



Inventaire automatique du patrimoine d'information géographique des Parcs nationaux

En partenariat avec François Prunayre, société Titellus

Historique

- Une réflexion sur le catalogage initiée au sein du GT des géomaticiens des Parcs nationaux dès 2007,

Historique

- Une réflexion sur le catalogage initiée au sein du GT des géomaticiens des Parcs nationaux dès 2007,
- En 2009, la décision est prise de mettre en place un service de catalogage commun avec comme objectifs à atteindre :
 - > Se mettre en conformité avec les principes de la Directive Inspire
 - > Éviter la création de données redondantes
 - > Ouvrir les catalogues aux partenaires et au grand public

Historique

- **Une réflexion sur le catalogage initiée au sein du GT des géomaticiens des Parcs nationaux dès 2007,**
- **En 2009, la décision est prise de mettre en place un service de catalogage commun avec comme objectifs à atteindre :**
 - > Se mettre en conformité avec les principes de la Directive Inspire
 - > Éviter la création de données redondantes
 - > Ouvrir les catalogues aux partenaires et au grand public
- **Très rapidement apparaissent des difficultés :**
 - > Techniques
 - > Fonctionnelles
 - > En découle une faible mobilisation des gestionnaires de données des parcs

Quel bilan pour cette première étape ?

● Un premier bilan très mitigé :

- > Très peu de fiches saisies
- > Le catalogage n'est pas ressenti comme une activité prioritaire des parcs nationaux
- > Manque de temps à y consacrer
- > Catalogage : activité fastidieuse, chronophage, peu valorisante
- > Une norme éloignée d'une gestion optimisée des données
- > Les modèles de données sont compliqués
- > Inadéquation de l'outil avec les besoins et les attentes des géomaticiens
- > Dynamique régionale pour les parcs d'outre-mer s'appuyant sur des solutions techniques différentes (géoguyane par exemple)

À partir de ce constat, une nouvelle réflexion est engagée en 2013

- **Remise à plat du précédent processus de catalogage**

- > Identification des points de blocage
- > Quels enseignements en tirer ?
- > État de l'art en 2013
- > Quelles solutions et quels moyens disponibles ?

À partir de ce constat, une nouvelle réflexion est engagée en 2013

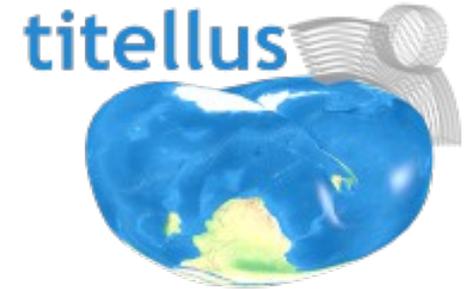
- **Remise à plat du précédent processus de catalogage**

- > Identification des points de blocage
- > Quels enseignements en tirer ?
- > État de l'art en 2013
- > Quelles solutions et quels moyens disponibles ?

- **Cibler l'effort sur la génération de métadonnées
... et non sur l'outil de catalogage**

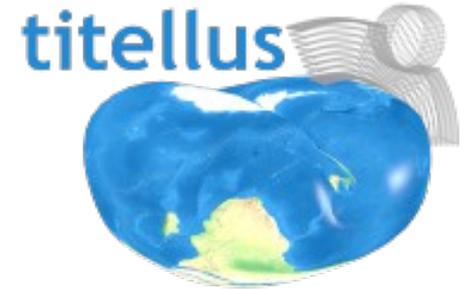
À partir de ce constat, une nouvelle réflexion est engagée en 2013

- Cette réflexion est menée en collaboration entre PNF et Francois Prunayre de la société Titellus



À partir de ce constat, une nouvelle réflexion est engagée en 2013

- Cette réflexion est menée en collaboration entre PNF et Francois Prunayre de la société Titellus



- **Objectif à atteindre :**

- > Mettre à disposition des Parcs nationaux un outil d'inventaire de leurs données géographiques
- > Générer le plus simplement possible des squelettes de métadonnées
- > Avoir un outil synchronisé aux données géographiques
- > S'appuyer sur une solution facilement dépersonnalisable

- **Solution retenue :**

- => **Talend (TDI) et son module spatial**

- => **génération de métadonnées par analyse de données SIG**



- **Talend ?**

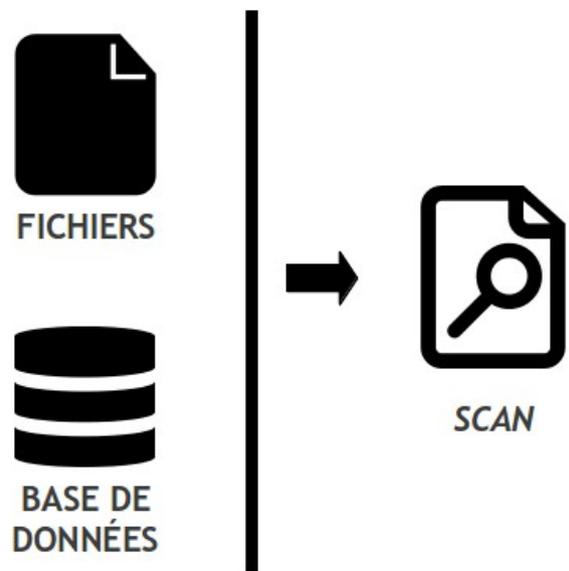
- > Outil ETL Open-source (Extract Transform Load)

- > Talend déjà utilisé par certains parcs

- > Talend bénéficie d'un module spatial s'appuyant sur GDAL/OGR

● Principe :

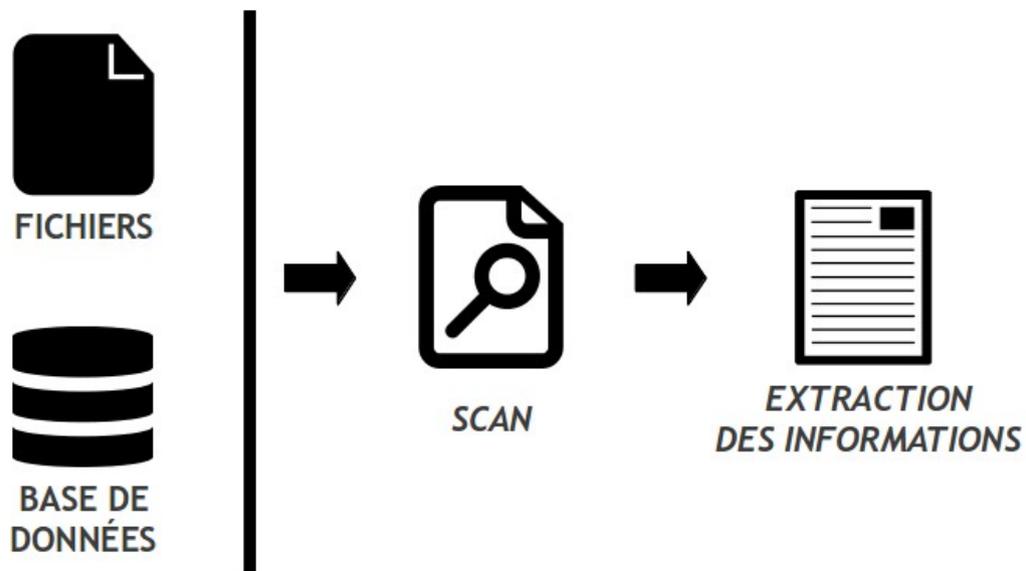
Scan de répertoires,
bases de données ou
de services web



● Principe :

Scan de répertoires,
bases de données ou
de services web

Extraction des caractéristiques
de la données, puis création de
fiches descriptives à partir d'un
modèle de saisie

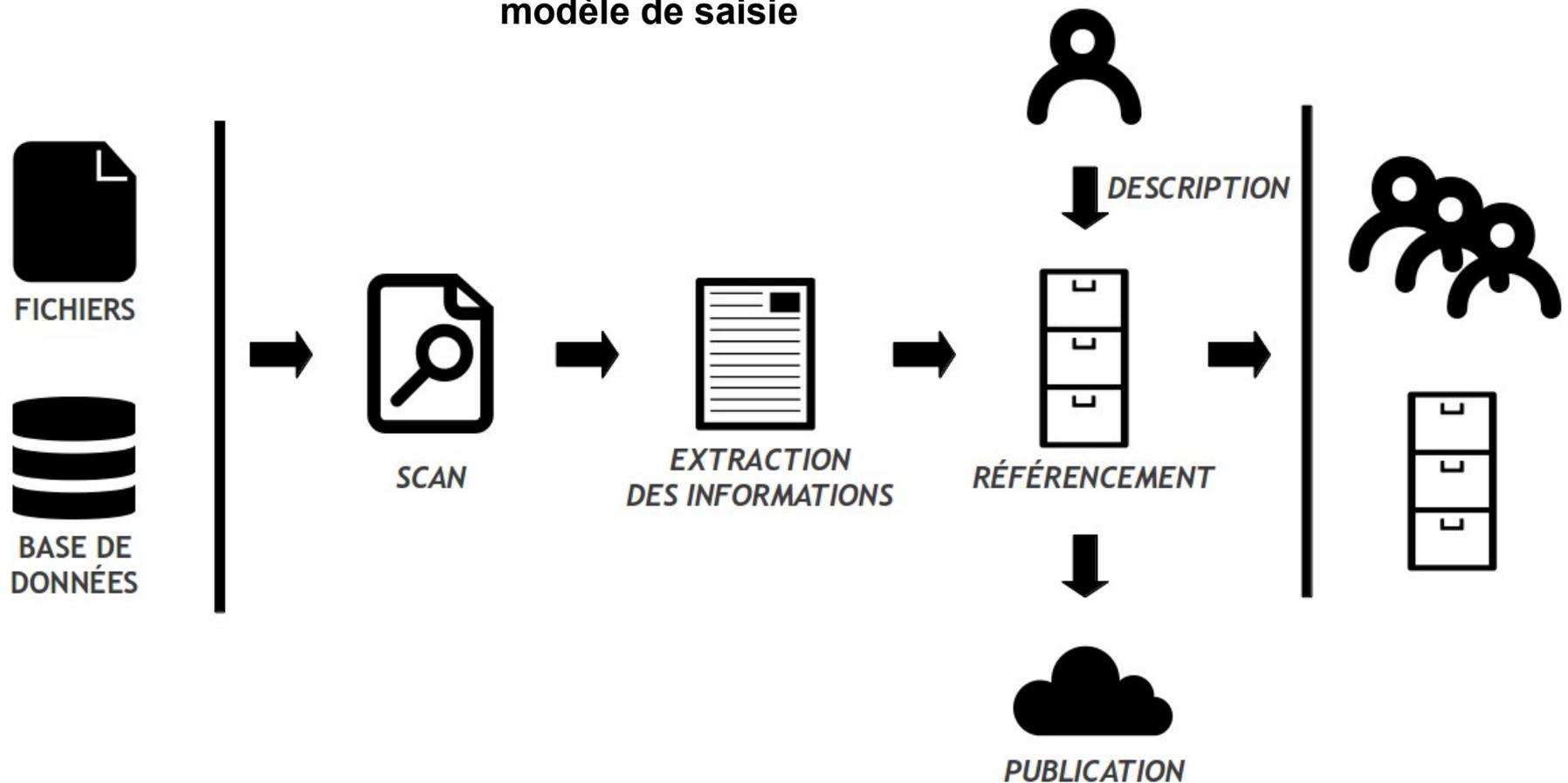


● Principe :

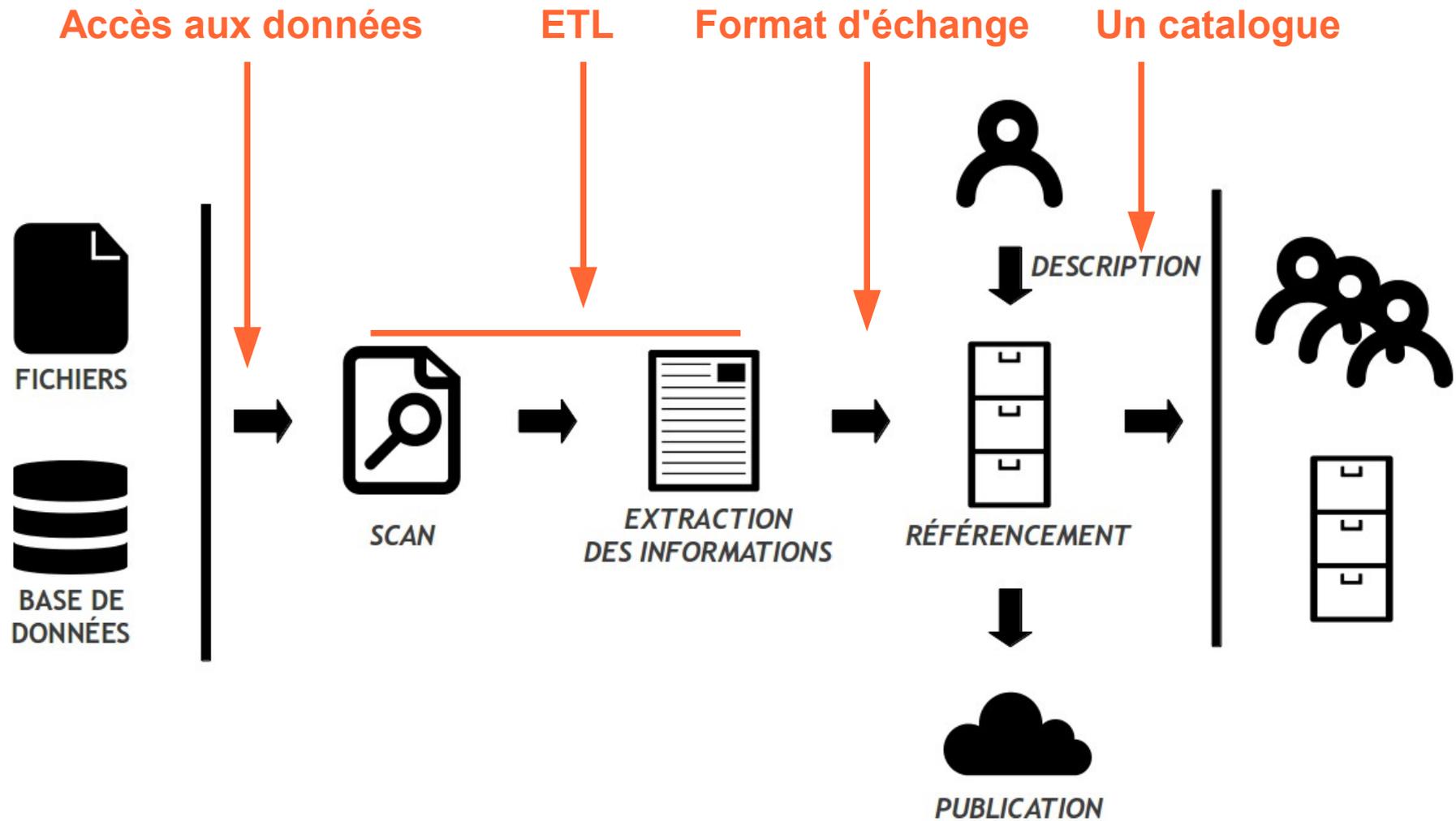
Scan de répertoires,
bases de données ou
de services web

Extraction des caractéristiques
de la données, puis création de
fiches descriptives à partir d'un
modèle de saisie

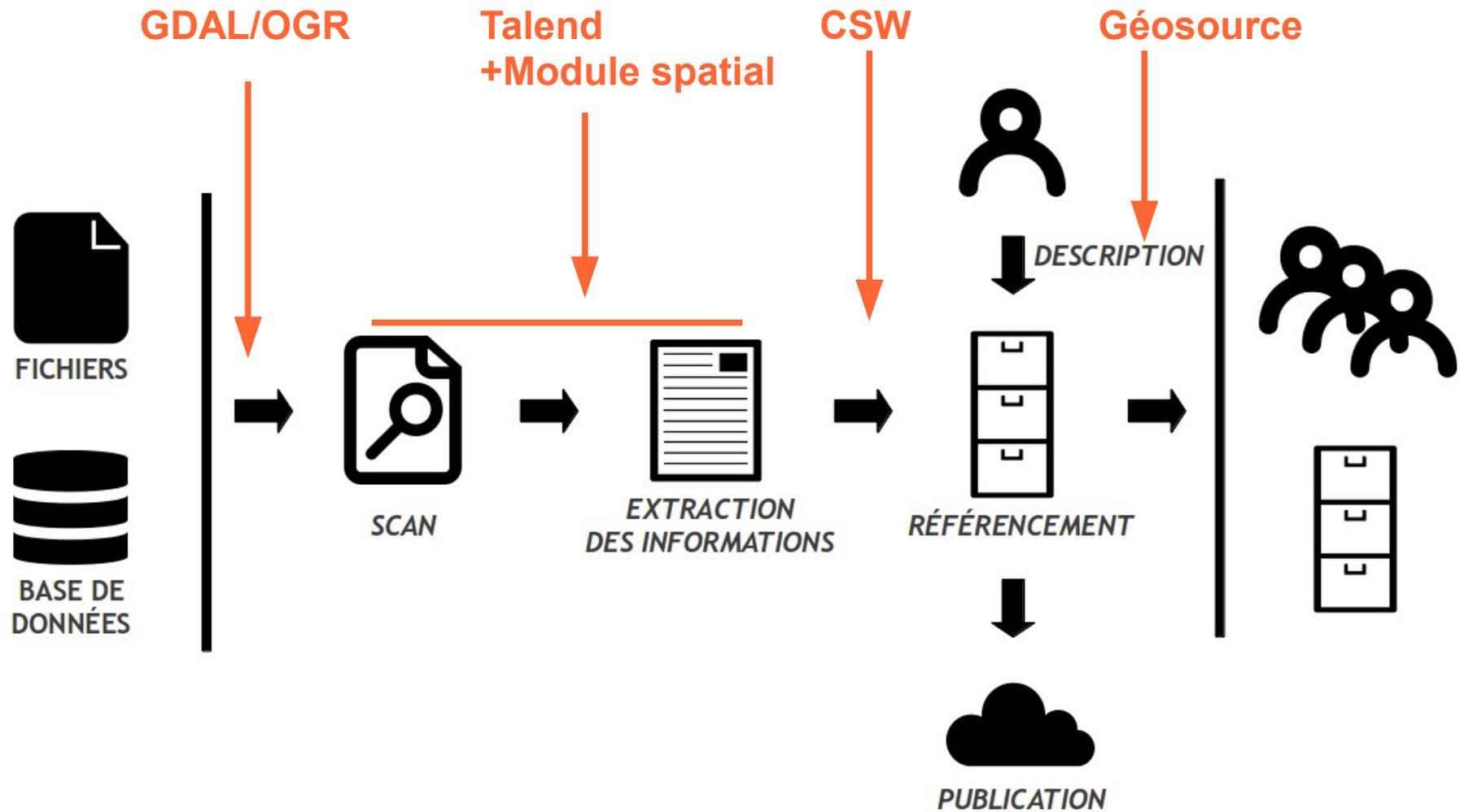
Publication des fiches dans
un service de catalogage



● Principe :



● Principe :



● Formats testés :

> Vecteurs

- Esri Shapefile
- Mapinfo Tab
- Table PostGIS
- GeoDatabase
- GPX
- KML
- DXF
- WFS

> Cartes

- QGIS project

> Raster

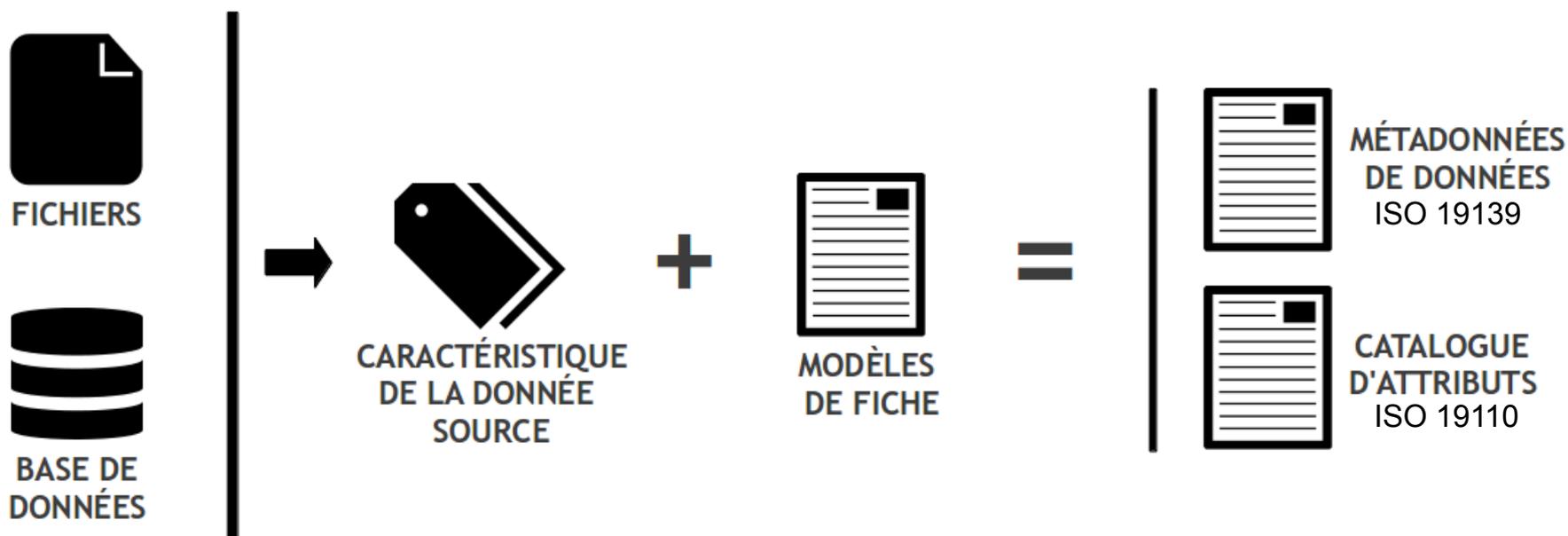
- TIFF
- ASC
- GéoPDF
- ECW
- AIG

> Source

- Fichiers
- Bases de données
- Services web

● Pour des données vecteurs :

- > Création d'une métadonnée de données au format ISO 19139
- > Création d'un catalogue d'attribut au format ISO 19110



● Pour des données raster :

- > Seule la fiche de métadonnées est générée

● Une métadonnée INSPIRE

- > Titre
- > Résumé
- > Contact pour la donnée
- > Contact pour la métadonnée
- > Contrainte d'accès
- > Thème
- > Un mot clé GEMET
- > Date
- > Résolution spatiale
- > Représentation spatiale
- > Ressource en ligne
- > Format de diffusion
- > Étendue géographique
- > Système de projection
- > Généalogie

30% de fait ...



EXTRACTION
DES INFORMATIONS

- > **Identifiant de la métadonnée = UUID**
- > **Identifiant de la ressource = <http://parcsnationaux.fr/UUID>**
- > **Langue de la ressource = fr**
- > **Langue de la métadonnée = fr**
- > **Jeu de caractère de la ressource = utf-8**
- > **Jeu de caractères de la métadonnée = utf-8**
- > **Nom de la norme = ISO 19139**
- > **Version de la norme = 2006**
- > **Date de la métadonnée = date de l'édition**

● Une métadonnée INSPIRE

Modèle de saisie : **60%**

- > Titre
- > Résumé
- > **Contact pour la donnée = bruno lafage**
- > **Contact pour la métadonnée = bruno lafage**
- > **Contrainte d'accès = pas de contrainte, convention, licence ...**
- > **Thème = Environnement**
- > Un mot clé GEMET
- > Date
- > Résolution spatiale
- > Représentation spatiale
- > Ressource en ligne
- > Format de diffusion
- > Étendue géographique
- > Système de projection
- > Généalogie
- > **Identifiant de la métadonnée = UUID**
- > **Identifiant de la ressource = <http://parcsnationaux.fr/UUID>**
- > **Langue de la ressource = fr**
- > **Langue de la métadonnée = fr**
- > **Jeu de caractère de la ressource = utf-8**
- > **Jeu de caractères de la métadonnée = utf-8**
- > **Nom de la norme = ISO 19139**
- > **Version de la norme = 2006**
- > **Date de la métadonnée = date de l'édition**



EXTRACTION
DES INFORMATIONS

● Une métadonnée INSPIRE

Analyse de la donnée : **85%**

- > Titre
- > Résumé
- > **Contact pour la donnée = bruno lafage**
- > **Contact pour la métadonnée = bruno lafage**
- > **Contrainte d'accès = pas de contrainte, convention, licence ...**
- > **Thème = Environnement**
- > Un mot clé GEMET
- > Date
- > Résolution spatiale
- > **Représentation spatiale**
- > **Ressource en ligne**
- > **Format de diffusion**
- > **Étendue géographique**
- > **Système de projection**
- > Généalogie
- > **Identifiant de la métadonnée = UUID**
- > **Identifiant de la ressource = <http://parcsnationaux.fr/UUID>**
- > **Langue de la ressource = fr**
- > **Langue de la métadonnée = fr**
- > **Jeu de caractère de la ressource = utf-8**
- > **Jeu de caractères de la métadonnée = utf-8**
- > **Nom de la norme = ISO 19139**
- > **Version de la norme = 2006**
- > **Date de la métadonnée = date de l'édition**

) calcul à partir de la donnée géographique



EXTRACTION
DES INFORMATIONS

● Une métadonnée INSPIRE

- > Titre
- > Résumé
- > Contact pour la donnée = bruno lafage
- > Contact pour la métadonnée = bruno lafage
- > Contrainte d'accès = pas de contrainte, convention, licence ...
- > Thème = Environnement
- > Un mot clé GEMET
- > Date
- > Résolution spatiale
- > Représentation spatiale
- > Ressource en ligne
- > Format de diffusion
- > Étendue géographique
- > Système de projection
- > Généalogie
- > Identifiant de la métadonnée = UUID
- > Identifiant de la ressource = <http://parcsnationaux.fr/UUID>
- > Langue de la ressource = fr
- > Langue de la métadonnée = fr
- > Jeu de caractère de la ressource = utf-8
- > Jeu de caractères de la métadonnée = utf-8
- > Nom de la norme = ISO 19139
- > Version de la norme = 2006
- > Date de la métadonnée = date de l'édition



↓ DESCRIPTION



EXTRACTION
DES INFORMATIONS

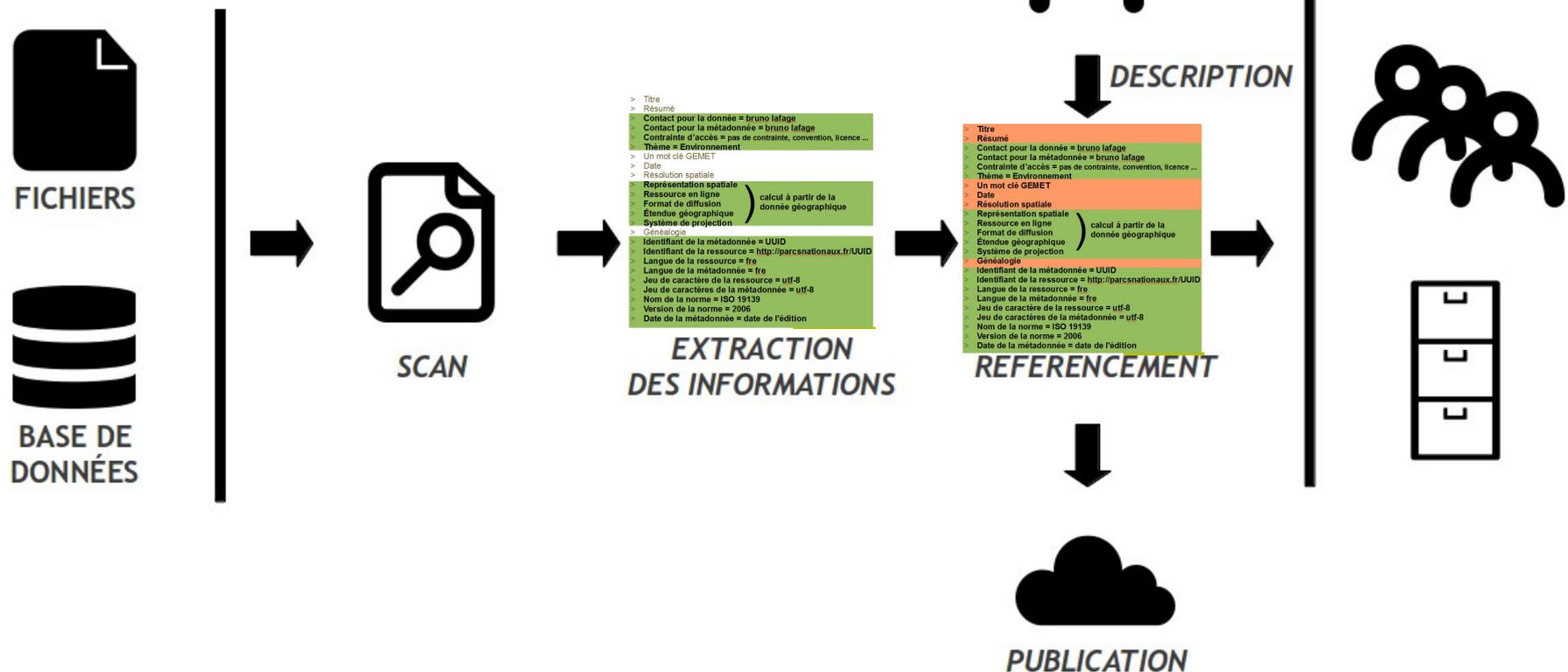
) calcul à partir de la
donnée géographique

● Principe :

Scan de répertoires,
bases de données ou
de services web

Extraction des caractéristiques
de la données puis création de
fiches descriptives à partir d'un
modèle de saisie

Publication des fiches dans
un service de catalogage



● Fenêtre Talend (TDI)

Check that the context is properly configured and the resources folder setup before running the job.

Prejob:Start
begin

OnComponentOk
Cleanup

If (order1)
callScan

If (order2)
publish_metadata

OnComponentOk
publish_overview

If (order2)
listAllVector
scan_vector

If (order1)
listAllRaster
scan_raster

If (order3)
forEachDb
scan_vector

If (order4)
forEachService
scan_vector

If (order5)
listAllMaps
scan_map_qgis

Return config (used in webservice mode usual...)
ping
row1 (Main)
pingResponse

Repository: LOCAL: Metadata_V2

- Business Models
- Jobs
 - overview
 - check_metadata_exists 0.1
 - cleanup 0.1
 - first_test_ogr 0.1
 - first_test 0.1
 - ping 0.1
 - publish_geoserver 0.1
 - publish_metadata 0.1
 - publish_overview 0.1
 - run 0.1
 - scan_map_qgis 0.1
 - scan_raster 0.1
 - scan_vector 0.1
- Contextes
 - MetadataCrawlerConfig 0.1
- Code
- Modèles SQL
- Métadonnées
 - Documentation
 - Corbeille

Structure | Aperçu du code

```

tBufferOutput_1 (pingResponse)
tFileList_1 (listAllRaster)
tFileList_2 (listAllVector)
tFileList_3 (listAllMaps)
tForEach_1 (forEachDb)
tForEach_2 (forEachService)
tLoop_1 (callScan)
tPrejob_1 (begin)
tRunJob_1 (scan_vector)
tRunJob_10 (scan_raster)
tRunJob_2 (scan_raster)
tRunJob_3 (publish_metadata)
tRunJob_4 (scan_map_qgis)
tRunJob_6 (scan_vector)
tRunJob_7 (publish_overview)
tRunJob_8 (Cleanup)
tRunJob_9 (ping)
    
```

Designer | Code

Contextes(MetadataCrawlerConfig 0.1) | Contextes(Job run 0.1) | Composant | Exécuter (Job run)

Job run

Exécution simple
Exécution Debug

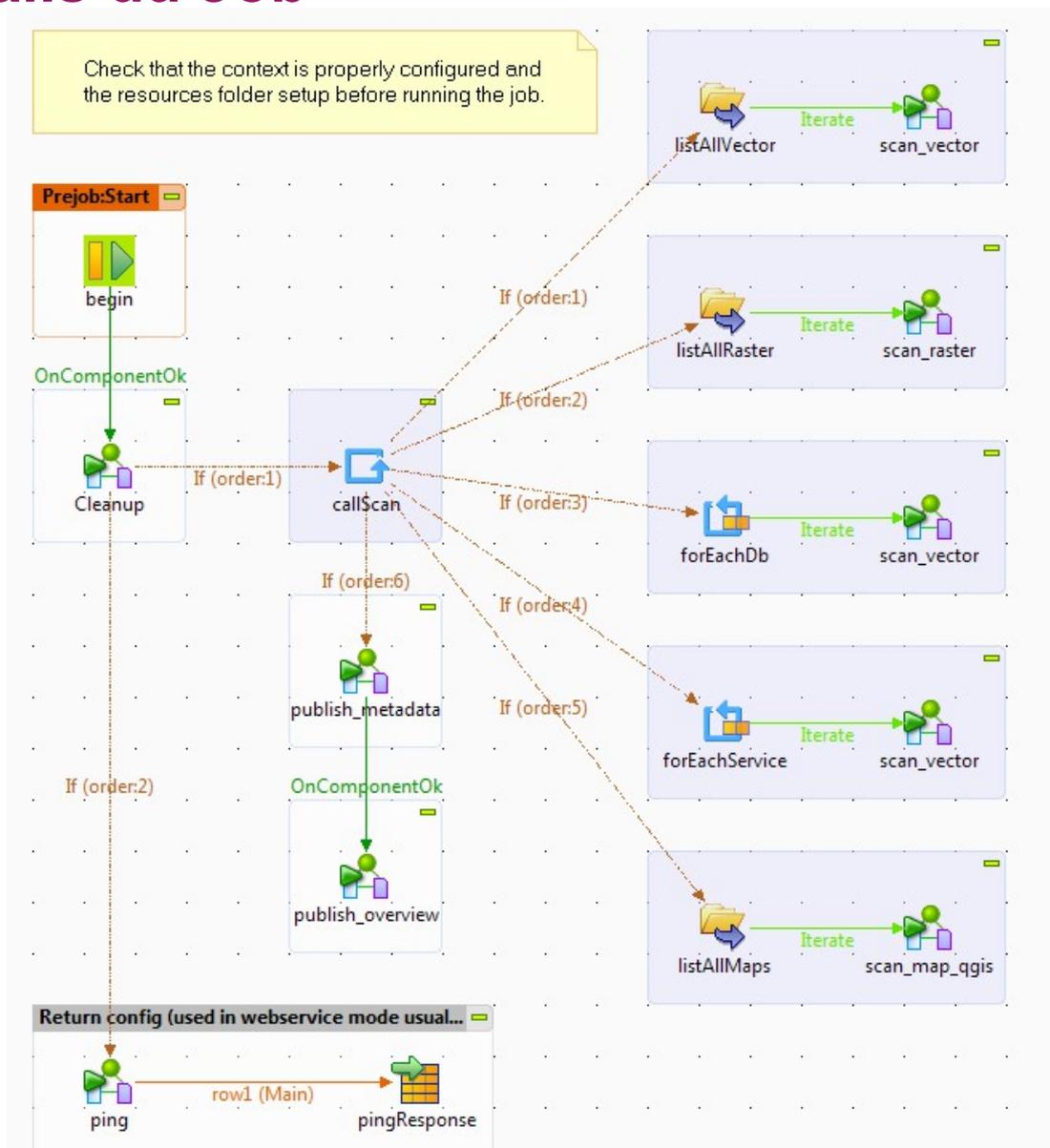
Exécution
▶ Exécuter | ⏹ Arrêter | 🗑 Effacer

Default

| Nom | Valeur |
|----------|---------------------|
| data_dir | G:/DonneesProfes... |
| db_list | "PG:dbname=obs... |

1 items sélectionnés

● Détails du Job



● Paramétrage des variables du Job

| Nom | Code du script | Commentaire |
|--------------------------------------|--|--|
| create_overview | context.create_overview | Create or not overview files |
| csw_password | context.csw_password | The CSW user password |
| csw_server_host | context.csw_server_host | The CSW server to push metadata records into |
| csw_username | context.csw_username | The CSW user name |
| data_dir | context.data_dir | The folder to scan for datasets |
| db_list | context.db_list | List of database to scan |
| default_projection | context.default_projection | An EPSG code of a projection to use in case the projection could not be defined from the file |
| default_xmax | context.default_xmax | |
| default_xmin | context.default_xmin | |
| default_ymax | context.default_ymax | |
| default_ymin | context.default_ymin | |
| ftp_dir | context.ftp_dir | The remove directory in the FTP server |
| ftp_host | context.ftp_host | The FTP host |
| ftp_password | context.ftp_password | The user password |
| ftp_port | context.ftp_port | The FTP port |
| ftp_username | context.ftp_username | The user name to log in the FTP server |
| gdal_path | context.gdal_path | Path to GDAL/OGR bin or empty if in system path |
| geoserver_password | context.geoserver_password | |
| geoserver_server_host | context.geoserver_server_host | |
| geoserver_username | context.geoserver_username | |
| geoserver_namespace | context.geoserver_namespace_prefix | |
| log_dir | context.log_dir | The log folder |
| map_data_dir | context.map_data_dir | The folder to scan for maps |
| map_file_extensions | context.map_file_extensions | Map file extensions |
| metadata_dir | context.metadata_dir | The folder to create metadata records into |
| metadata_file_with_source | context.metadata_file_with_source | Store the generated metadata file next to the analyzed file |
| overview_copy_mode | context.overview_copy_mode | Define the way to copy overview (ssh, ftp, copy) |
| overview_folder | context.overview_folder | The folder to create overview into |
| overview_height | context.overview_height | The height of the overview to generate |
| overview_width | context.overview_width | The width of the overview to generate |
| ping | context.ping | True to ping the job and return status response (use when deploy as webservice) |
| publish_metadata | context.publish_metadata | Publish metadata or not in the CSW catalogue defined |
| publish_metadata_full_update | context.publish_metadata_full_update | Always run update by sending the complete metadata record (not only some properties like bounding box) |
| raster_file_extensions | context.raster_file_extensions | RASTER file extensions |
| raster_file_extensions_for_geoserver | context.raster_file_extensions_for_geoserver | RASTER file extensions supported for publication in mapservers |
| resource_dir | context.resource_dir | The folder with property files |
| scan_map | context.scan_map | Scan for maps |
| scan_map_and_datasources | context.scan_map_and_datasources | For each datasources find in a map, create metadata |
| scan_raster | context.scan_raster | Scan for RASTER files |

● Installation

- > Téléchargement et installation
 - JAVA
 - Talend Data Integration
 - Module spatial de Talend
 - GDAL/OGR
- > Configuration de Talend
 - Importation des jobs
 - Paramétrage des jobs et des variables

● Exécution

- > Directement dans Talend
- > En ligne de commande
- > Webservice

● Besoin de documentation ?

> Google doc

- **Inventaire automatique du patrimoine d'information géographique avec Talend Spatial**
- Auteurs et contributeurs :

François Prunayre (Titellus)

Mathieu Coudert (Astrium Geo-Information Services)

Bruno Lafage (Parcs nationaux de France)

<https://docs.google.com/document/d/10GyQLL9XORawurLI6AUCpP7PDfcQ3D3nkBO1Le3UI8/edit?pli=1#>

> Téléchargement depuis github.com (documentation + jobs)

<https://github.com/talend-spatial/talend-spatial/wiki/Data-inventory>

<https://github.com/talend-spatial/workspace-metadata-crawler/archive/master.zip>

Inventaire automatique du patrimoine d'information géographique avec Talend Spatial

[Auteurs & contributeurs](#)

[Principe](#)

[Téléchargements & installation](#)

[Télécharger JAVA](#)

[Télécharger GDAL/OGR](#)

[Installer TDI](#)

[Premier lancement](#)

[Importer les jobs dans l'espace de travail](#)

[Valider la bonne installation du module Spatial](#)

[Configurer GDAL/OGR](#)

[Valider l'installation GDAL/OGR](#)

[Chaîne de traitements](#)

[Les formats supportés](#)

[L'extraction des propriétés par analyse des fichiers](#)

[Règles générales](#)

[Le jeu de test](#)

[Configurer les jobs](#)

[Répertoire de fichier de propriétés](#)

[Configuration des fichiers de propriétés](#)

[Lancement des jobs depuis le Studio](#)

[Job "run"](#)

[Job "scan vector"](#)

[Job "scan raster"](#)

[Job "publish metadata"](#)

[Job "publish overview"](#)

[Export du job](#)

[FAQ](#)

[Conclusions](#)

[Annexes](#)

[Autres travaux similaires & références](#)

[Corrections & améliorations apportées à TDI](#)

[Corrections & améliorations apportées au module Spatial](#)

- Une solution présentée lors d'événements majeurs



FOSS4G
NOTTINGHAM 2013

OSGeo's Global Conference for Open Source Geospatial Software

The annual gathering of Open Source Geospatial Developers, Users and Leaders. At Nottingham's new East Midlands Conference Centre in the heart of the country, 17th to 21st September

"Geo For All"

● **Quelle suite pour les Parcs nationaux**

- > Une solution déployée dans tous les Parcs nationaux
- > Mobilisation de chacun afin de mettre son patrimoine de données en conformité avec INSPIRE
- > Soins laissés à chaque parc de choisir son service de catalogage (moissonable)
- > Mise en place à PNF d'un catalogue des Parcs nationaux
- > Connexion envisagée avec les catalogues nationaux

Merci de votre attention

et bon catalogage !

