

# VALIDATION OF INSPIRE DATA



Modèle TN-02.018-1.6

Dominique Laurent  
September 2018 – INSPIRE conference

© IGN-SHOM





# INTRODUCTION

# PLAN

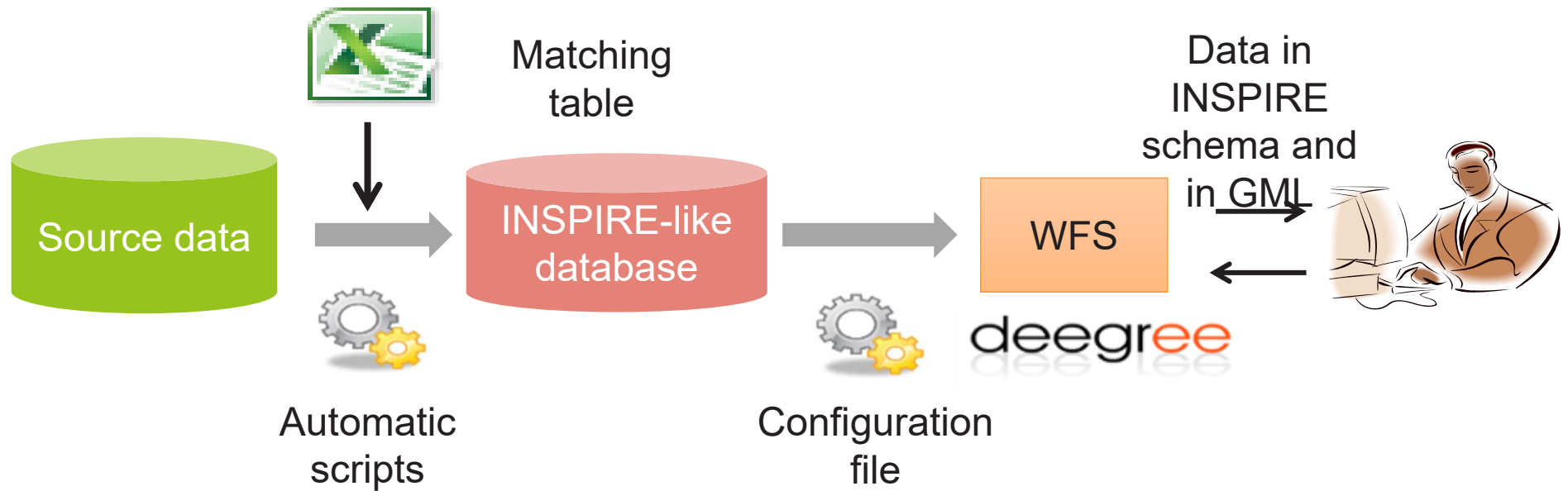
Transformation process

Validation against INSPIRE

Validation against source data

Conclusions

# TRANSFORMATION PROCESS



A very common two-steps approach



## VALIDATION AGAINST INSPIRE

# Context

 Tool: INSPIRE official validator





- on-line

 Which tests?

- Application schema
- Reference systems
- Data consistency (if automatic)

 Who is testing?

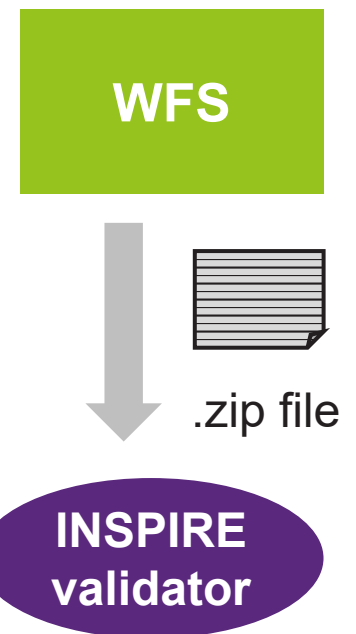
- The GeoPortal team (in charge of WFS)

Data Theme: Addresses (Data Specification version 3.1)	
	Conformance class: Application schema, Addresses
	Conformance class: Data consistency, Addresses
	Conformance class: Information accessibility, Addresses
	Conformance class: Reference systems, Addresses

# Methodology

## 📍 Select sample data

- **General case: all themes except networks**
  - **n first features** of each feature type (around 10)
  - In case of associations, validator is able to find related target features on Internet
- **Specific case: networks (TN, HY)**
  - Rule about nodes that should be end/start of links
  - **Use of bounding box** (to get the links related to the nodes)



# Methodology

 Run the test

 Analyse the results

- **No error => no action**
- **If error**
  - Understand where it comes from
  - Correct it directly or send it to the team in charge of previous step in the transformation process



# Learnings

## Nice:

- **Automatic tool**
- **Stable version since 2017**
- **Understandable error message (even if not always simple)**

## Not so nice

- **validation time : a few minutes to check around 10 features!**
- **if use of tool on Internet, data has to be published on Internet before validation**

# Potential improvements

 Use the INSPIRE validator in our local environment

- we might validate data before publication on Intranet
- we might automate the validation process
  
- **but tool is not easy to be installed**



## TESTING AGAINST SOURCE DATA

# Context and objectives

## Objectives

- **Ensure data content quality**
- **Validate our transformation process**

## Context

- **Depends on source data and on matching rules**
- **Very specific => home-made solution**

# Methodology

- 📍 Testing plan
  - 📍 Prepare test suite on FME
  - 📍 Extract sample data
  - 📍 Run the tests
  - 📍 Send report test for validation or error corrections
- } Making IGN validator

# Methodology: testing plan

## Principle:

- Transform matching table into tests
- Done by INSPIRE expert
- Not fully exhaustive

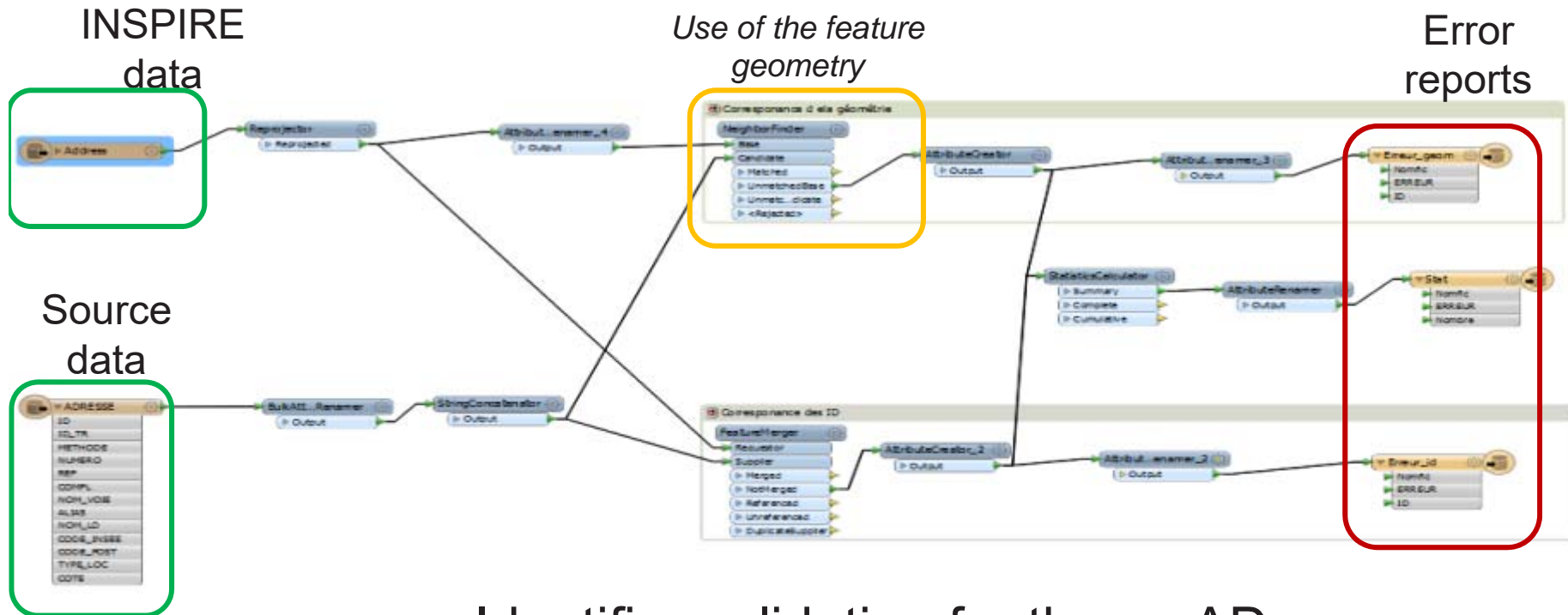
**Tests proposés sur la classe RoadLink**

1. Extraire les objets de SD UNI correspondants à des RoadLink d'INSPIRE : il s'agit des tronçons de route tels que [NATURE <-> 'Bac auto and NATURE <-> 'Bac piéton']
2. Vérifier le nombre d'objets transformés  
Nombre de RoadLink (INSPIRE) = nombre d'objets SD UNI
- 3- Vérifier le traitement de quelques attributs « vides »
  - attribut beginLifespanVersion : l'attribut INSPIRE est vide mais le voidValueReason, or nilReason est mis à « unpopulated »
  - attribut endLifespanVersion : l'attribut INSPIRE est vide et le voidValueReason, or nilReason est aussi vide

NamedPlace	beginLifespanVersion	"VoidValueReason" = unpopulated	(INDUSCOM, ZONEHAB, EQUADML, CULTLOIS, SCIENENS, GESTEAUX, RELIGIE, SANTE, SPORT, TRANSPOR, HYDROGRA, OROGRAPH, ESPACNAT) where 'GRAPHEP' is not null.		
	endLifespanVersion				
	geometry	Case 1: If "Lien vers SURACTIN" is not null, then SURACTIN.the_geom ; Case 2: If "Lien vers SURACTIN" is null, then the_geom			
	inspireId	see Types complexes			
	leastDetailedViewingResolution	equivalentScale=50000			



# Methodology: test suite (FME)



## Identifiant validation for theme AD

# Methodology: extract sample data

## Principle:

- **Get representative data with as many cases as possible (e.g. values in code list)**

## Method for each theme or sub-theme

- **Selection in source data**
- **Based on the main feature type (with more instances)**
  - Ex: Tronçon Route (RoadLink) for Road TN
  - => Find an area with around 5 000 instances (limit of WFS request)



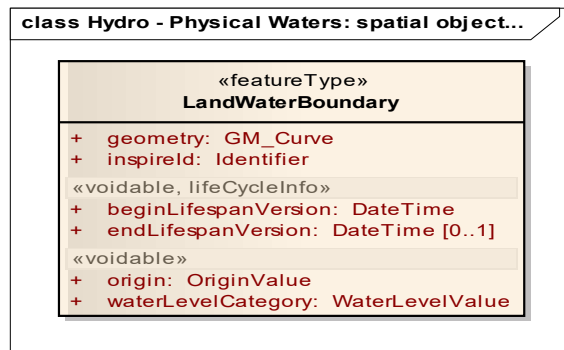
**Get a  
bounding  
box**



# Methodology: extract sample data

## 📍 Challenge : extract related INSPIRE data

- Case 1: isolated feature type (with geometry, no association)

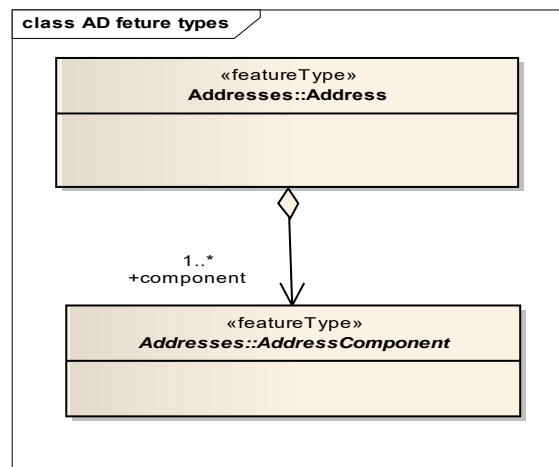


Simple URL request  
to get relevant  
features

# Methodology: extract sample data

## 📍 Challenge : extract related INSPIRE data

- Case 2: feature type with geometry and with associated features

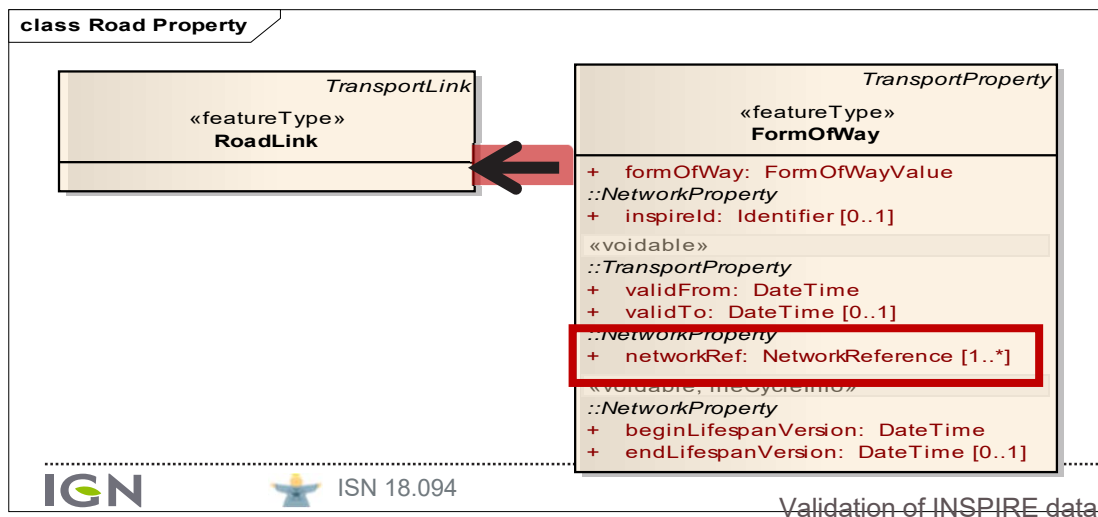


Simple URL request with  
**ResolveDepth = 1**  
=> Get features "Address"  
with their related  
AddressComponents

# Methodology: extract sample data

## 📍 Challenge : extract related INSPIRE data

- Case 3.A : feature type without geometry and big number of instances (e.g. most of Transport Properties)

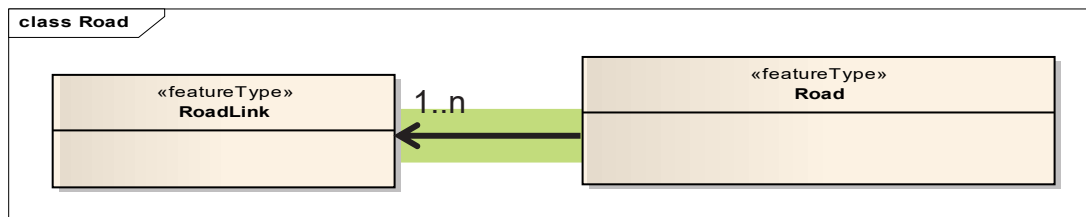


Complex HTML request  
“scanning” all instances of  
feature type FormOfway to  
get those associated to the  
RoadLinks of sample area  
=> doesn't work smoothly!

# Methodology: extract sample data

## 📍 Challenge : extract related INSPIRE data

- **Case 3.B : feature type without geometry and limited number of instances (n x 5 000)**



Sequence of simple URL requests  
⇒ Get all instances of “Roads” on whole France  
⇒ Make relevant selection on FME

# Results

 For each theme or sub-theme, we have found a few errors

- feature type forgotten (rare)
- transformation errors
- **errors in the matching table (rare)**
- “editorial” errors : writing conventions

Validation against source data : a necessary exercise



## CONCLUSIONS

# Main learnings

 Data should be validated:

- **Against INSPIRE => conformity**
- **Against source data => content quality**

 Validation against INSPIRE is quite easier with the INSPIRE validator than with our previous tool (XML Spy)

 **Main difficulty is to get INSPIRE data from WFS**

- **various strategies to extract relevant sample data in order to perform the tests**