



HAL
open science

Création de bases de connaissances topographiques à partir de sources hétérogènes

Helen Mair Rawsthorne

► **To cite this version:**

Helen Mair Rawsthorne. Création de bases de connaissances topographiques à partir de sources hétérogènes. 30e Journées de la Recherche de l'IGN-ENSG en ligne: L'INFORMATION GÉOGRAPHIQUE ENCORE +, May 2021, Saint-Mandé, France. , 2021. hal-03239957

HAL Id: hal-03239957

<https://hal.science/hal-03239957>

Submitted on 27 May 2021

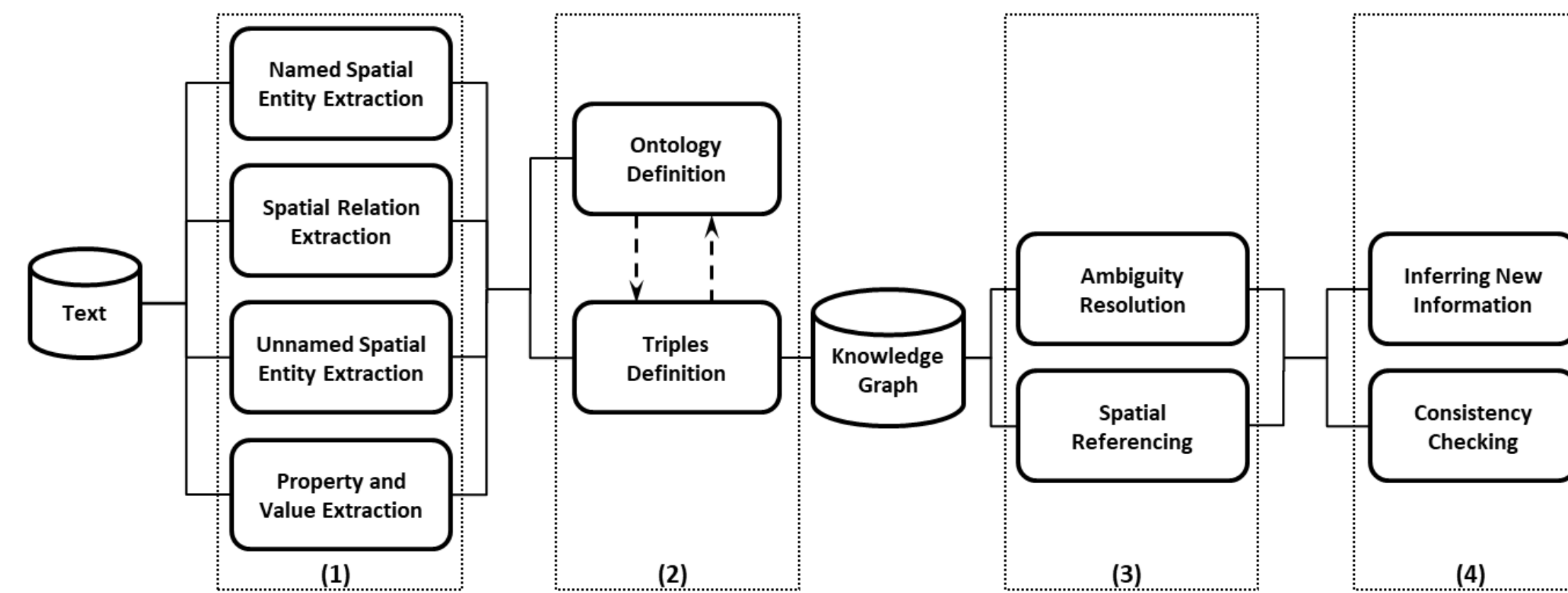
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Objectifs et enjeux de la thèse

- **Objectif** : proposer une approche pour construire, peupler et évaluer une base de connaissances géoréférencées à partir de sources hétérogènes, notamment des bases de données vectorielles et des textes.
- **Enjeu méthodologique** : produire des informations géographiques structurées à partir de textes en langage naturel et les intégrer aux référentiels existants.
- **Enjeux applicatifs** : développer des applications combinant informations géographiques à **références spatiales directes et indirectes** et un **raisonnement spatial quantitatif et qualitatif**.

Approche globale



1. Dans le texte, annoter les entités spatiales, leurs propriétés et leurs relations spatiales.
2. Générer une base de connaissances à partir de ces informations.
3. Géoréférencer les entités spatiales afin de réduire les ambiguïtés, éviter des doublons et enrichir la base de connaissances.
4. Axiomatiser les classes et les propriétés afin de vérifier la cohérence de la base de connaissances et inférer des connaissances qui ne sont pas explicitées dans le texte (par exemple : les relations spatiales inverses, etc.).

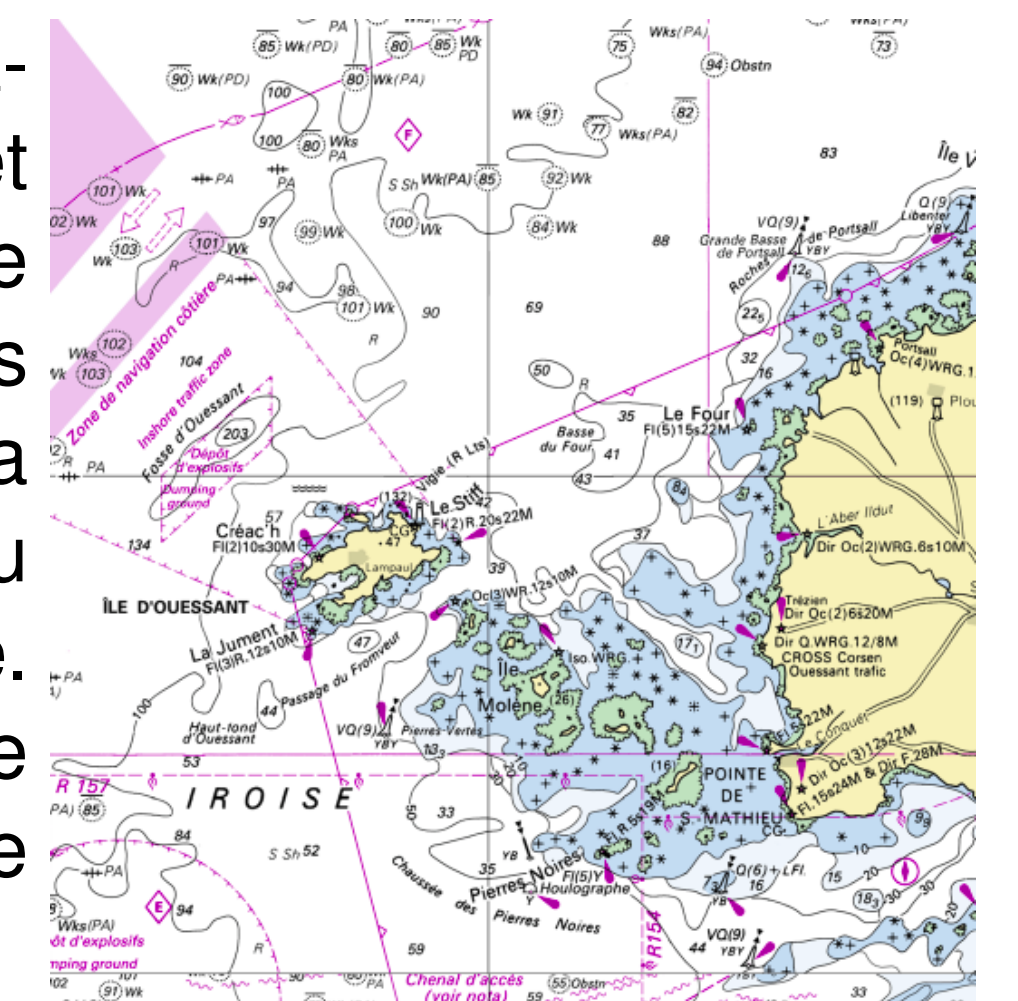
Questions de recherche

- Comment extraire des informations géographiques à partir de textes (entités spatiales nommées ou non, relations spatiales) ?
- Comment désambigüiser et géoréférencer des entités spatiales extraites de textes (entités absentes des référentiels et non nommées) ?
- Comment créer et axiomatiser une ontologie modélisant les entités extraites ? (méthodologie adaptée, (semi-)automatisation)

Une première étude de cas

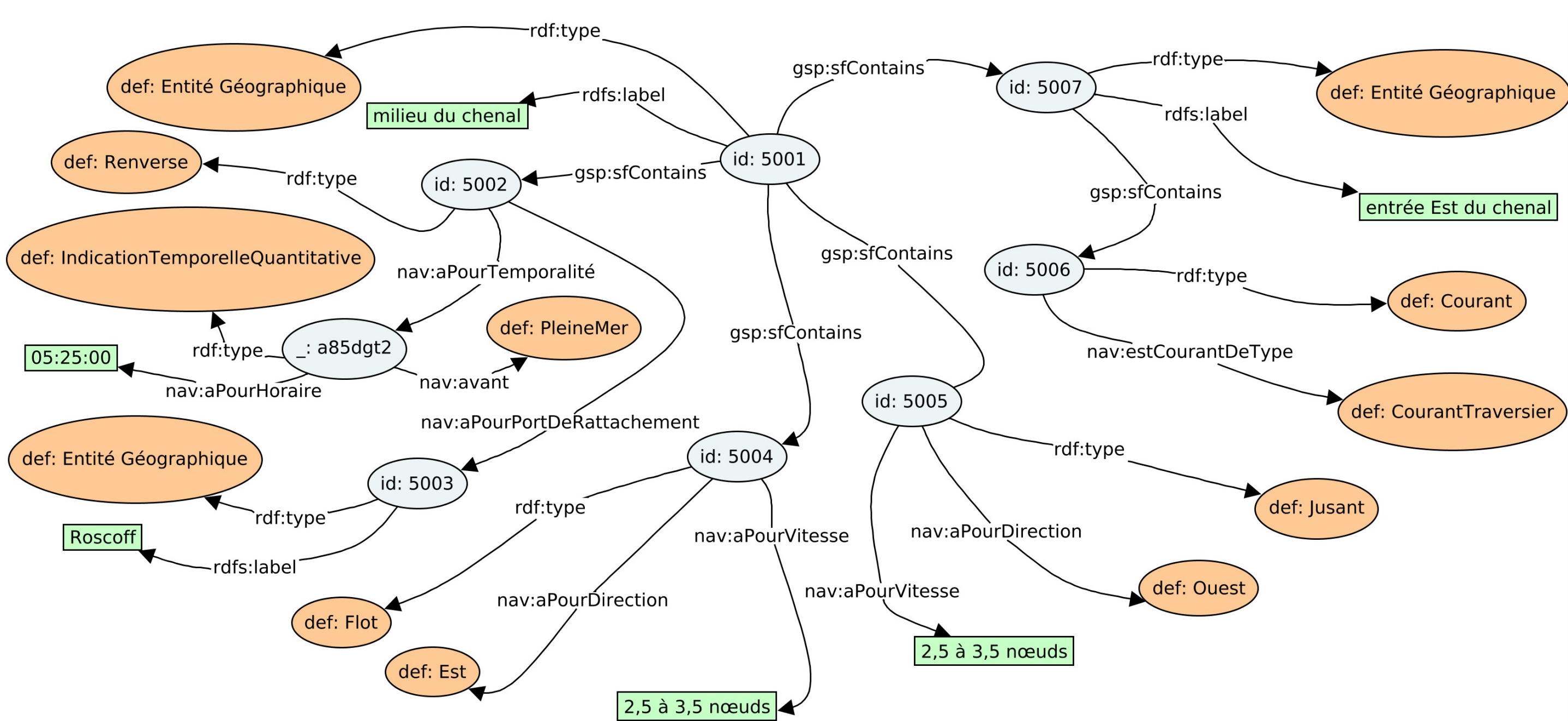


Les **Instructions nautiques** (IN) sont des ouvrages textuels papier et électroniques publiés par le Service hydrographique et océanographique de la Marine (Shom). Elles décrivent le paysage côtier depuis le point de vue d'un navire et donnent des indications pour naviguer près de la côte et accéder aux ports. Les IN aident à la navigation côtière en complément des **cartes nautiques** papier ou électroniques (ENC) pour planifier un itinéraire de navigation côtière. La figure à gauche montre la couverture des IN C22 France (côte Nord et Ouest) et celle à droite montre un extrait de l'Assemblage des cartes RasterMarine (diffusion.shom.fr).



Du texte des IN à une base de connaissances

COURANTS DE MARÉE. — Dans le milieu du chenal, les reverses ont lieu vers - 0525 et + 0020 PM Roscoff. Le flot porte à l'Est et le jusant à l'Ouest ; les vitesses varient selon les endroits, mais sont de 2,5 à 3,5 nœuds en général. Le courant est traversier à l'entrée Est du chenal, à l'Est de la ligne joignant l'Île Pigued à la tourelle « Men Guen Bras ».



Le texte à gauche est un extrait d'un ouvrage des IN (Shom 2010). Les informations géographiques à extraire sont identifiées selon un code couleur : entités géographiques, phénomènes océanographiques, propriétés et relations spatiales. Une fois extraites, ces informations sont représentées sous la forme d'un graphe, conformément aux propositions de Kim et al. 2015.

- Le vocabulaire utilisé pour décrire les informations extraites et permettre leur interrogation est en cours de développement. L'intérêt de cette étape est double : guider l'extraction et faciliter l'exploitation des connaissances extraites.
- Nous avons adopté la méthodologie SAMOD : "Simplified Agile Methodology for Ontology Development" (Peroni 2016).
 - 1ère étape : analyse des besoins. Analyser le contenu des IN, analyser des fiches de préparations de quart, mener une enquête auprès des vrais usagers des IN (des élèves et des instructeurs militaires de l'École Navale à Brest, et des membres de la marine marchande).
 - 2ème étape : définition des questions de compétences auxquelles la base doit permettre de répondre.
 - 3ème étape : proposition de sous-modèles par thème (ex. les courants, les amers, etc.).
 - 4ème étape : évaluation des sous-modèles, corrections éventuelles, itération.

Bibliographie

- Kim, Junchul, Maria Vasardani, and Stephan Winter (2015). "Harvesting large corpora for generating place graphs". In: International Workshop on Cognitive Engineering for Spatial Information Processes (CESIP). Vol. 12. Santa Fe, NM, USA.
- Peroni, Silvio (2016). "A Simplified Agile Methodology for Ontology Development". In: OWL: Experiences and Directions – Reasoner Evaluation. 13th International Workshop, OWLED 2016, and 5th International Workshop, ORE 2016. Bologna, Italy.
- Shom (2010). Instructions nautiques. Vol. C2A France (côtes Nord et Ouest).

Remerciements

Ce travail est co-financé par le Shom et l'IGN et mené au LaSTIG, une unité de recherche de l'Université Gustave Eiffel.

Informations

- Thèse débutée le 1 novembre 2020
- Encadrée par Cécile DUCHÊNE, Nathalie ABADIE, Eric KERGOSIEN et Eric SAUX
- Université Gustave Eiffel, École Doctorale MSTIC
- Équipe STRUDEL, LASTIG, Université Gustave Eiffel, ENSG, IGN, F-94160 Saint-Mandé, France
- Contact : helen.rawsthorne@ign.fr