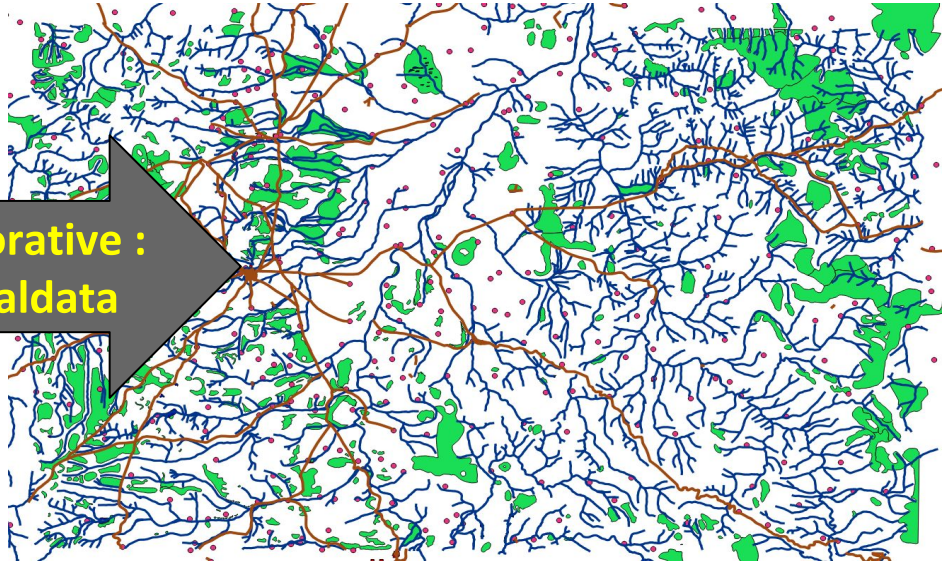
Création d'une base de connaissances sur les redécoupages administratifs durant la Révolution française: l'exemple des paroisses constitutionnelles

Antoine Keller (Ecole Navale), **Nathalie Abadie** (IGN-LaSTIG-Strudel), Bertrand Duménieu (EHESS-CRH), Stéphane Baciocchi (EHESS-CRH), Eric Kergosien (Univ. Lille3 - Geriico)

28èmes Journées de la Recherche IGN
Cité Descartes, Marne la Vallée
Session « Base de données et processus temporels »
18 avril 2019

Un référentiel géographique produit à partir de la carte de Cassini

Saisie collaborative :
Geohistoricaldata



Carte de Cassini, feuille 52,
Puy-de-Dôme, Clermont-Ferrand,
1759 -1777

Référentiel géographique vectoriel,
consultable et téléchargeable sur
geohistoricaldata.org

Les lieux nommés sur la carte de Cassini

Toponymes vectorisés :

- référentiel pour la fin du 18ème siècle (1759 -1777),
- la carte couvre l'ensemble du Royaume de France,
- vectorisation complète sur la feuille 52. Sur l'Auvergne, on dispose d'un référentiel de:
  - 1471 chefs-lieux
  - 4418 lieux-dits
- niveau de détail élevé : hameaux, fermes, moulins, vacheries, etc.

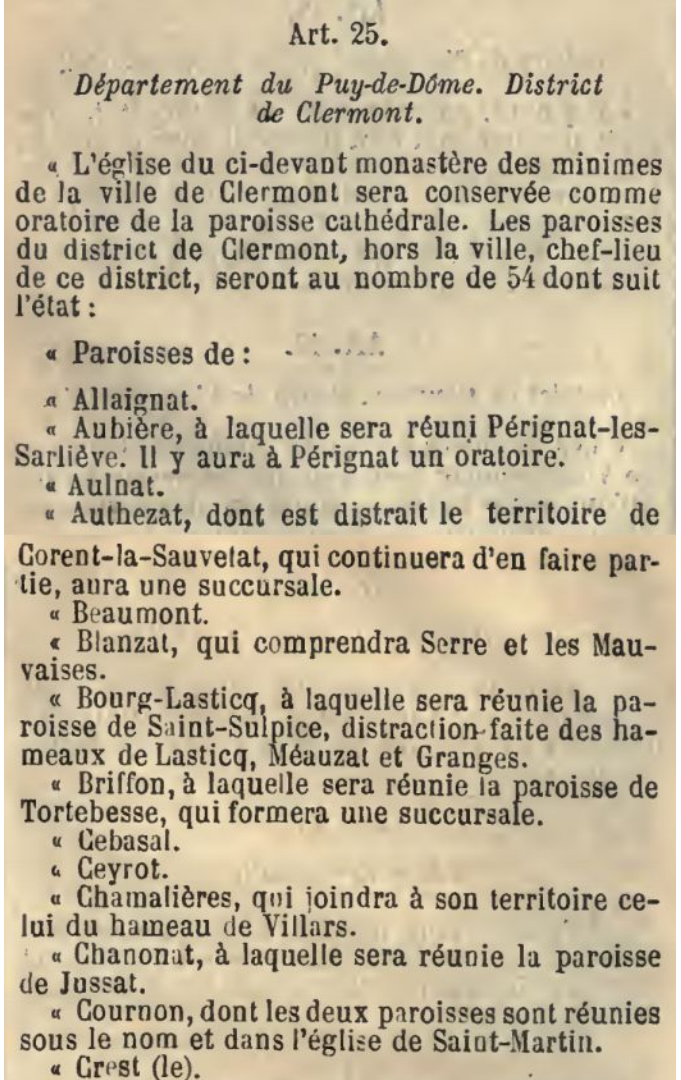
Carte de Cassini, feuille 52, Source: https://remonterletemps.ign.fr/

Les décrets de circonscription des paroisses de l'Assemblée Nationale

Décrets structurés par articles:

- Organisation géographique des textes: un article = un district
- Décrivent les paroisses constitutionnelles en listant les paroisses de l'ancien régime et les lieux-dits qui les composent
- Précisent l'organisation du culte dans ces paroisses

Pour le Puy-de-Dôme : ~ 10 pages
Pour la France entière : ~1000 pages
→ Traitement manuel peu envisageable



Art. 25.

Département du Puy-de-Dôme. District de Clermont.

« L'église du ci-devant monastère des minimes de la ville de Clermont sera conservée comme oratoire de la paroisse cathédrale. Les paroisses du district de Clermont, hors la ville, chef-lieu de ce district, seront au nombre de 54 dont suit l'état :

« Paroisses de :

« Allaignat.

« Aubière, à laquelle sera réuni Pérignat-les-Sarliève. Il y aura à Pérignat un oratoire.

« Aulnat.

« Authezat, dont est distrait le territoire de Corent-la-Sauvetat, qui continuera d'en faire partie, aura une succursale.

« Beaumont.

« Blanzat, qui comprendra Serre et les Mauvaises.

« Bourg-Lasticq, à laquelle sera réunie la paroisse de Saint-Sulpice, distraction faite des hameaux de Lasticq, Méauzat et Granges.

« Briffon, à laquelle sera réunie la paroisse de Tortebesse, qui formera une succursale.

« Cebasal.

« Ceyrot.

« Chamalières, qui joindra à son territoire celui du hameau de Villars.

« Chanonat, à laquelle sera réunie la paroisse de Jussat.

« Cournon, dont les deux paroisses sont réunies sous le nom et dans l'église de Saint-Martin.

« Crest (le).

Extrait de l'article 25 - 1er juin 1791

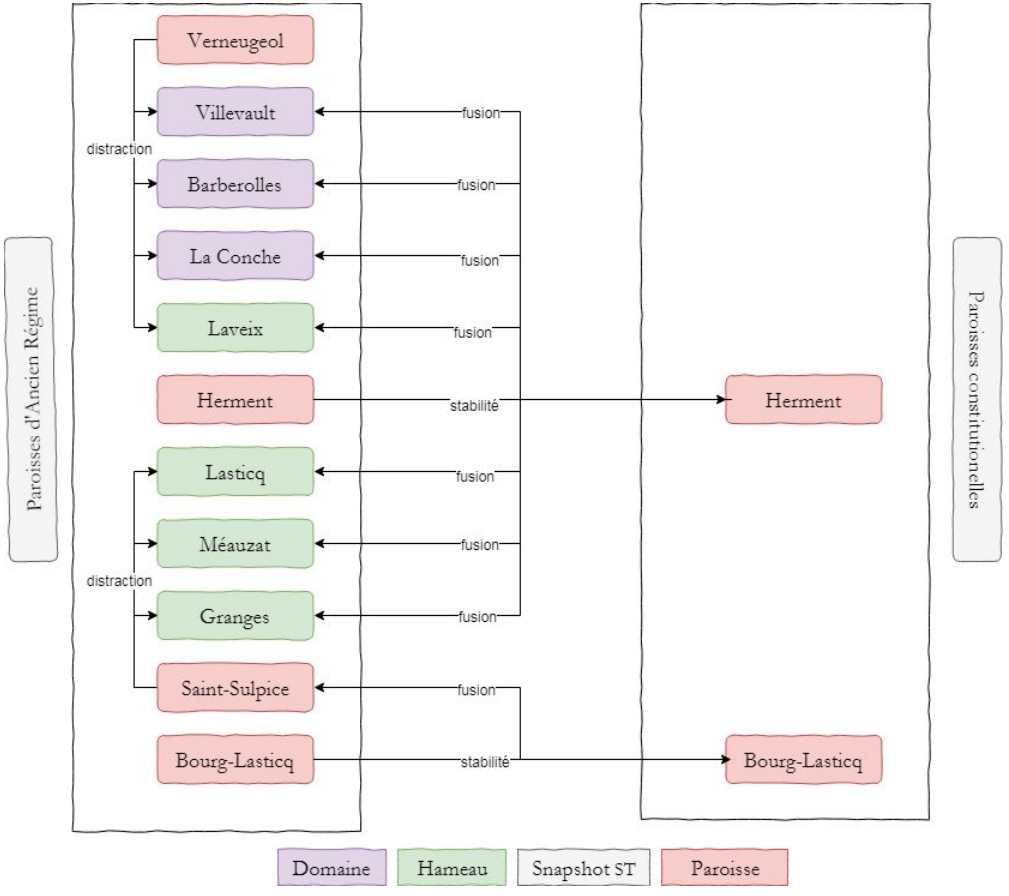
Créer une base de connaissances sur la réorganisation des paroisses de l'ancien régime pour créer les paroisses constitutionnelles ?

Hypothèses et questions de recherche

Un graphe des évolutions spatio-temporelles du découpage administratif construit à partir des décrets

« Bourg-Lasticq, à laquelle sera réunie la paroisse de Saint-Sulpice, distraction faite des hameaux de Lasticq, Méauzat et Granges.

« Herment, qui comprendra, outre son territoire, les hameaux de Lasticq, Méauzat et Granges, distraits de Saint-Sulpice ; et le hameau de Laveix, les domaines de la Conche, Barberolles et Villevault, distraits de la paroisse de Verneugeol.



Un graphe des évolutions spatio-temporelles du découpage administratif construit à partir des décrets

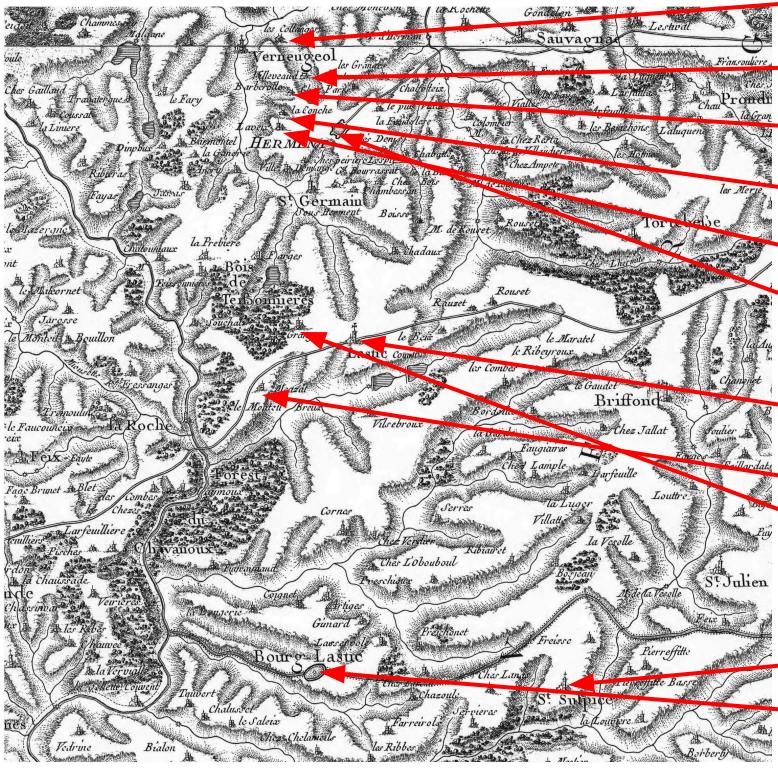
Extraction d'informations à partir de textes:

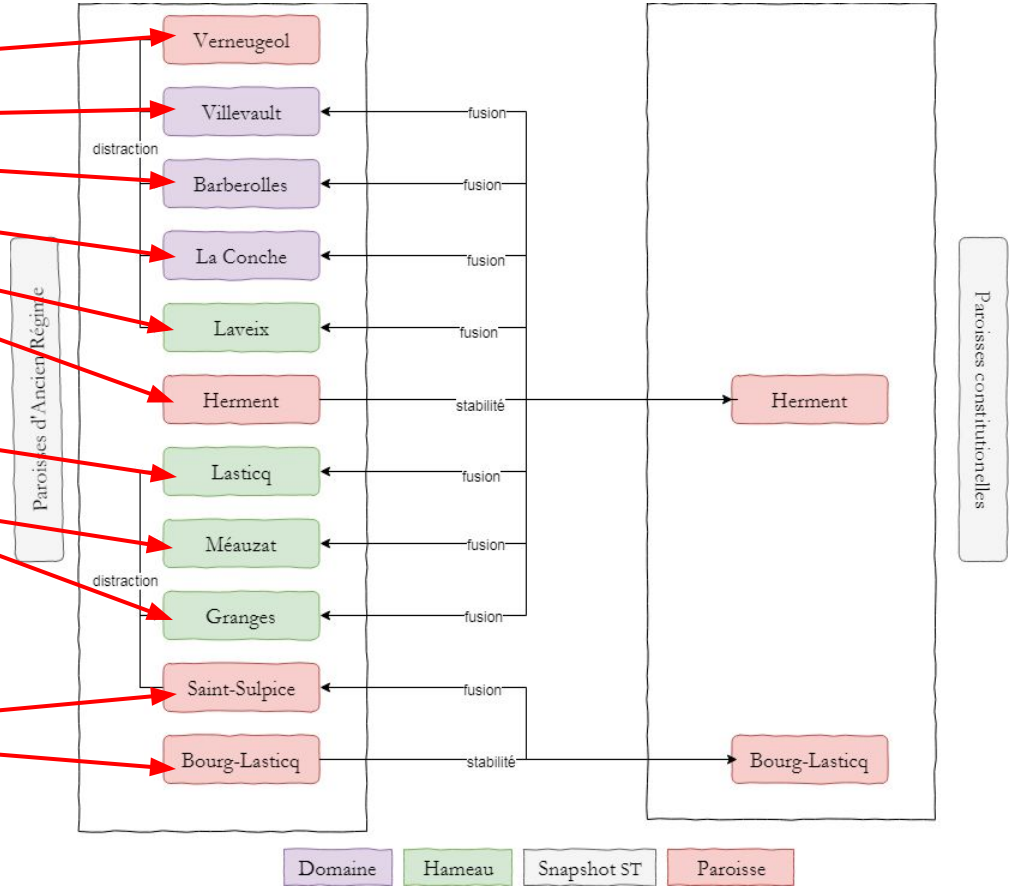
- Textes anciens + textes de lois
- Entités Spatiales Nommées (ESN) historiques + un niveau de détail très fin
- Relations spatio-temporelles entre ESN

Structuration des informations extraites (et inférences):

- Informations incomplètes et hétérogènes → Besoin d'un modèle de données qui autorise des informations lacunaires et des propriétés hétérogènes pour un même type d'entités.

Un graphe des évolutions spatio-temporelles du découpage administratif géoréférencé à l'aide du référentiel "Cassini"





Un graphe des évolutions spatio-temporelles du découpage administratif géoréférencé à l'aide du référentiel "Cassini"

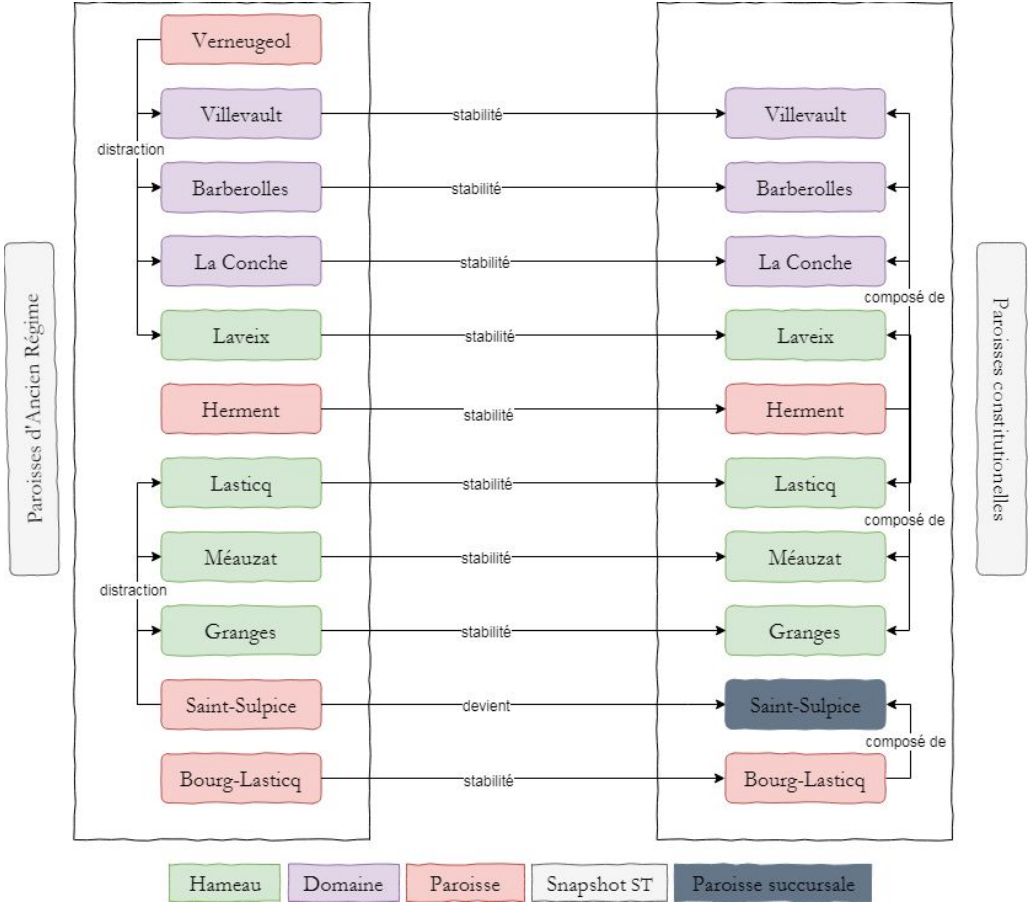
Désambiguisation d'ESN extraites de textes anciens:

- Référentiel de toponymes adapté (même actualité, même niveau de détail, etc.)
- Beaucoup de micro-toponymes: forte densité spatiale et beaucoup d'homonymes
- Graphie non fixée

Un graphe des évolutions spatio-temporelles du découpage administratif enrichi à l'aide d'inférences

« Bourg-Lasticq, à laquelle sera réunie la paroisse de Saint-Sulpice, distraction faite des hameaux de Lasticq, Méauzat et Granges.

« Herment, qui comprendra, outre son territoire, les hameaux de Lasticq, Méauzat et Granges, distraits de Saint-Sulpice ; et le hameau de Laveix, les domaines de la Conche, Barberolles et Villevault, distraits de la paroisse de Verneugeol.



Un graphe des évolutions spatio-temporelles du découpage administratif enrichi à l'aide d'inférences

Structuration des informations extraites et inférences :

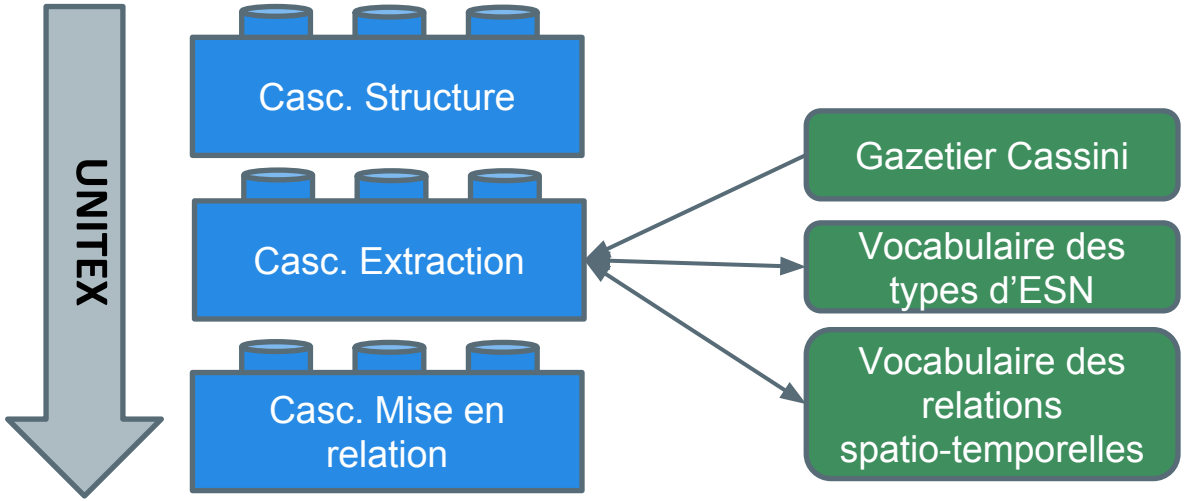
- Informations implicites → besoin de formaliser des connaissances pour vérifier la cohérence des faits extraits des textes et déduire de nouveaux faits à partir des faits extraits.

Approche proposée pour l’extraction des informations à partir des textes

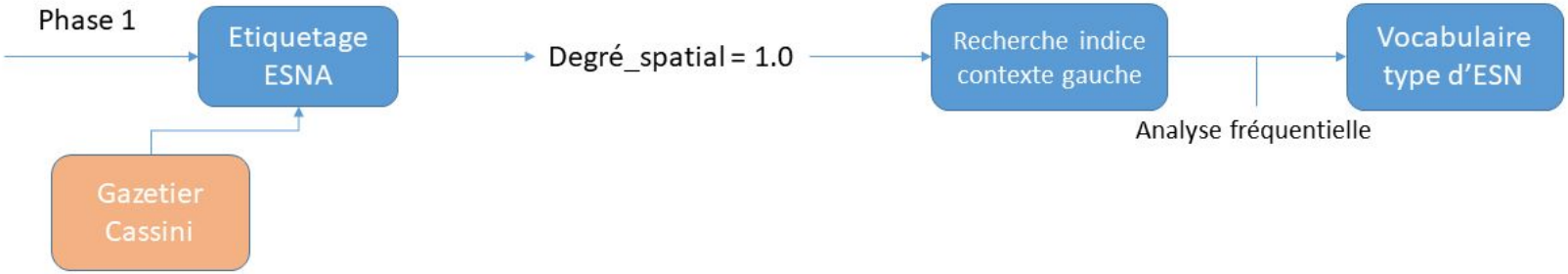
Extraction et structuration des informations géohistoriques à partir des textes des décrets

Approche à base de règles :

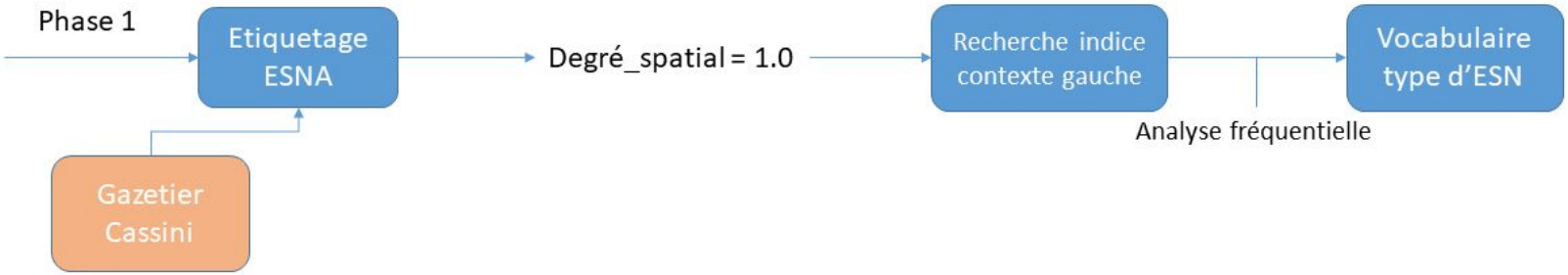
- Corpus de petite taille
- Absence de corpus d'entraînement annoté
- Textes fortement structurés
- Syntaxe régulière, vocabulaire quasi contrôlé
- Référentiel de toponymes disponible



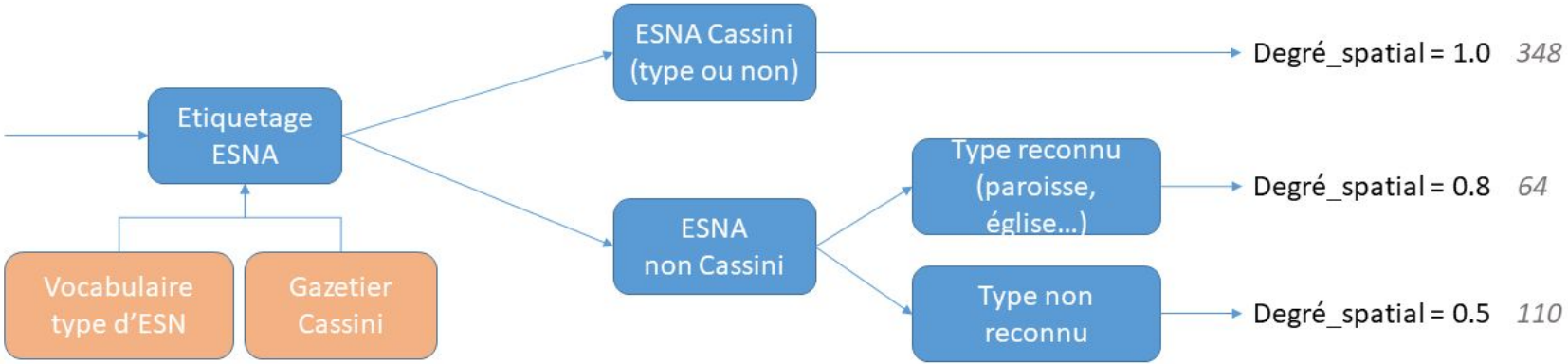
Etiquetage des Entités Spatiales Nommées “Cassini” et extraction de leurs types



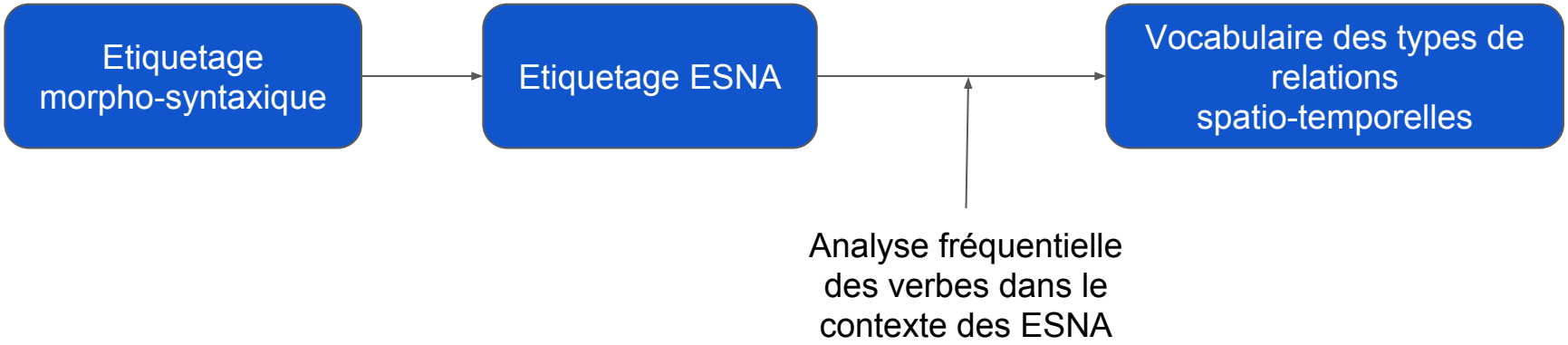
Extraction de l'ensemble des Entités Spatiales Nommées



Phase 2 Adaptation de l'approche d'extraction d'ESA proposée dans (Moncla, 2015)



Extraction des évolutions spatio-temporelles des Entités Spatiales Nommées



Extraction des évolutions spatio-temporelles des Entités Spatiales Nommées

Processus spatio-temporel*	**Evolution**: concerne une entité spatiale	**Stabilité**	« Paroisses de: - Allaignat. »
		Transformation	« Il y aura, à Peyrol, une succursale .»
		Suppression	« Les paroisses de Menestrol, de Saint-Jean et de Pessat-Villette sont supprimées. »
	Restructuration: concerne plusieurs entités spatiales	**Fusion**	« Aubière, à laquelle sera réunie Pérignat-les-Sarlève. »
		Distraction	« Lussat, qui conservera son territoire actuel, excepté le hameau de Lignat .»

* Sous-ensemble de la typologie proposée dans Duménieu B., « Un système d'information géographique pour le suivi d'objet historiques urbains à travers l'espace et le temps » Thèse de doctorat de l'EHESS (2015)
Exemples tirés de l'article 25 - 1er juin 1791

Evaluation des résultats d'extraction

Extraction d'ESN :

398 entités extraites uniques / 402 référencées manuellement

Précision : **92** % Rappel : **90** % F-mesure = **91** %

Extraction des relations :

73.9% des ESN sont en relation (fusion : 36.8% | séparation 5.0% | évolution : 32.2%)

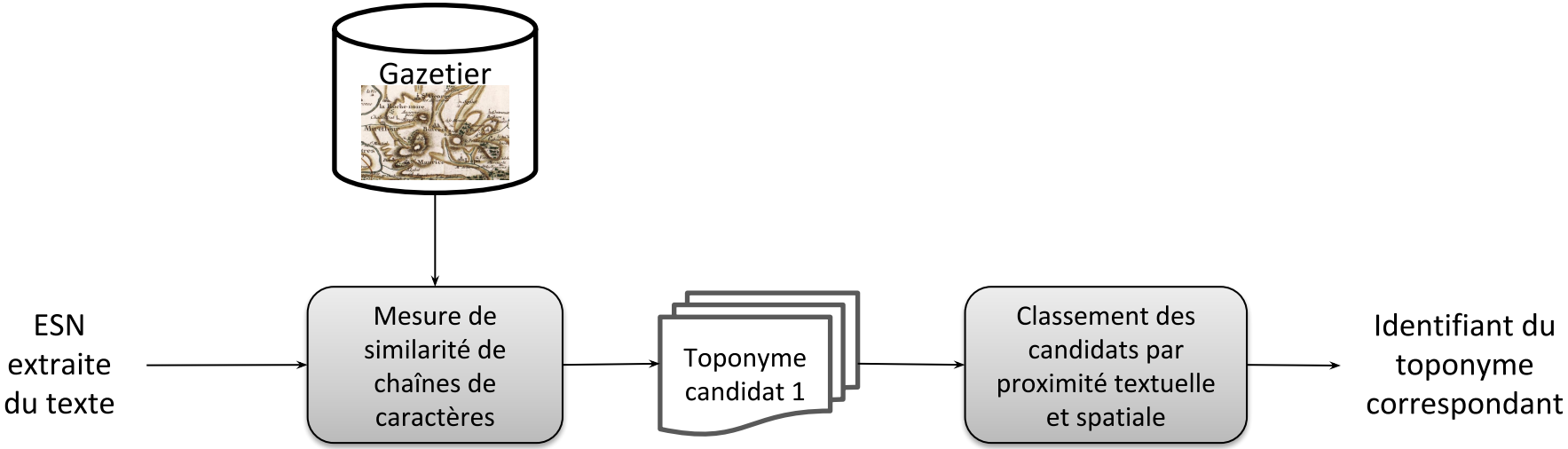
Sources d'erreurs :

Textes effacés, OCR, grammaires d'extraction.

Approche proposée pour le géoréférencement des informations extraites

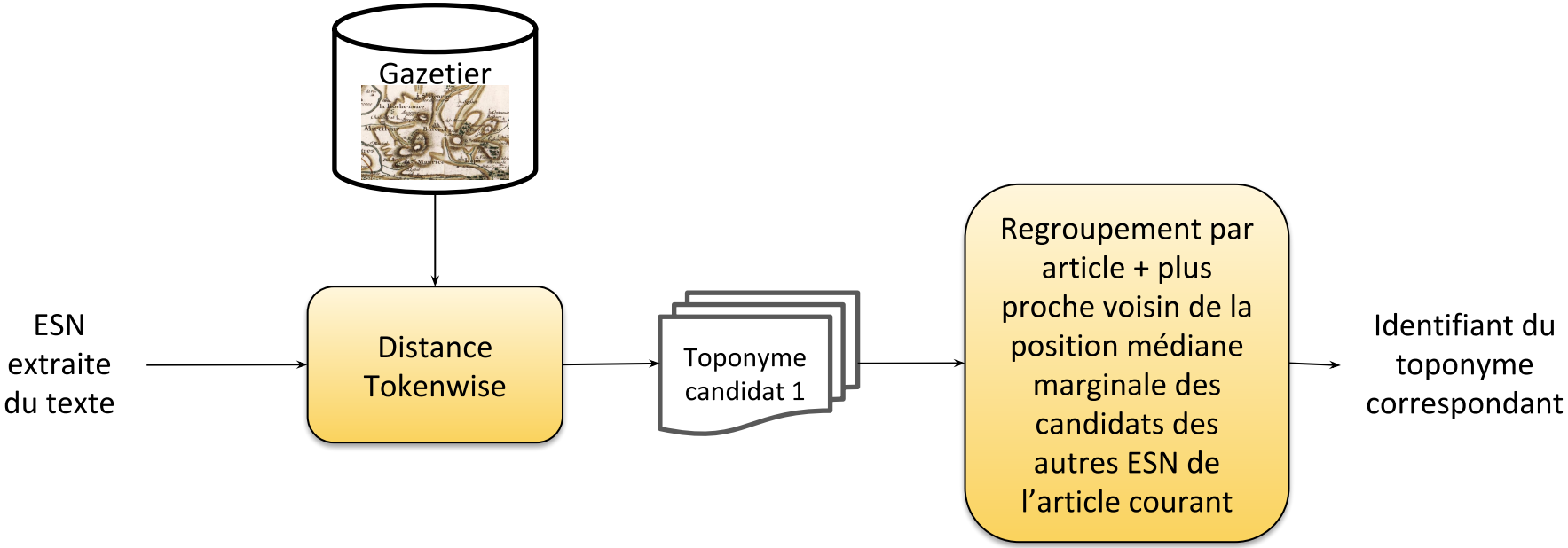
Désambiguïsation et géoréférencement des ESN à l'aide du référentiel Cassini

Hypothèse classique : Entités proches dans les textes <=> Lieux proches sur le terrain



Désambiguïsation et géoréférencement des ESN à l'aide du référentiel Cassini

Hypothèse classique : Entités proches dans les textes <=> Lieux proches sur le terrain



Désambiguïsation et géoréférencement des ESN à l'aide du référentiel Cassini: évaluation

	Score obtenu
Sélection des candidats — Cardinalité Moyenne	1.78
Précision Candidats	89 %
Rappel Candidats	71 %
Précision Nil	12 %
Rappel Nil	58 %
Calcul du meilleur candidat — Exactitude de désambiguisation	85 %
Exactitude globale	61 %

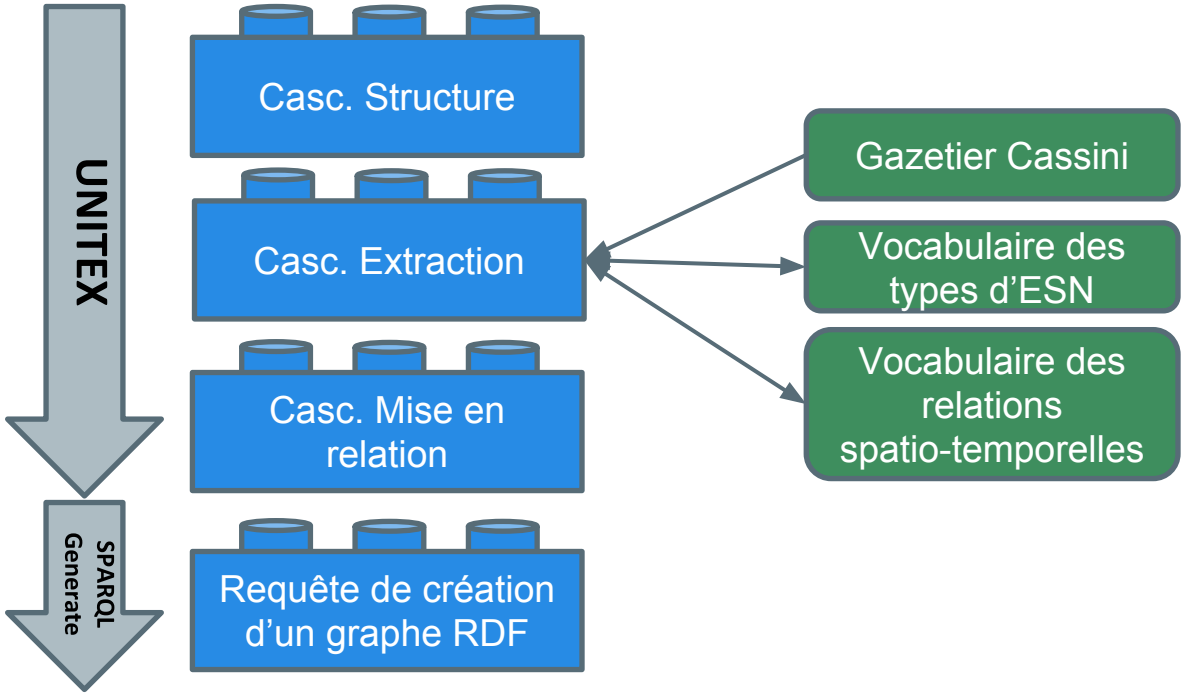
Approche proposée pour la structuration des informations extraites et l’explicitation des informations implicites

Structuration des informations extraites et inférences

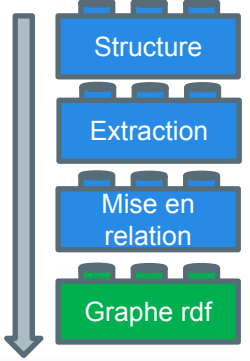
Objectifs:

- Structurer et typer les informations extraites du texte: modèle de graphe spatio-temporel (snapshot), schéma de données adaptable
- Vérifier leur cohérence
- Expliciter les informations implicites

→ Structurer les informations en RDF et formaliser les connaissances nécessaires à la vérification de cohérence et aux inférences au sein d'une ontologie

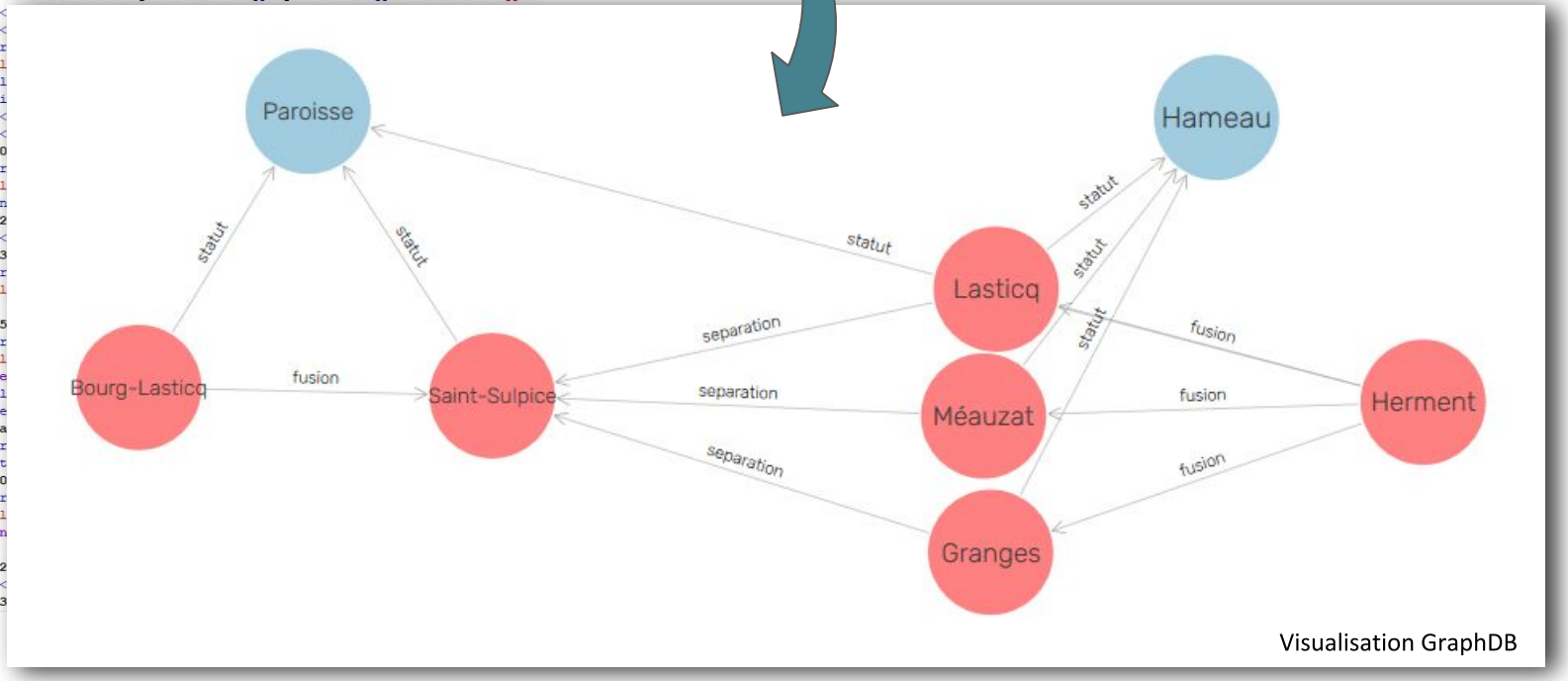


Création du graphe RDF

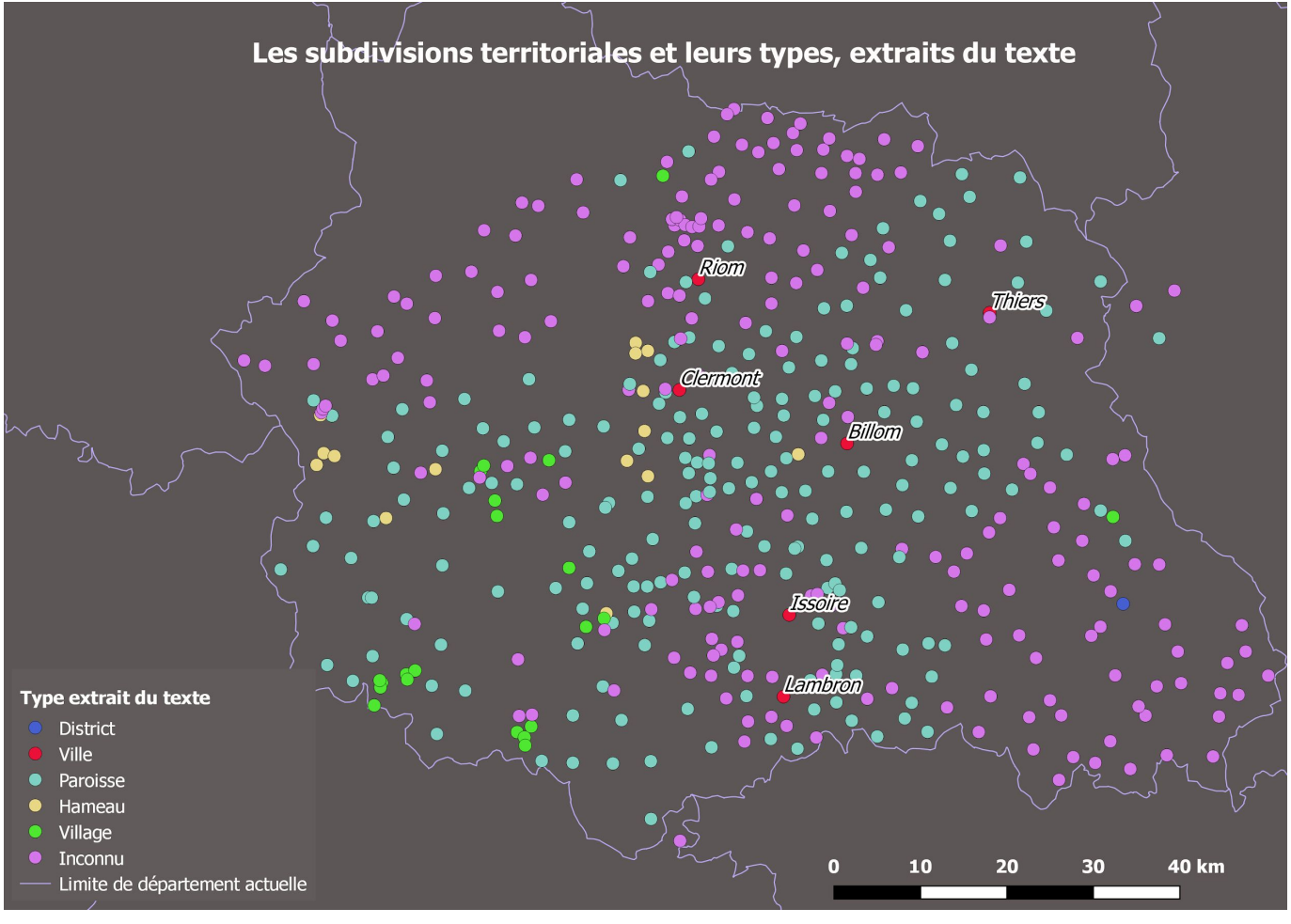



```
<tei><article><headart><csc><form>Art. 25.</form><code>25</code></csc></headart>
<localisation><csc><form>Département du Puy-de-Dôme. District de <esn spatial_degree="1.0"><id>1</id>
Clermont</label></esn></form><code>Puy-de-Dôme</code><code><esn spatial_degree="1.0"><id>2</id><lab
Clermont</label></esn></code></csc></localisation>.
<headerlist>« L'église du ci-devant monastère des minimes de la <esn spatial_degree="1.0"><id>3</id><label>
Clermont</label><type>ville</type></esn> <continuation>conservée</continuation> comme oratoire de la
paroisse cathédrale. Les paroisses du <esn spatial_degree="1.0"><id>4</id><label>Clermont</label><type>
district</type></esn>, hors la ville, chef-lieu de ce district, seront <nombreParoisse><csc><form>au nombre
de 54</form><code>54</code></csc></nombreParoisse> dont suit l'état : « Paroisses de :</headerlist>
<item><esn spatial_degree="0.5"><id>5</id><label>Allaignat</label><type>paroisse</type><evolution type=
"stabilité">stable</evolution></esn>.<
<item><esn spatial_degree="1.0"><id>6<
"stabilité">stable</evolution></esn><r
</range></restructuration><esn spatial
<esn spatial_degree="0.5"><id>8</id><l
</sujet><nouv>oratoire</nouv></evoluti
<item><esn spatial_degree="1.0"><id>9<
"stabilité">stable</evolution></esn>.<
<item><esn spatial_degree="1.0"><id>10
"stabilité">stable</evolution></esn><r
</range></restructuration><esn spatial
<continuation>continuera</continuation
<item><esn spatial_degree="1.0"><id>12
"stabilité">stable</evolution></esn>.<
<item><esn spatial_degree="1.0"><id>13
"stabilité">stable</evolution></esn><r
</range></restructuration><esn spatial
</item>
<item><esn spatial_degree="1.0"><id>15
"stabilité">stable</evolution></esn><r
</range></restructuration><esn spatial
</type></esn><restructuration type="se
spatial_degree="0.8"><id>17</id><label
"separation"><domain>18</domain><range
</id><label>Méauzat</label><type>hamea
</domain><range>16</range></restructur
</label><type>hameau</type></esn>.</it
<item><esn spatial_degree="0.5"><id>20
"stabilité">stable</evolution></esn><r
</range></restructuration><esn spatial
</type><evolution type="transformation
</item>
<item><esn spatial_degree="0.5"><id>22
"stabilité">stable</evolution></esn>.<
<item><esn spatial_degree="0.5"><id>23
```

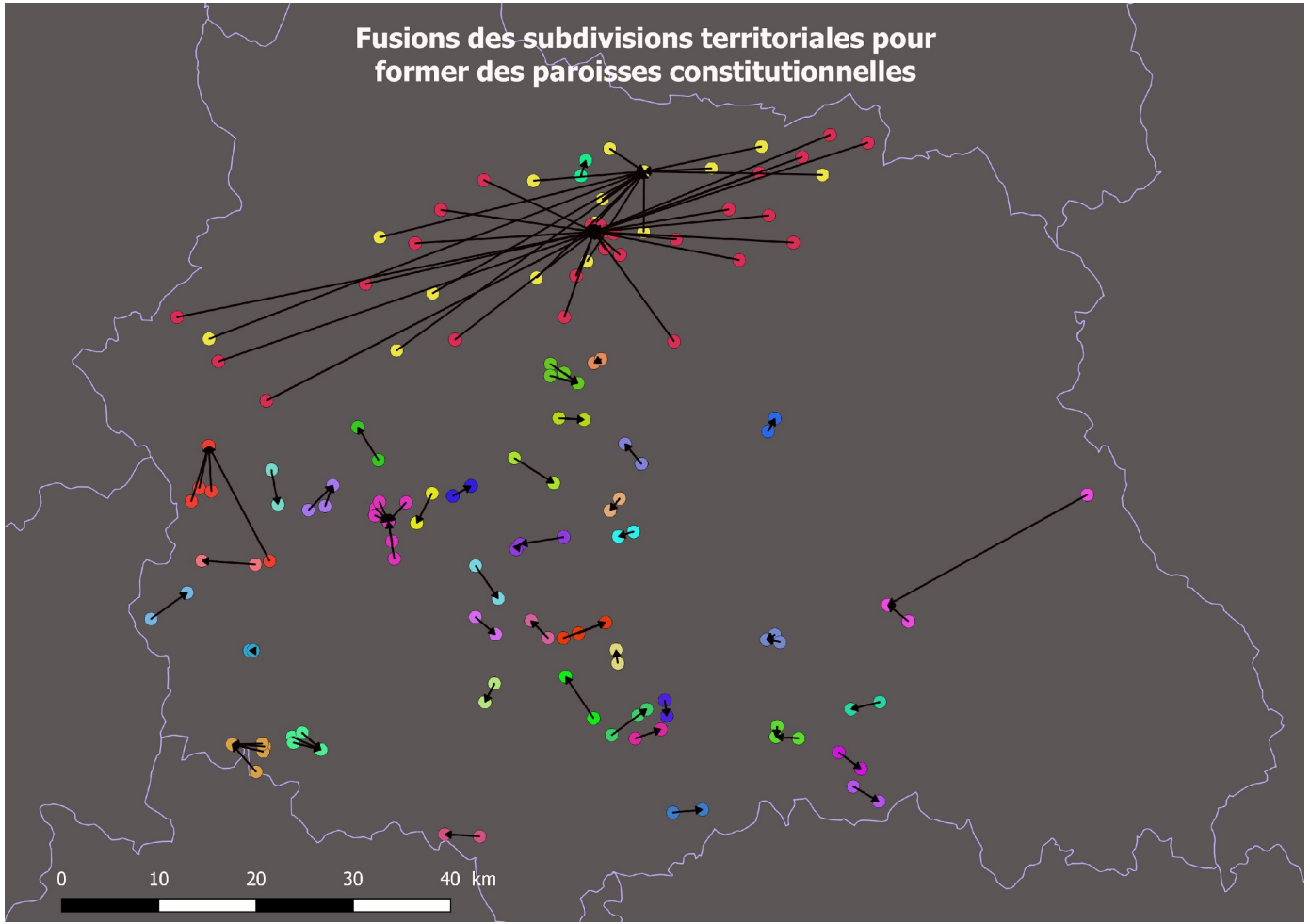

SPARQL-Generate



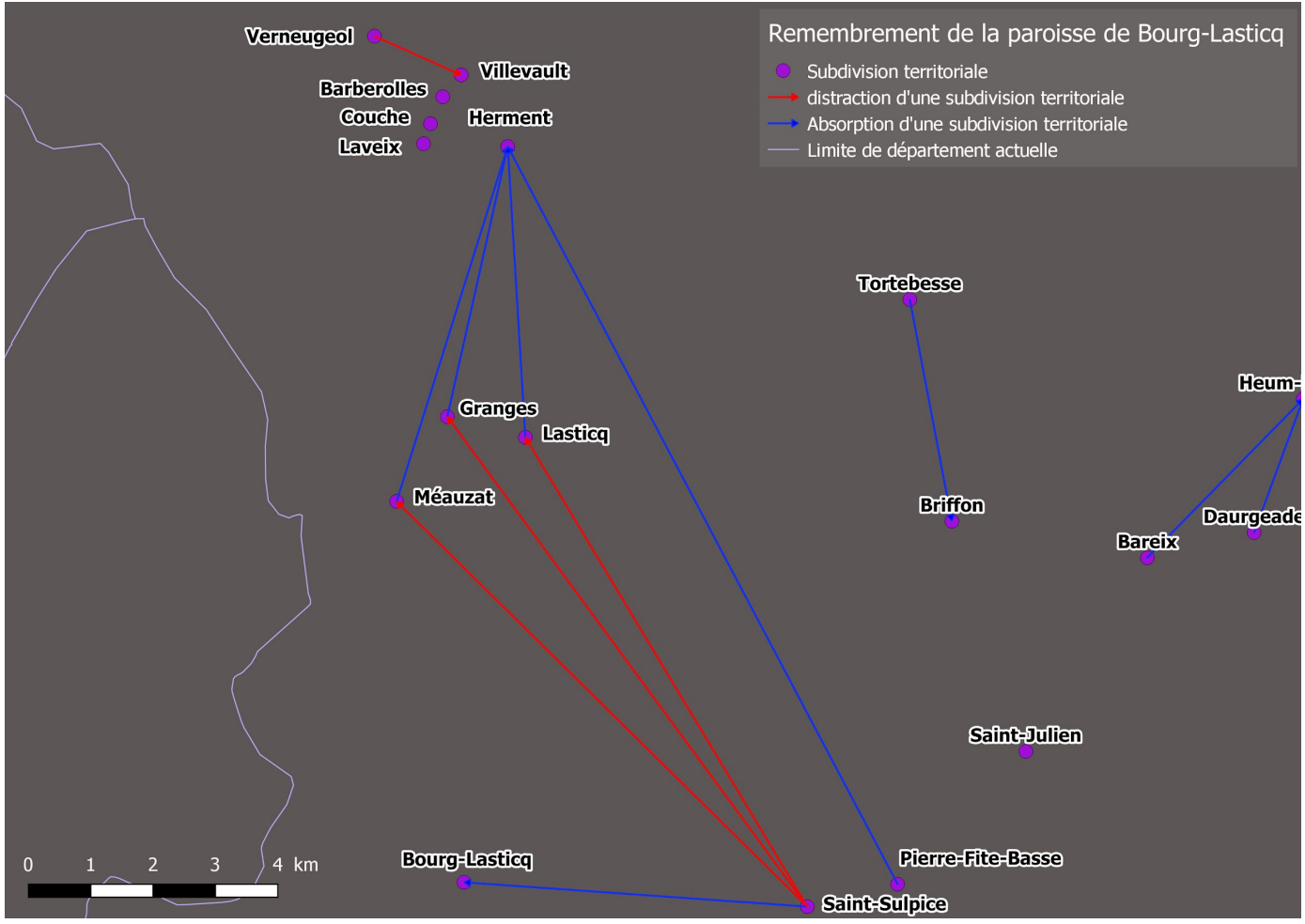
Les ESN extraites du textes, géoréférencées avec le mapping de référence



Les fusions de paroisses, géoréférencées avec le mapping de référence



Quid de Bourg-Lasticq?...



Formalisation des conditions nécessaires et suffisantes pour qu'une entité soit considérée comme une paroisse de l'ancien régime

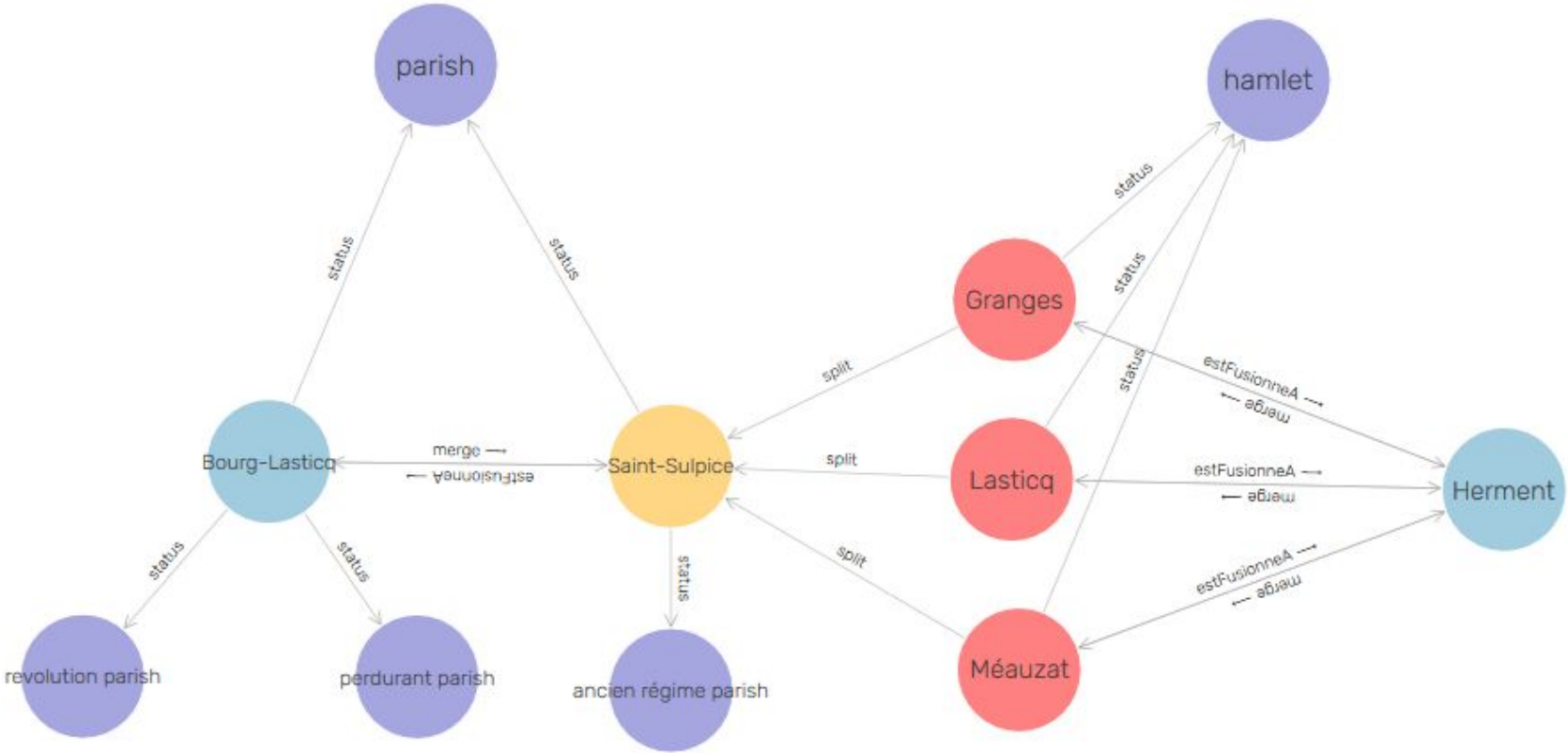
Class: prg:ParoisseAncienRegime

Annotations:
rdfs:comment "Cette classe contient l'ensemble des entités spatiales nommées extraites du texte et identifiées comme des paroisses de l'ancien régime."@fr,
rdfs:label « Ancien régime parish"@en,
rdfs:label "Paroisse de l'ancien régime"@fr

EquivalentTo:
(prg:statut value subterr:Paroisse) and (prg:hasDisappeared value true),
prg:lieuDeCulte some (prg:LieuDeCulte and (prg:ancienStatut value egl:Eglise)),
(prg:estFusionneA some (prg:ParoisseConstitutionnelle or prg:ParoissePerdurante)) and
(prg:statut value subterr:Paroisse),
prg:statut value subterr:ParoisseAncienRegime

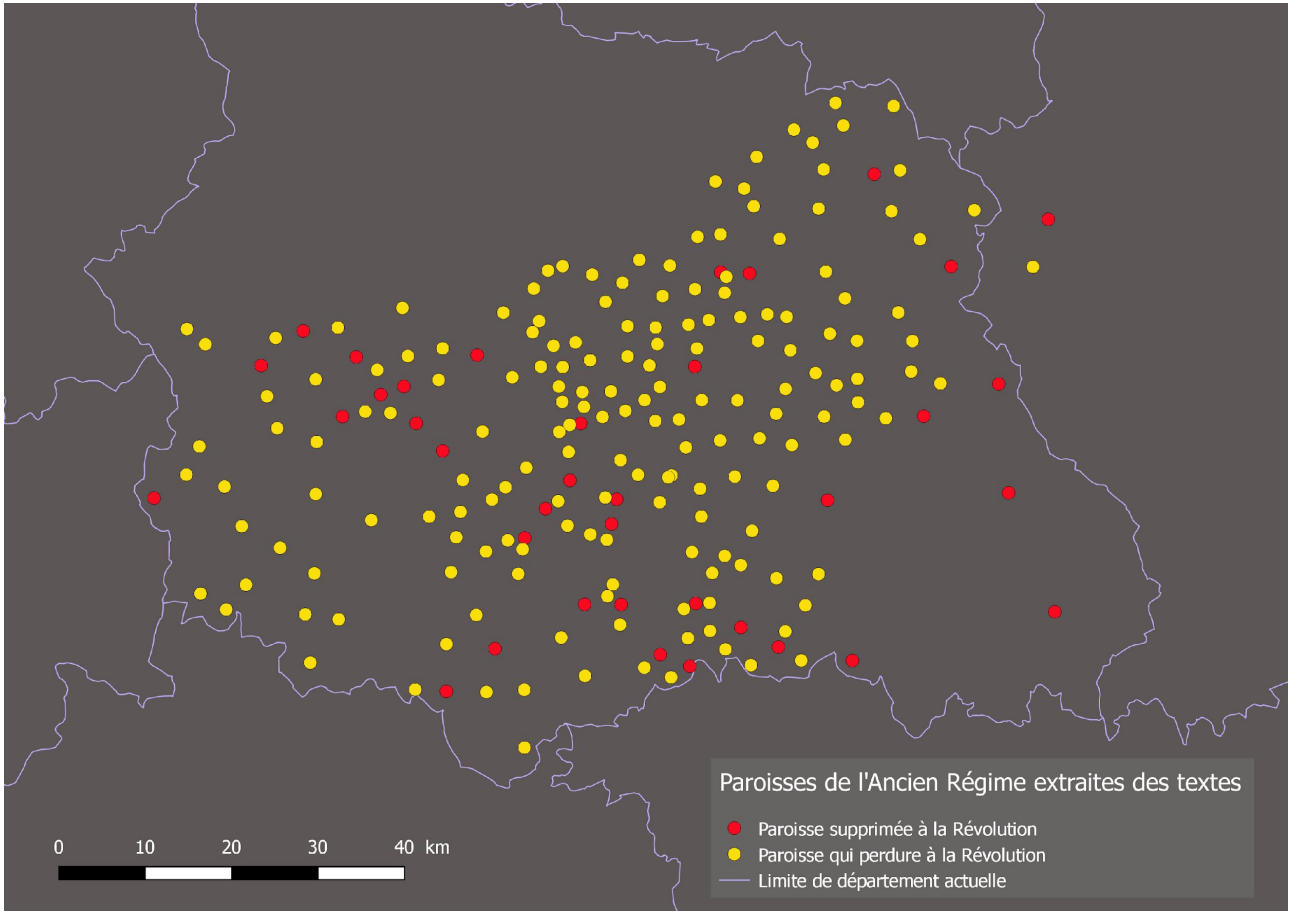
SubClassOf: prg:Paroisse

Inférences pour la restructuration du graphe selon le modèle snapshot

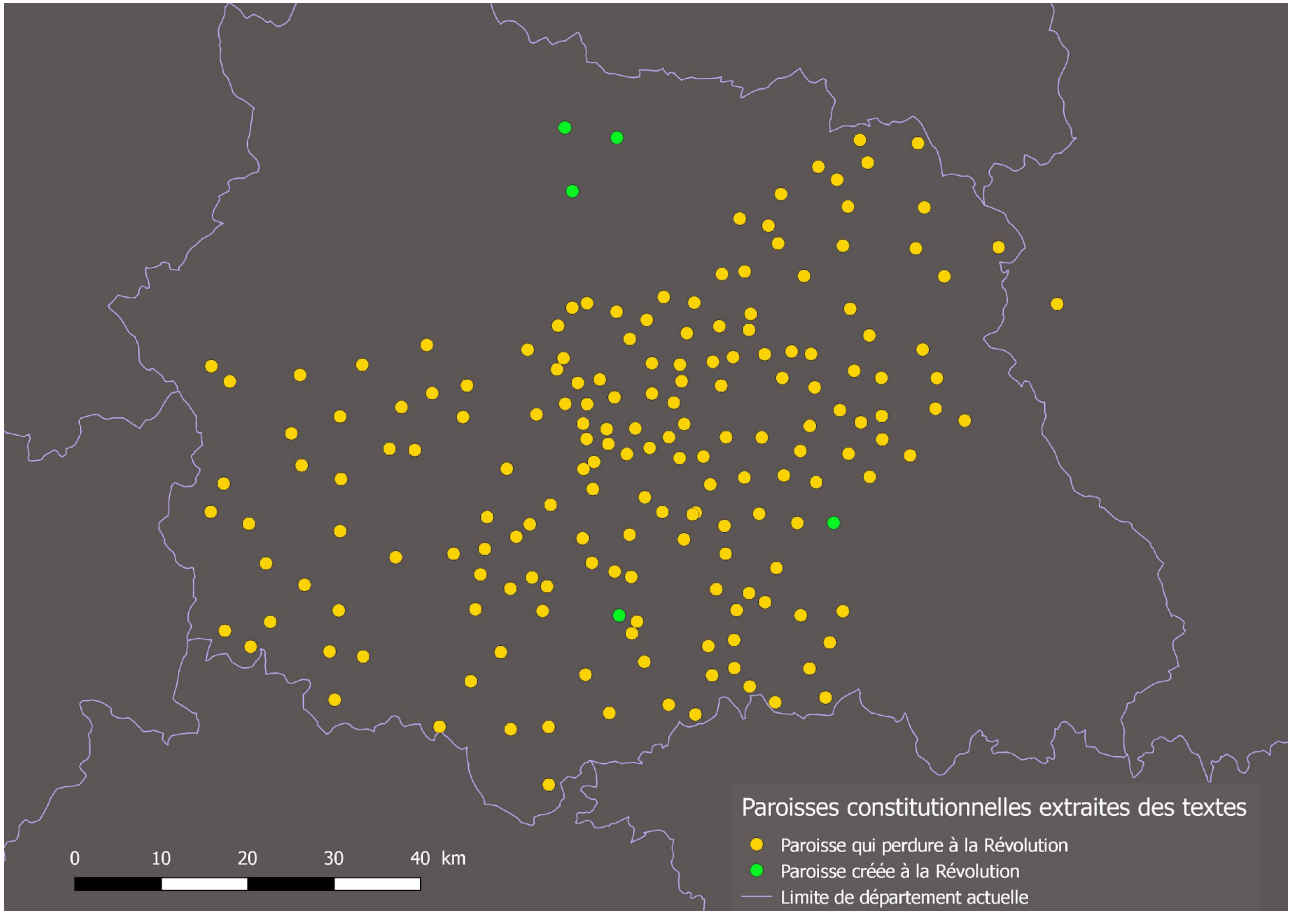


Visualisation GraphDB

Classification des ESN en paroisses de l'ancien régime et paroisses qui perdurent à la Révolution

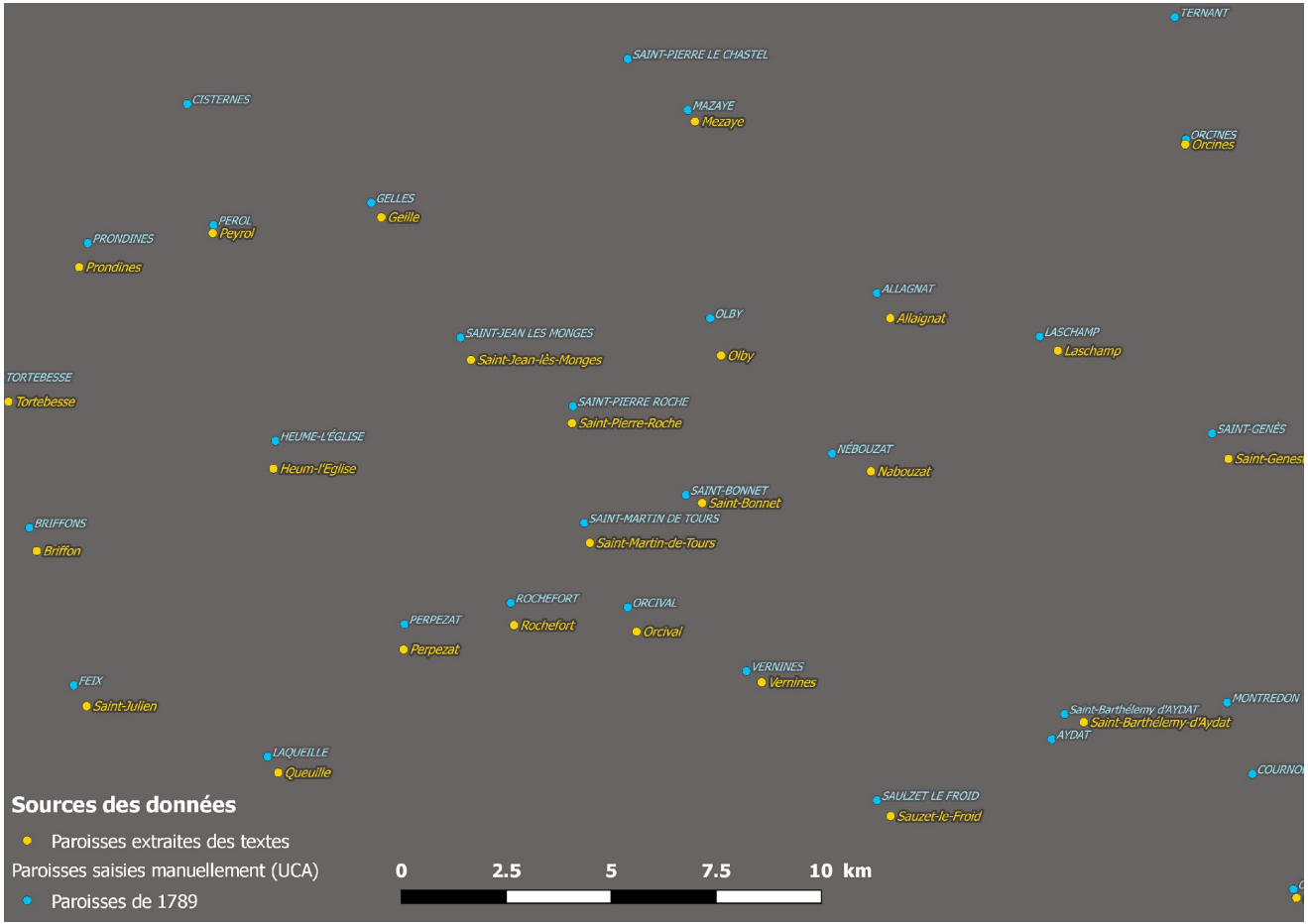


Classification des ESN en paroisses qui perdurent et paroisses créées à la Révolution



Les paroisses de l'Ancien Régime extraites du textes et celles produites à la main

(source: université de Clermont-Ferrand)



Conclusions

Preuve de concept de chaîne de traitement (presque complète) permettant de créer une base de connaissances spatio-temporelles sur l'évolution des paroisses à la Révolution à partir de sources textuelles et cartographiques.

Perspectives d'améliorations pour l'extraction :

- Gérer les paroisses urbaines (pas toujours nommées ou pas toujours présentes sur Cassini)
- Tester la chaîne de traitement sur les arrêtés du directoire des districts pour lesquels on n'a pas la description du redécoupage dans les décrets de l'Assemblée Nationale.
- Définir des métriques adéquates pour l'évaluation de l'extraction des relations d' évolution: extraction complète, extraction partielle (erreur sur le sujet/prédicat/objet), relation partielle dans le texte/référentiel (rdf:nil), etc.

Conclusions

Améliorations pour la désambiguïsation :

- Proposer une mesure de similarité adaptée aux toponymes anciens pour la sélection des candidats,
- Donner la possibilité de ne pas proposer de résultat lors du classement des candidats.

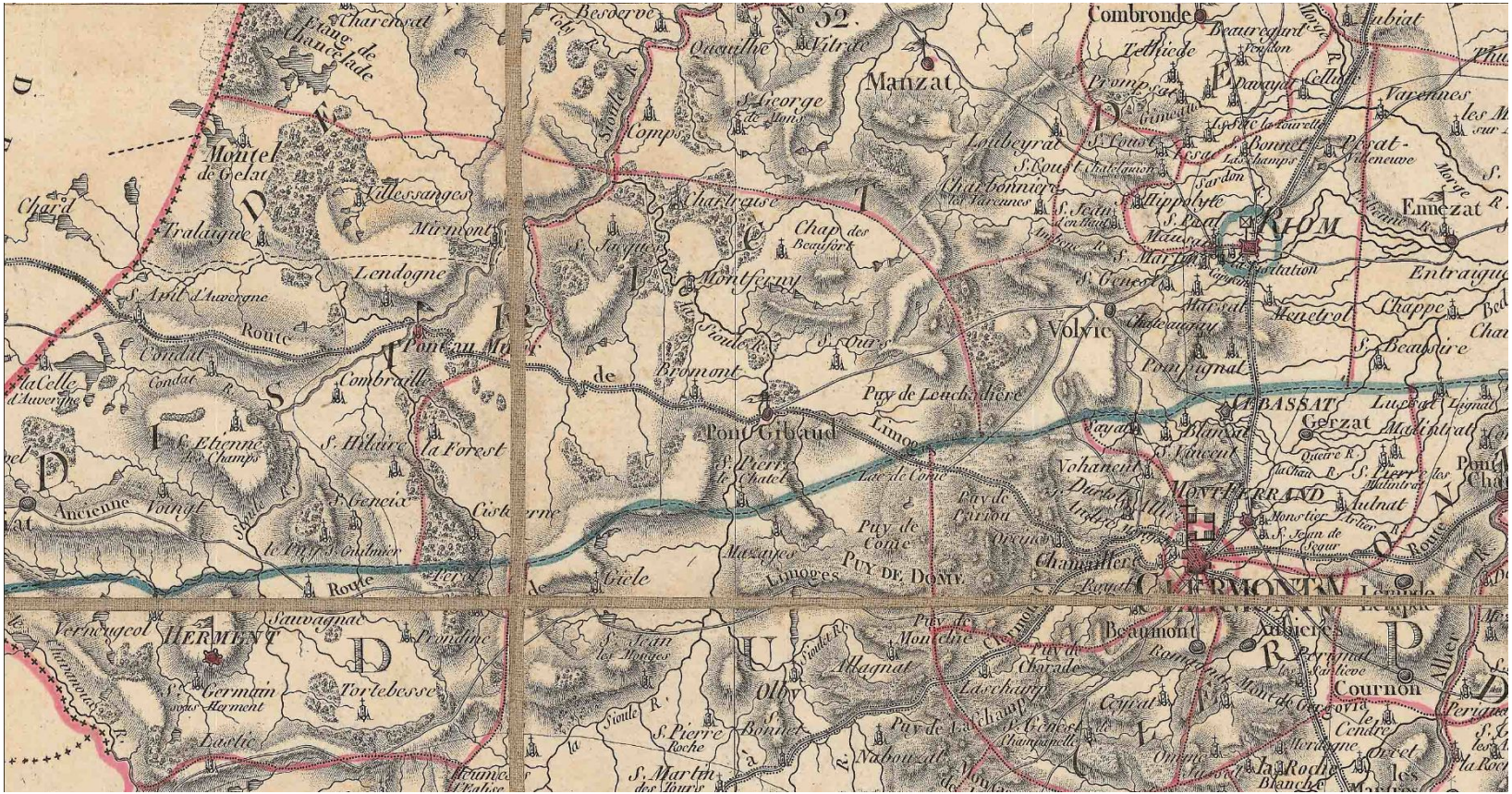
Inférence de nouveaux faits :

- Enrichir l'ontologie pour inférer plus de nouveaux faits, notamment sur les lieux de culte.

Une base de connaissances sur l'évolution du découpage administratif français de l'Ancien Régime à l'Empire?

- Les paroisses constitutionnelles sont une étape de l'évolution entre les paroisses de l'ancien régime et les communes de 1804 : reconstruire l'ensemble du graphe d'évolution ?
- Utiliser des sources cartographiques complémentaires pour reconstituer les limites des paroisses de l'ancien régime (limites de districts, etc.)?
- Croiser les paroisses extraites pour l'ancien régime avec des données démographiques?

Pour aller plus loin…



Source: Collection cartothèque IGN

Merci pour votre attention