

# TERRE-ISTEX :

## vers un modèle pour identifier des terrains d'études

**Atelier VADOR**

Toulouse, le 31 mai 2017

**GÉRIICO**



avoirs  
**STL**  
langage  
extés



# Les partenaires

TERRE-ISTEX



Groupe d'Études et de Recherche Interdisciplinaire en Information et Communication, Université de Lille

Chercheurs impliqués : Stéphane Chaudiron (PR), Bernard Jacquemin (MCF), Marta Severo (MCF), Joachim Schöpfel (MCF), Eric Kergosien (MCF)



Laboratoire Savoirs, Textes, Langage associé au CNRS

Chercheurs impliqués : Natalia Grabar



UMR Territoires, Environnement, Télédétection et Information Spatiale – TETIS, Montpellier, attachement GDR MAGIS

Chercheurs impliqués : Mathieu Roche, Maguelonne Teisseire, Jean-Philippe Tonneau



Laboratoire Informatique de l'Université de Pau et des Pays de l'Adour – LIUPPA, Pau

Chercheurs impliqués : Marie Noëlle Bessagnet (MCF), Annig Le Parc-Lacayrelle (MCF), Christian Sallaberry (MCF, HDR)



Atelier National de Reproduction des Thèses (ANRT), Lille

Chercheurs et personnels impliqués : Joachim Schöpfel (directeur), Rachid Berbache (informaticien, adjoint au directeur), Jérémy Berthe (technicien, chargé de projet).

# Objectifs du projet

- Identifier les territoires dans un corpus de document
  - territoire : ensemble d'informations géographiques associant entité spatiale + entité thématique + entité temporelle

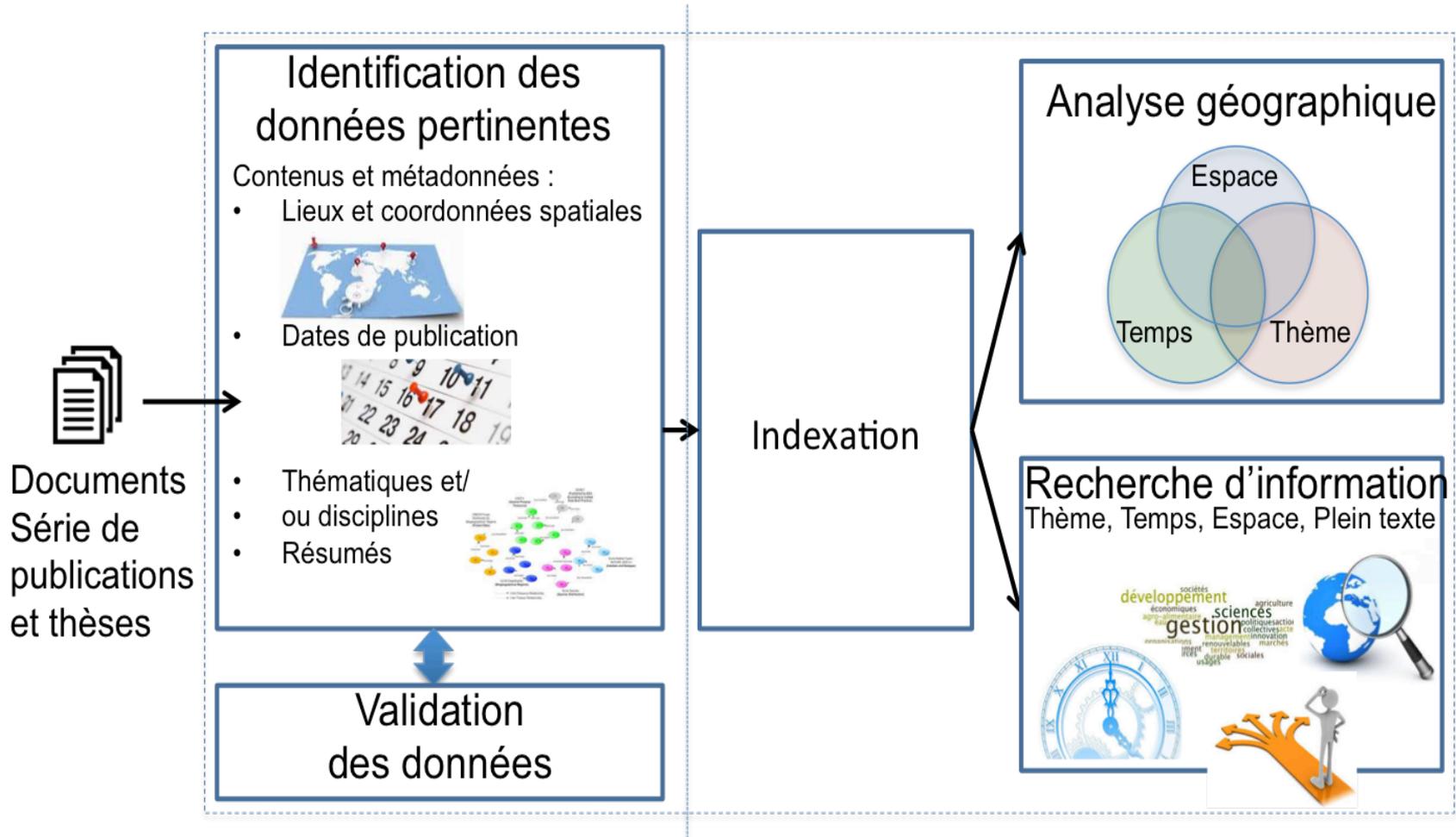
*Une étude du changement climatique menée dans le sud de Madagascar en 1981*

- Proposer une analyse géographique du corpus
  - comprendre quelles recherches, sur quels territoires, selon quelles approches et à quels moments

*Quel est le territoire d'études associé à la thématique "changement climatique" ?  
Pour les territoires Lac Aloatra (Madagascar) et Fleuve Sénégal (Sénégal), quelles sont les thématiques traitées ?*

- Recherche d'information multidimensionnelle
  - combinant temps, espace, thème

# Démarche générale

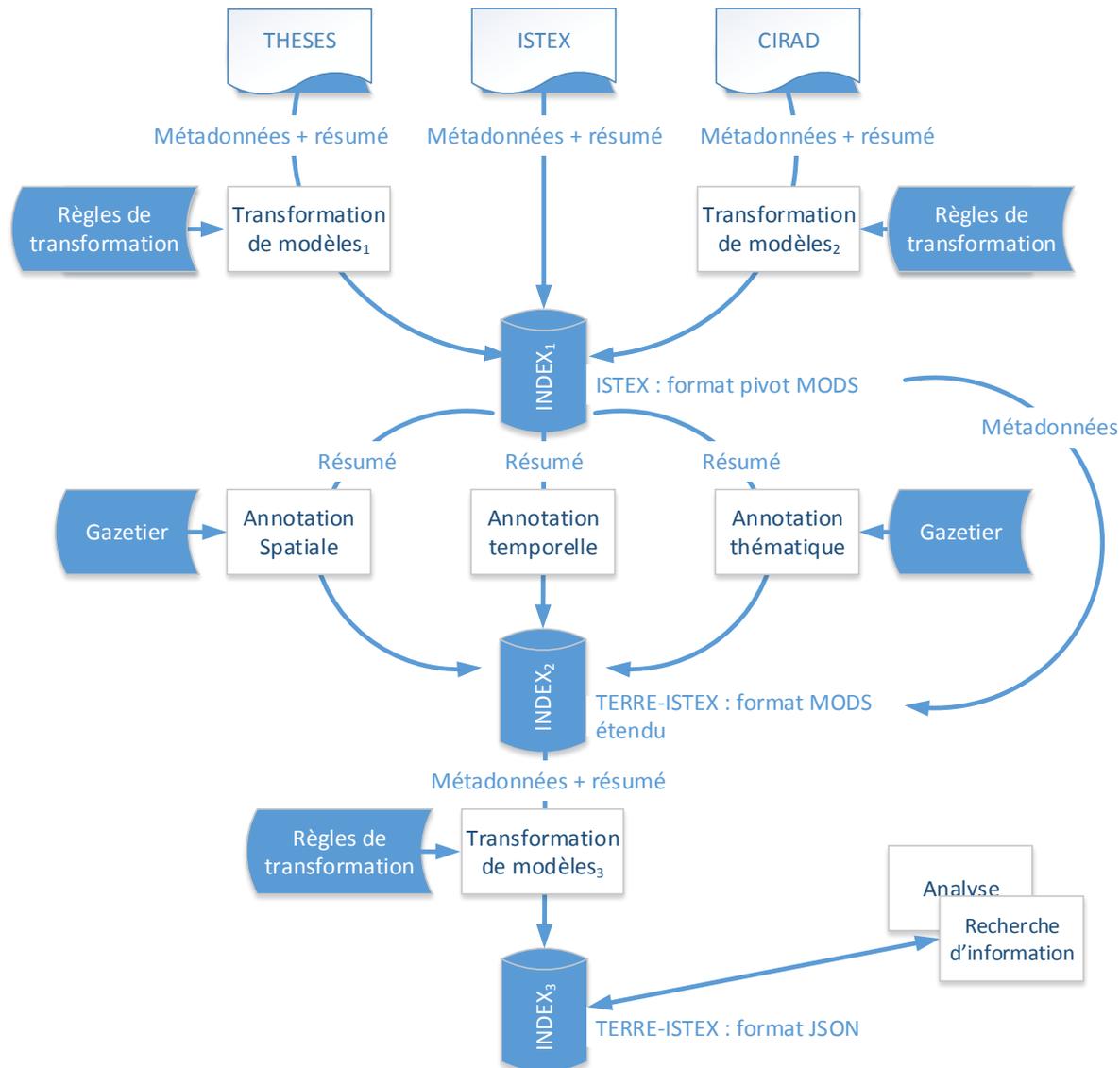


# Le corpus

- Données issues de la plateforme ISTEEX (170000 documents)
  - climate change, changement climatique, Sénégal, Senegal, Madagascar
  - méta-données : format MODS
  
- Données issues de la plateforme Agritrop (25000 documents)
  - Madagascar, Fleuve Sénégal
  - méta-données : format XML inspiré du Dublin Core
  
- Thèses de l'ANRT (400 thèses)
  - changement climatique
  - méta-données : RDF

→ Corpus **hétérogène** et **multilingue** (*français, anglais*)

# Mise en oeuvre de la démarche



# Modèle de données TERRE-ISTEX



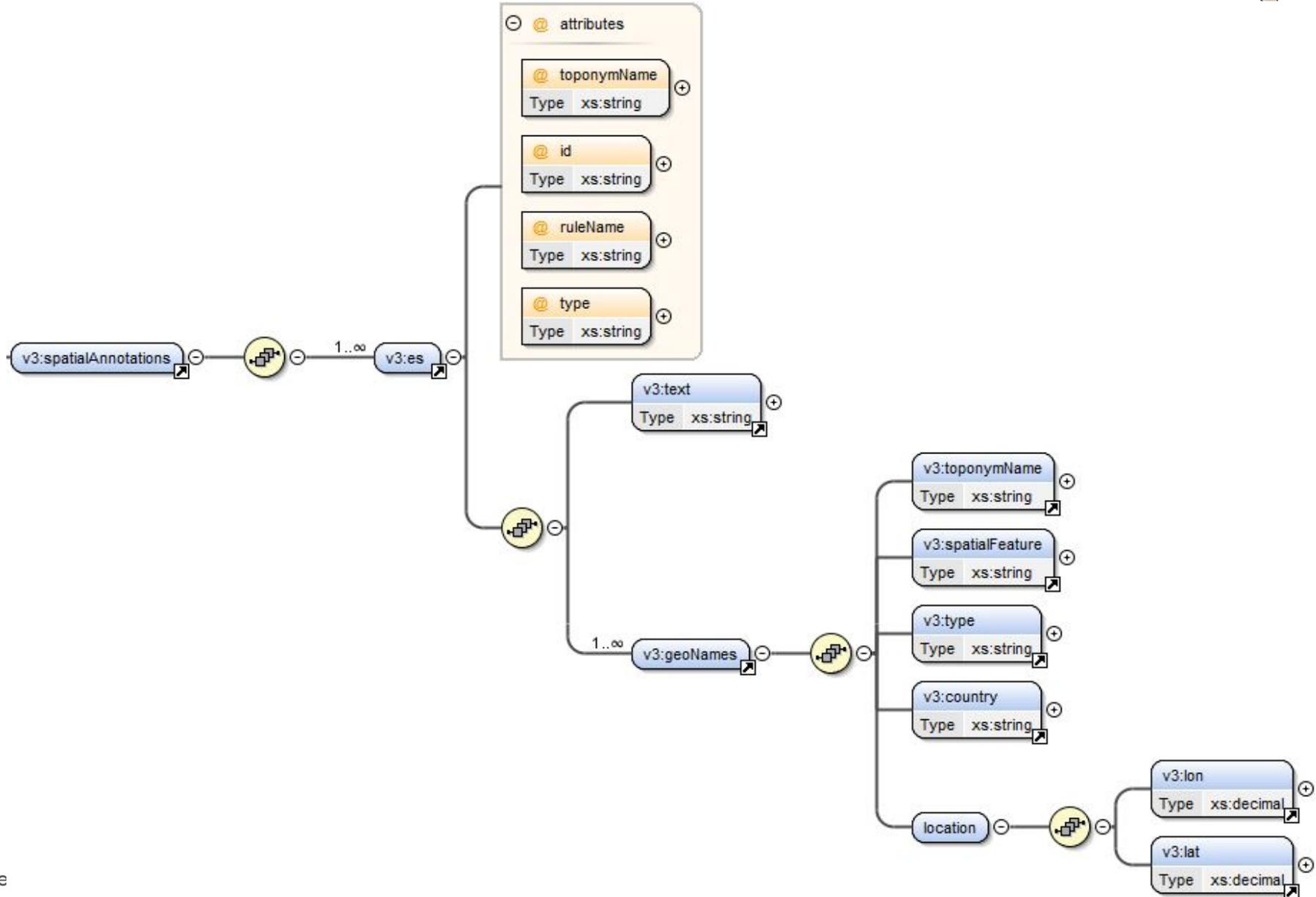
- Etend le format MODS
- Pourquoi MODS ?
  - utilisé sur la plateforme ISTEEX
  - approprié à la description de tout type de documents
  - plus riche que le Dublin Core
- Extension pour décrire
  - les informations spatiales
  - les informations temporelles
  - les informations thématiques

} annotées

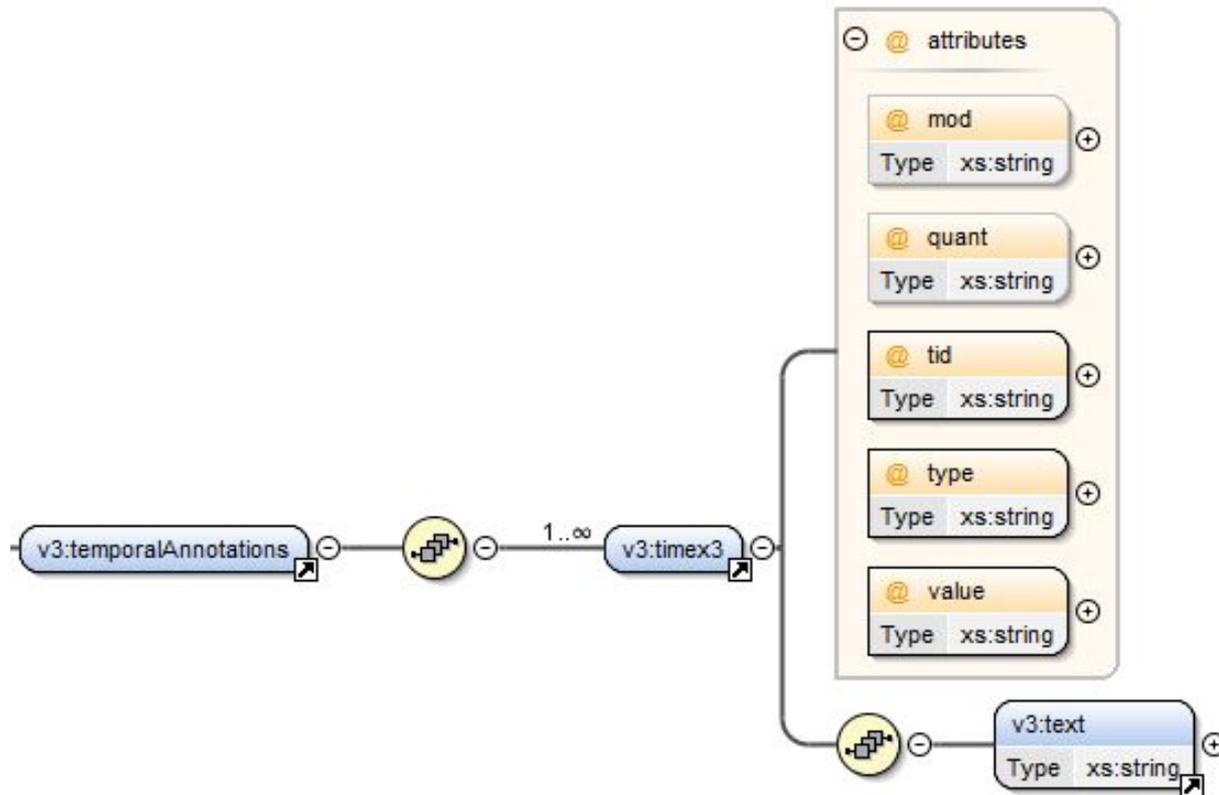
→ 3 balises

```
<spatialAnnotations>  
<temporalAnnotations>  
<thematicAnnotations>
```

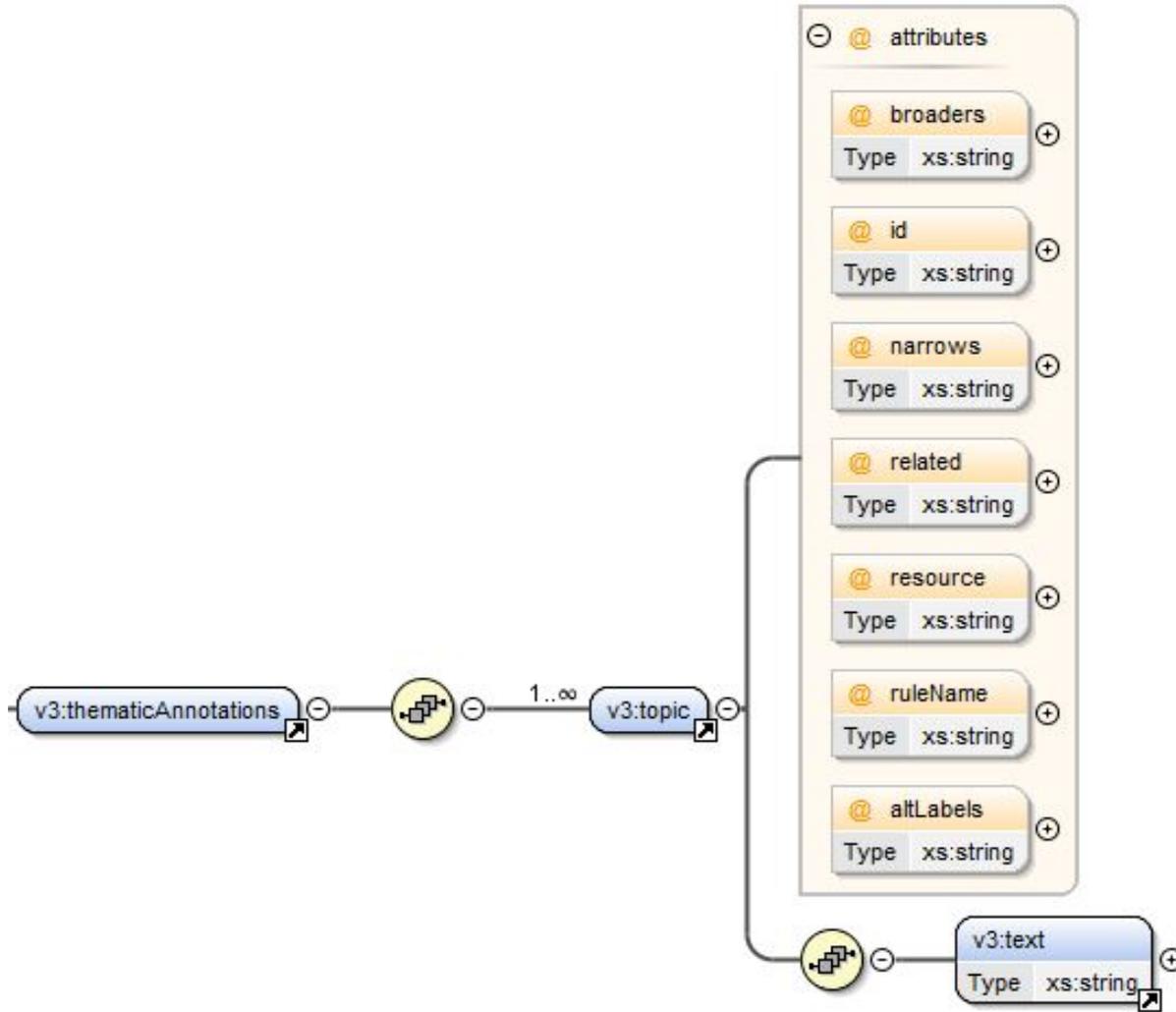
# <spatialAnnotations>



# <temporalAnnotations>



# <thematicAnnotations>



# Conclusion

## ■ Projet TERRE-ISTEX

### ■ Proposition d'une chaîne de traitement

- actuellement fin de construction de l'index
- appliquée uniquement sur les méta-données et les résumés des documents

### ■ Prochaines étapes

- Production d'analyses
- Développement d'un outil de recherche d'information multidimensionnelle
- Application de la chaîne de traitement aux documents en entier



*Merci de votre attention.*

Des questions ?

<https://terreistex.hypotheses.org>

