-crawler](#)

WebService / Add ping parameter to get current config (#4). Add param...



fxprunayre authored 3 days ago

latest commit 80475ff2af



METADATA-CRAWLER

WebService / Add ping parameter to get current config (#4). Add param...

3 days ago



README.md

Introduction

14 days ago



README.md

# Automatic geospatial data inventory with Talend Spatial

The data inventory is made by scanning folders, databases or webservises. For each geospatial resource (ie. file, table, layers) found, main properties (eg. bounding box, projection) are computed and metadata records created. After metadata generated, they could be published into a catalog.

For vector resources, 2 metadata records are created :

- one metadata for the dataset following ISO19139 standards
- one metadata for the feature catalog (ie. list of columns, datatype, ...) following ISO19110 standards

For RASTER dataset, only the dataset metadata record is created.

## Data formats

<> Code

Issues 3

Pull Requests 0

Wiki

Pulse

Graphs

Network

Settings

HTTPS clone URL

<https://github.com/>



You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).



Download ZIP