

# Encyclopedia2GeoKG : un outil pour l'extraction d'informations et la génération de graphes de connaissances géo-historiques à partir d'articles encyclopédiques

Bin Yang, Ludovic Moncla, Fabien Duchateau, Frédérique Laforest (prenom.nom@liris.cnrs.fr)  
INSA Lyon, CNRS, UCBL, LIRIS, UMR5205, F-69622 Villeurbanne, France

## 1. CONTEXTE & MÉTHODOLOGIE

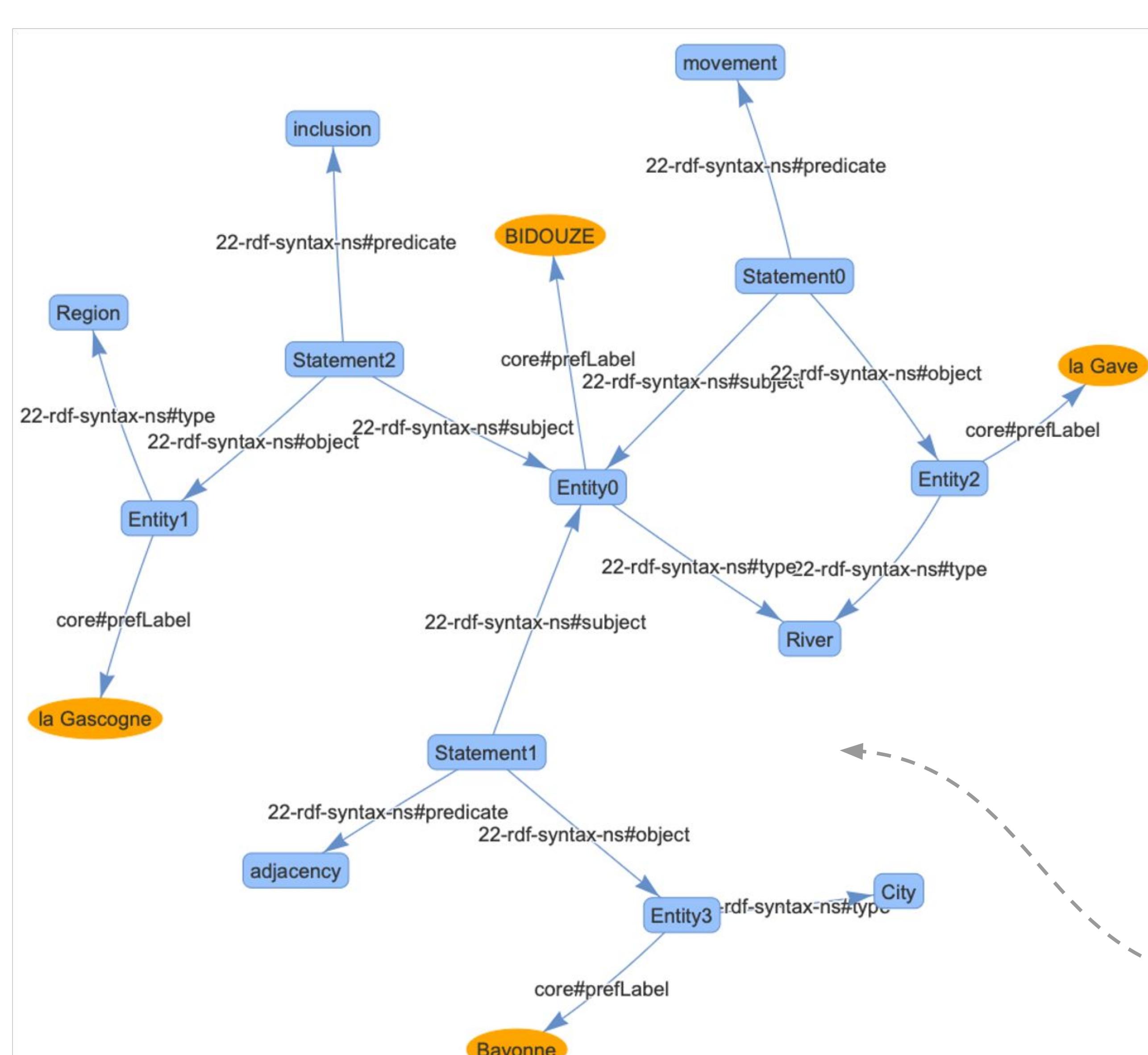
Le projet ECoDA a pour objectif de proposer une modélisation et une chaîne de traitement pour la construction automatique d'un **graphe de connaissances** géo-historiques à partir d'**articles encyclopédiques**. Le corpus d'étude est l'Encyclopédie de Diderot et d'Alembert (1751-1772).

### Modélisation

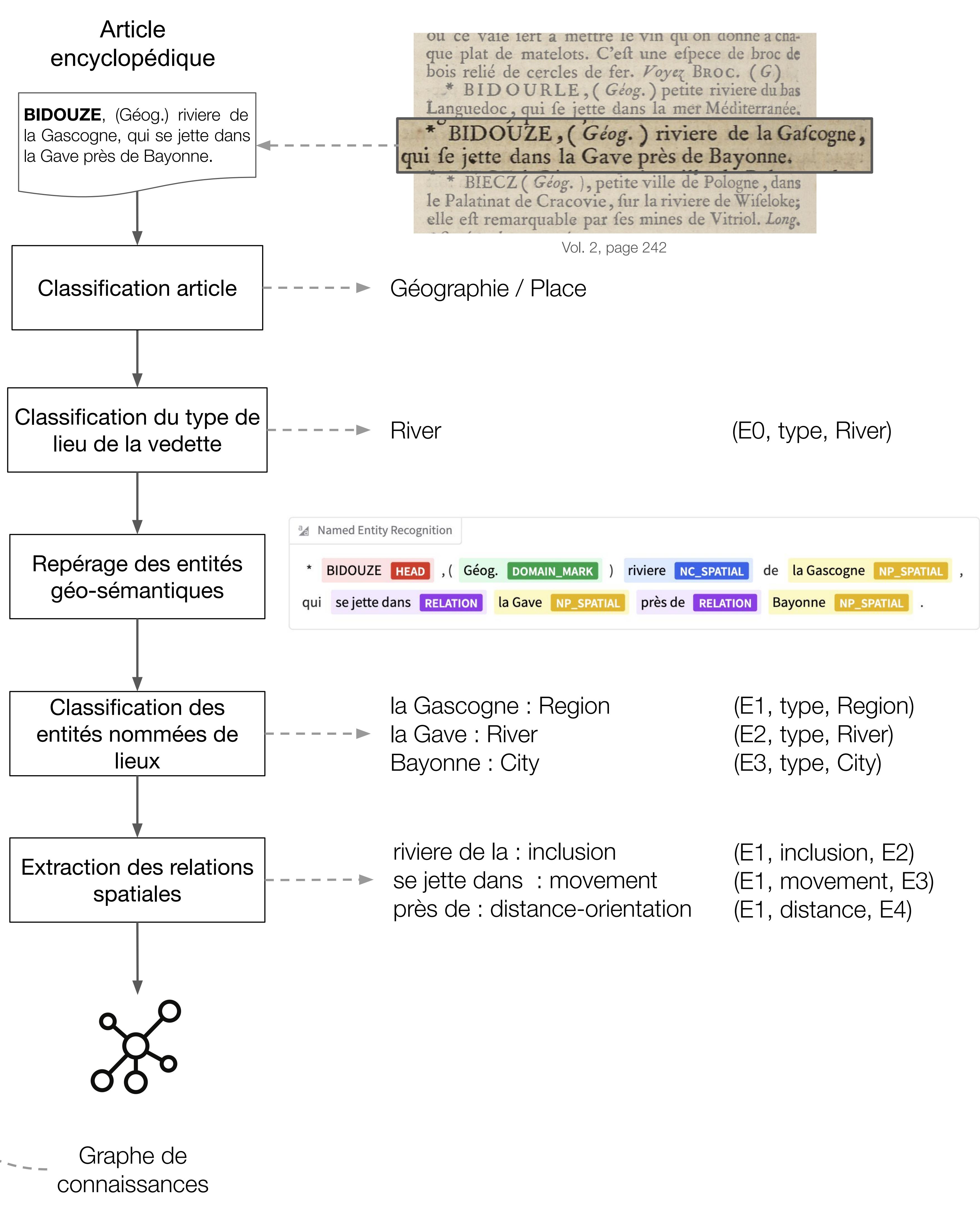
Une **ontologie spatiale** est proposée pour modéliser les **concepts géographiques** (e.g., ville, pays, rivière, etc.), **les relations** (e.g., inclusion, adjacence, etc.) et **les propriétés spatiales** (e.g., latitude, longitude, surface, etc.).

### Extraction d'informations et peuplement

Notre proposition repose sur une chaîne de traitement hybride (ci-contre) composée de différents types de **modèles de langue** (BERT, T5, GPT) fine-tunés pour des tâches de **repérage et classification** d'entités et de relations.



## 2. CHAÎNE DE TRAITEMENT



## 3. RÉSULTATS

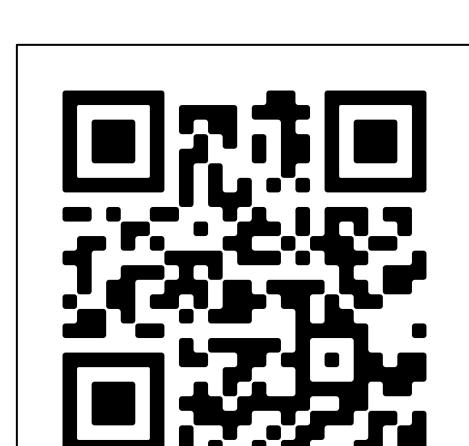
### Quelques chiffres

- 15 384 articles (domaine Géographie)
- 35 552 entités
- 46 584 relations

### Distribution des principales entités

City	17 159	48.3%
Region	5 167	14.5%
River	4 930	13.9%
Island	2 541	7.1%

### RESSOURCES



Modèles & jeux de données



Code



Démonstration

### Répartition des types de relations par type de sujets

