

Comment créer un graphe de connaissances géospatial à partir de texte ?

Une approche et son application aux *Instructions nautiques*

Helen Mair RAWSTHORNE

Ingénieure de recherche, Mines Saint-Étienne
helenmair.rawsthorne@emse.fr

11 mars 2024

Séminaire IXXI, ENS Lyon

Enjeu

- Informations spatiales précieuses mais sous-exploitées contenues dans des textes
 - Historiques, techniques, littéraires, etc.

Pourquoi l'exploiter ?

- Analyser contenu spatial
- Proposer un accès basé sur la localisation
- Découvrir des nouvelles informations spatiales
 - Lieux, attributs, localisations
 - Créer et enrichir ressources géographiques : bases de données, gazetiers

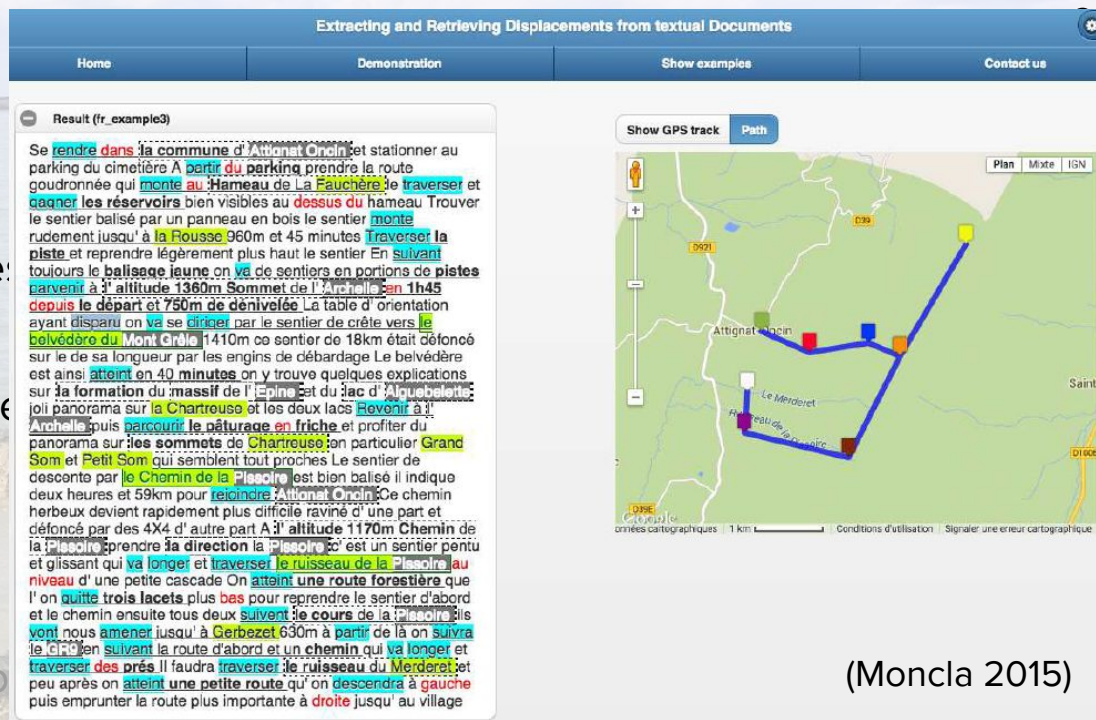
Introduction

Enjeu

- Informations spatiales précieuses contenues dans des textes
 - Historiques, techniques, littéraires

Pourquoi l'exploiter ?

- Analyser contenu spatial ➡
- Proposer un accès basé sur la localisation
- Découvrir des nouvelles informations spatiales
 - Lieux, attributs, localisations
 - Créer et enrichir ressources géographiques : bases de données, gazetteurs



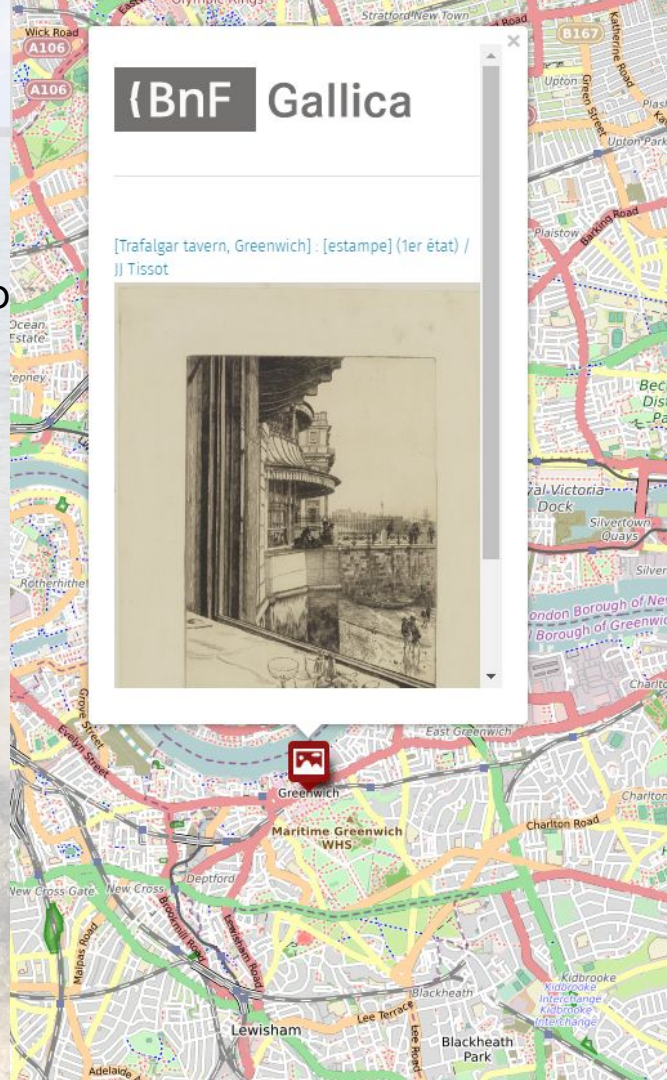
Introduction

Enjeu

- Informations spatiales précieuses mais sous-exploitées contenues dans des textes
 - Historiques, techniques, littéraires, etc.

Pourquoi l'exploiter ?

- Analyser contenu spatial
- Proposer un accès basé sur la localisation ➡
- Découvrir des nouvelles informations
 - Lieux, attributs, localisation
 - Créer et enrichir ressources
 - bases de données, gazetteer



Enjeu

- Informations spatiales précieuses mais sous-exploitées contenues dans des textes
 - Historiques, techniques, littéraires, etc.

Pourquoi l'exploiter ?

- Analyser contenu spatial
- Proposer un accès basé sur la localisation
- Découvrir des nouvelles informations spatiales →
 - Lieux, attributs, localisations
 - Créer et enrichir ressources géographiques : bases de données, gazetiers



Institut national
de l'information
géographique
et forestière



Service
hydrographique et
océanographique
de la Marine

Pourquoi un graphe de connaissances ?

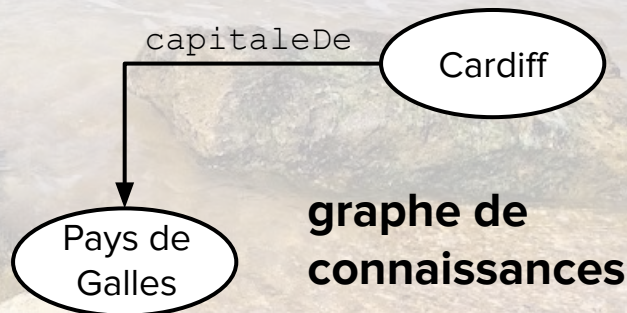
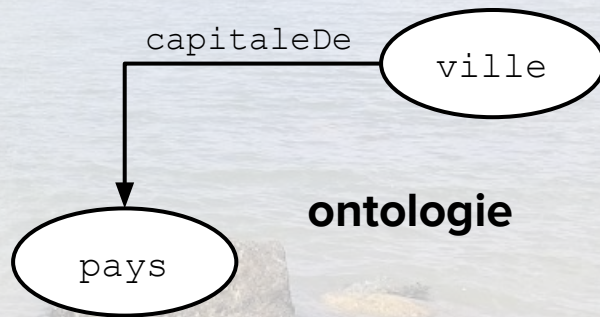
6

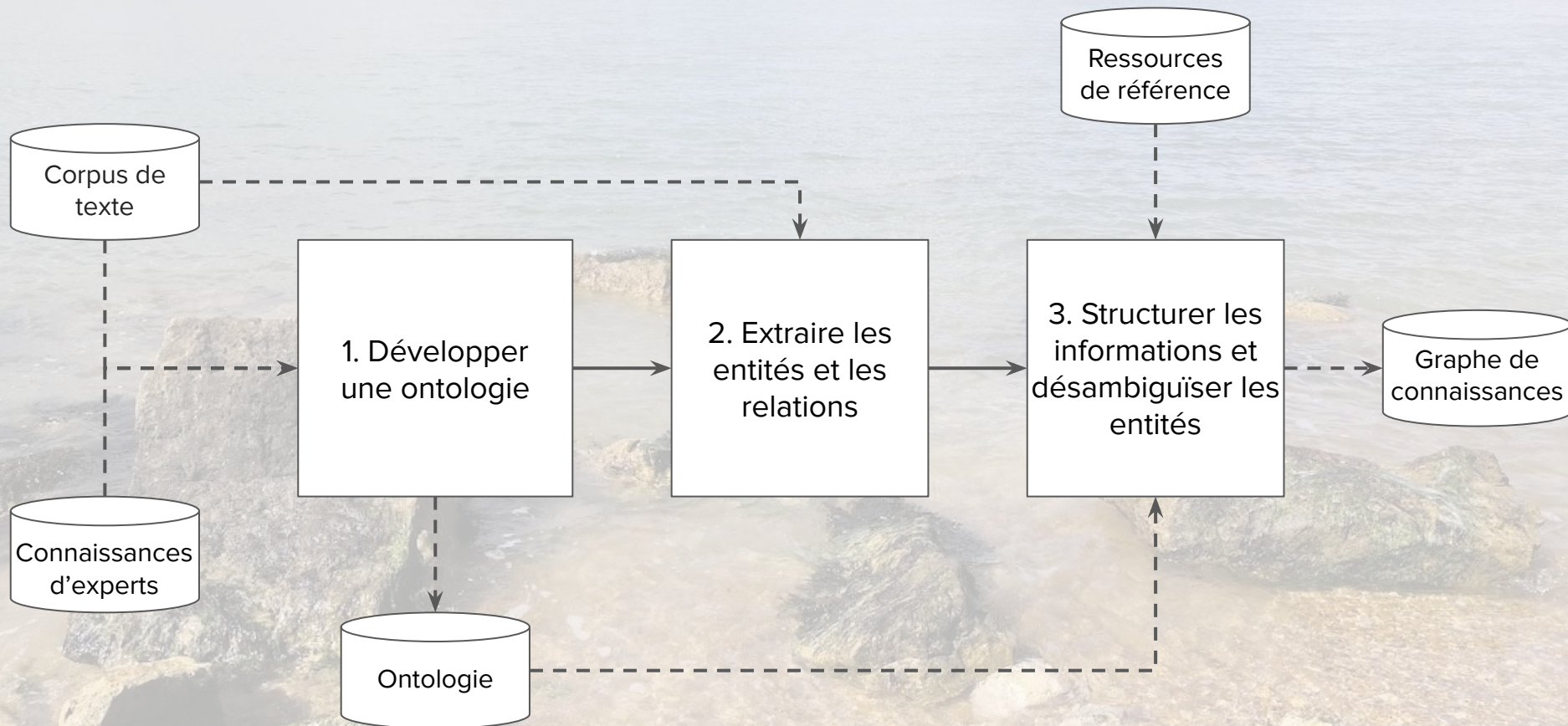
- Hétérogénéité des connaissances géospatiales

- Informations incomplètes
- Points de vue différents
- Détail variable
- Ambiguïtés

- Graphes de connaissances géospatiaux

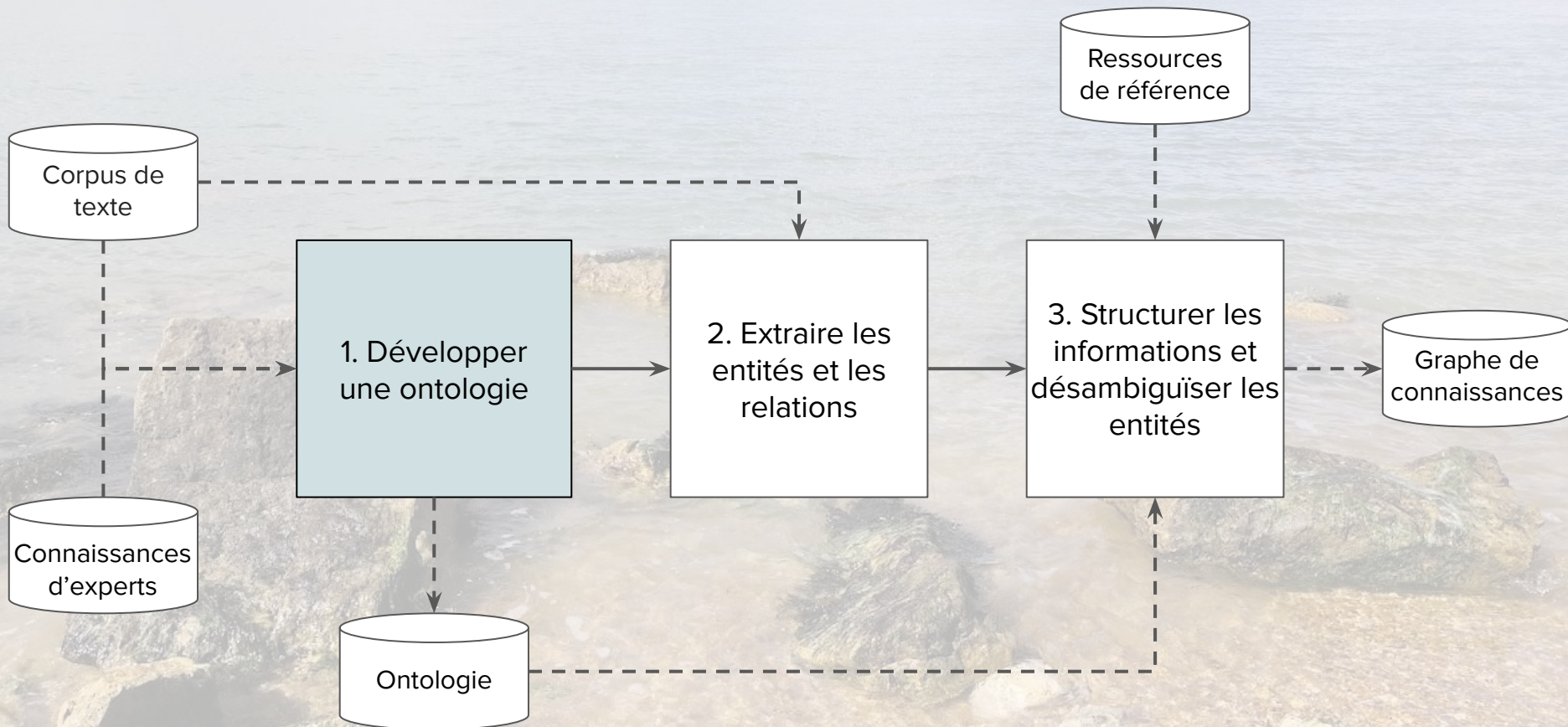
- Requêtage
- Raisonnement spatial
- Liens vers d'autres ressources





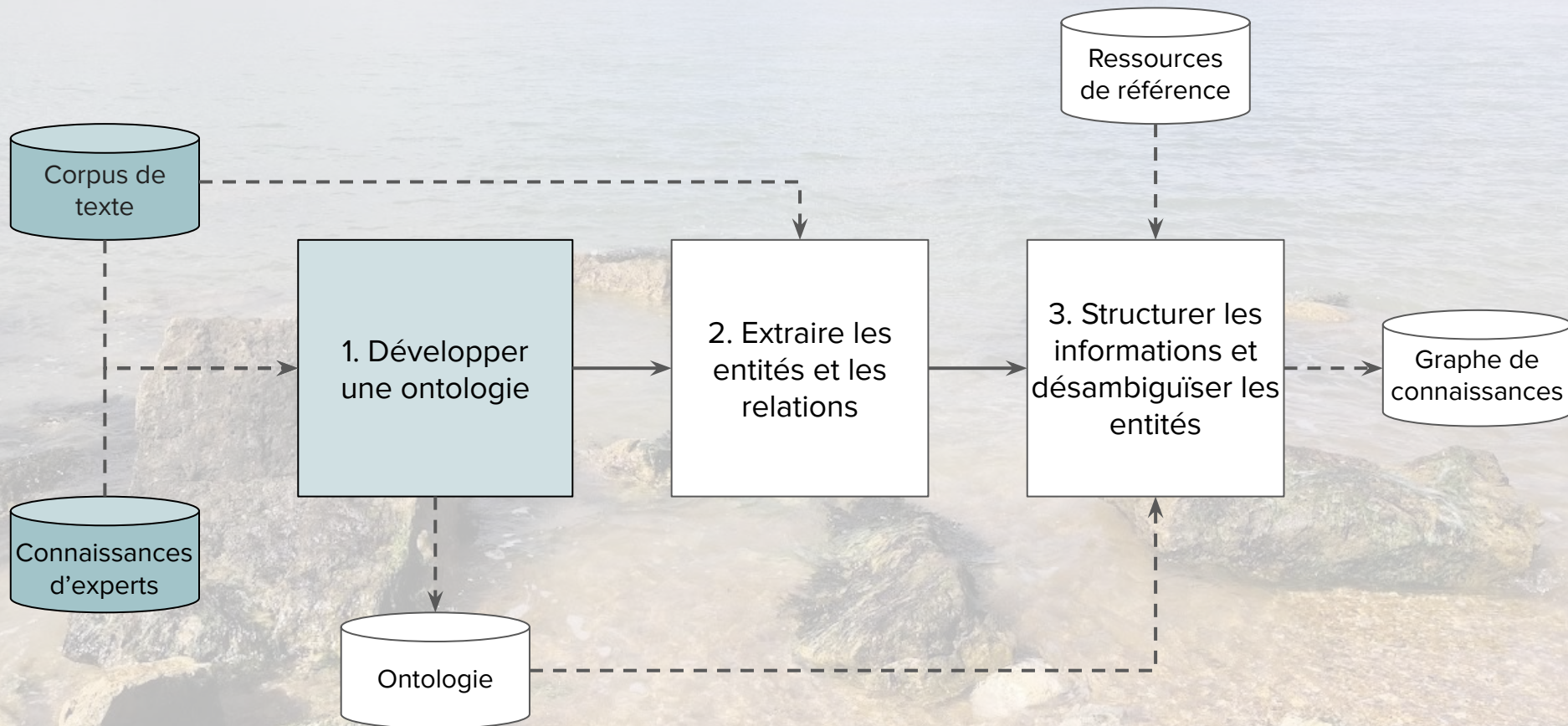
ATONTE étape 1 : développer une ontologie

8



ATONTE étape 1 : développer une ontologie

9



Contexte d'application

Instructions nautiques

- 16 PDFs, 100-800 pages
- Langage naturel, français
- Couverture globale
- Information essentielle pour la navigation côtière



01 5.2.5.2. Amers d'Alger

- 07 À environ 0,5 M au SE de la roche M'Tahen, le petit **îlot de la Marine**, relié à la côte par une courte et large digue, porte en son centre la tour d'un ancien phare.
- 13 La jetée Nord du port d'Alger, dite jetée Kheir Eddine, s'enracine à la pointe Sud de l'îlot. Son extrémité porte le phare d'Alger ($36^{\circ} 46,6' \text{ N} - 3^{\circ} 04,7' \text{ E}$), tour blanche haute de 17 m. D'autres feux marquent les différentes passes dans le port.
- 19 L'immeuble de l'hôtel El Aurassi se remarque en $36^{\circ} 46,6' \text{ N} - 3^{\circ} 03,1' \text{ E}$ avec à proximité les deux minarets de la mosquée Ennour ($36^{\circ} 46,4' \text{ N} - 3^{\circ} 02,9' \text{ E}$) ainsi que les immeubles de la cité des Tagarins ($36^{\circ} 46,3' \text{ N} - 3^{\circ} 02,8' \text{ E}$). La pointe du toit du grand théâtre ($36^{\circ} 46,9' \text{ N} - 3^{\circ} 03,6' \text{ E}$) peut être utile pour se présenter au port.
- 25 À 1,9 M au SSW du phare d'Alger, un monument ($36^{\circ} 44,7' \text{ N} - 3^{\circ} 04,2' \text{ E}$) haut de 80 m est remarquable. Le monument et les pylônes environnants portent des feux d'obstacle aérien. À 1 M au SE du monument, le petit dôme blanc de l'ancien séminaire de **Kouba** ($36^{\circ} 44,1' \text{ N} - 3^{\circ} 05,3' \text{ E}$) est visible de toute la baie.
- 31 Le tronçon Nord de la jetée détachée est remarquable par les grands réservoirs qu'il porte.

37



5.2.5.2.A. — Monument remarquable d'Alger au 200° et 2 M (2009) (photo Omar Zerouk <http://www.flickr.com/photos/lomaxe/>).



01 **5.2.5.2. Amers d'Alger**

07 À environ 0,5 M au SE de la roche M'Tahen, le petit **îlot de la Marine**, relié à la côte par une courte et large digue, porte en son centre la tour d'un ancien phare.

13 La jetée Nord du port d'Alger, dite jetée Kheir Eddine, s'enracine à la pointe Sud de l'îlot. Son extrémité porte le phare d'Alger ($36^{\circ} 46,6' \text{ N} - 3^{\circ} 04,7' \text{ E}$), tour blanche haute de 17 m. D'autres feux marquent les différentes

13 **INSTRUCTIONS.** — Par mauvais temps, il est déconseillé d'entrer dans le port ou d'en sortir.

19 Si le feu de guidage n'est pas utilisable, se présenter en route à 070° vers le djebel ach Zerfet (§ 4.6.4.2.) en gardant ce sommet à droite de l'extrémité Est du brise-lames détaché. Puis venir rondement vers la droite entre les musoirs des jetées, distants de 150 m. Cette manœuvre est délicate par fort vent de NW et le pilote peut demander de mouiller l'ancre de tribord pour faciliter et sécuriser la giration.

25 Des épaves se trouvent devant l'entrée du port.

01 **4.6.4.6. Accès**

07 Un feu marque chaque extrémité du brise-lames détaché et le musoir de chacune des deux jetées brise-lames formant le bassin du port. La ville de Ténès, à 1 M au SW du port, est entourée de remparts.



Monument en
 $36^{\circ} 44,7' \text{ N} - 3^{\circ} 04,2' \text{ E}$
à 200° et 2 M.

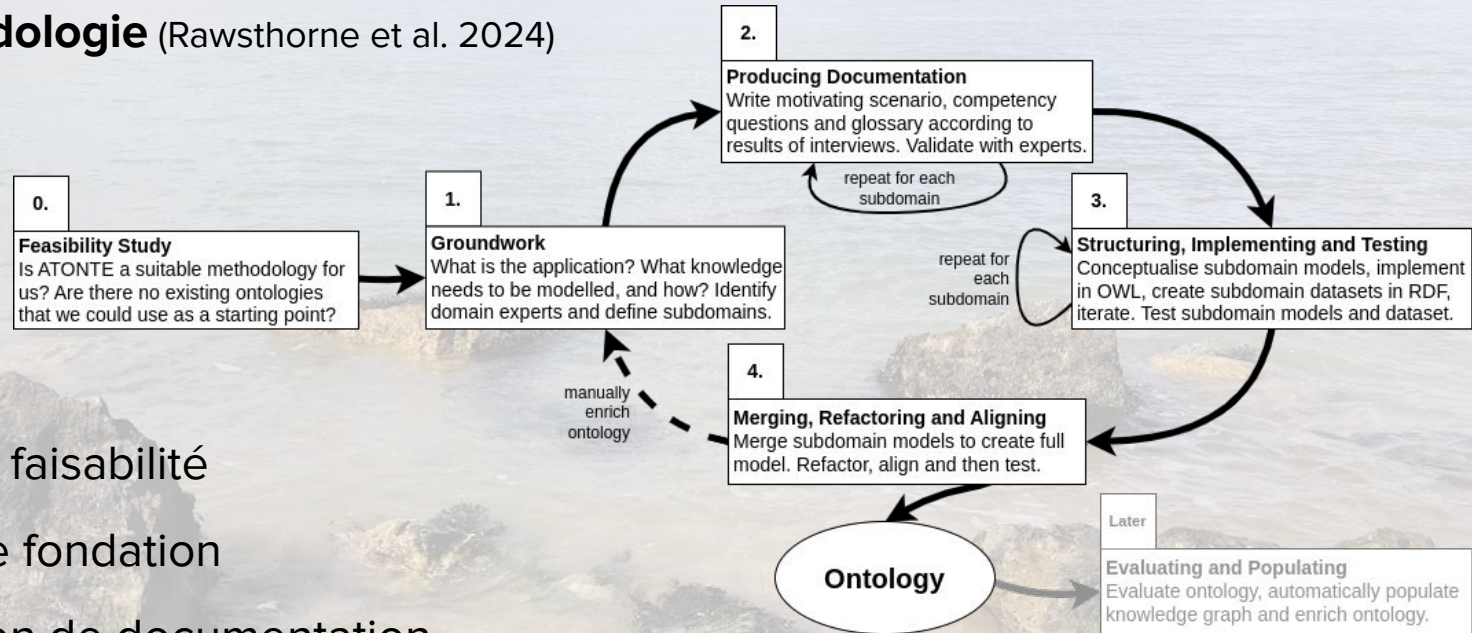
<http://www.flickr.com/photos/lomaxe/>

5.2.5.2.A. — Monument remarquable d'Alger au 200° et 2 M (2009) (photo Omar Zerouk <http://www.flickr.com/photos/lomaxe/>).

ATONTE étape 1 : développer une ontologie

13

Notre méthodologie (Rawsthorne et al. 2024)



0. Étude de faisabilité

1. Travail de fondation

2. Production de documentation

3. Structurer, implémenter et tester les modèles de sous-domaine

4. Fusionner, remanier et aligner

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ?
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ?
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ? ➡ oui
- Experts du domaine disponibles ?
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ?
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ? ➡ oui
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ?
- Ontologies existantes ?

0. Étude de faisabilité

- Modélisation d'un corpus + les connaissances d'experts ?
- Experts du domaine disponibles ?
- Ontologies existantes ? ➡ oui mais pas disponibles
 - Navigation maritime : non-publiée
 - Sécurité maritime
 - Vie marine et environnement

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus ➡ analyser le contenu, la production
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine ➡ Shom, OHI
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine ➡ rédacteurs, utilisateurs
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

Moderniser la production, le maintenance et l'utilisation des *Instructions nautiques*

- Pas entièrement fixé
- Structurer le contenu
- Lier le texte et les cartes nautiques
- Proposer un accès alternatif aux utilisateurs: plateforme Web cartographique ?
- Système de recommandation d'itinéraire (Laddada 2018)
- Production et mise à jour automatiques : langage contrôlé (Sauvage-Vincent 2017)

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques ➡
- Diviser le domaine en sous-domaines

Extrait

"À 30 M à l'Est de Makemo, Nihiru, atoll de 5 M de diamètre, est bien boisé, sauf au Sud."

Triplets sémantiques triples

entité 1 - a pour nom - Makemo
entité 2 - a pour nom - Nihiru
entité 2 - est un - atoll
entité 2 - est à 30 M de - entité 1
entité 2 - est à l'est de - entité 1
entité 2 - a pour diamètre - 5 M

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines

1. Travail de fondation

- Familiarisation avec le corpus
- Analyser d'autres sources de connaissances du domaine
- Identifier les experts du domaine
- Définir l'application de l'ontologie
- Créer un premier jeu de triplets sémantiques
- Diviser le domaine en sous-domaines ➡

Consignes de navigation	Entités et relations spatiales maritimes
Temporalité et phénomènes météorologiques et océanographiques	Navires

2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux
- Questions informelles de compétence
 - Réponses
- Glossaire

2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux
- Questions informelles de compétence
 - Réponses
- Glossaire

2. Production de documentation

- Consulter des experts du domaine ➡ au même temps que la documentation
 - Entretiens semi-structurés, 30'-60'
 - 15 utilisateurs des *Instructions nautiques*
 - Étudiants, instructeurs, militaires, civiles
 - Façon d'utiliser les *Instructions nautiques*
 - Leurs besoins et envies
 - Repères, informations sur les ports, coordonnées
 - Accès au texte intégral



2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux
- Questions informelles de compétence
 - Réponses
- Glossaire

2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux ➡
- Questions informelles de compétence
 - Réponses
- Glossaire

MOTIVATING SCENARIO p. 3

Main Concepts and Characteristics

Spatial Entity

- name [optional if type known]
- type [optional if name known]
- geographic coordinates [optional]
- spatial relations involved in [optional]
- visual characteristics [optional]
 - colour [optional]
 - pattern [optional]
 - height [optional]

2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux
- Questions informelles de compétence
 - Réponses
- Glossaire

2. Production de documentation

- Consulter des experts du domaine
- Argumentaire
 - Nom et description
 - Extraits et triplets sémantiques
 - Concepts et caractéristiques principaux
- Questions informelles de compétence
 - Réponses
- Glossaire

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine ➡
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques

Extrait

“À 30 M à l’Est de Makemo. Nihiru, atoll de 5 M de diamètre, est bien boisé, sauf au Sud.”

Triplets sémantiques

entité 1 - a pour nom - Makemo
entité 2 - a pour nom - Nihiru
entité 2 - est un - atoll
entité 2 - est à 30 M de - entité 1
entité 2 - est à l’est de - entité 1
entité 2 - a pour diamètre - 5 M



Conceptualisation des prédicats

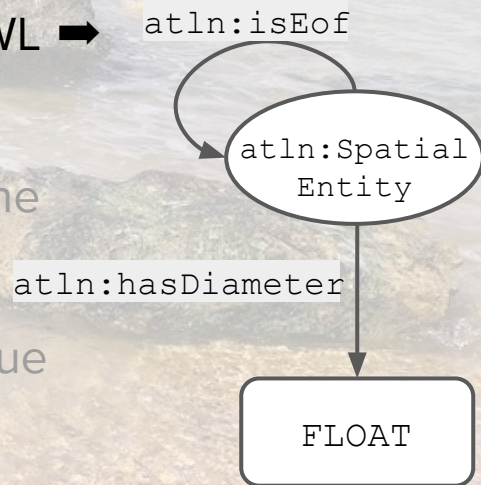
[INDIVIDUAL] est un [entité spatiale]
[entité spatiale] a pour nom [STRING]
[entité spatiale] est à l’est de [entité spatiale]
[entité spatiale] est à 30 M de [entité spatiale]
[entité spatiale] a pour diamètre [NUMBER + UNIT]
[entité spatiale] a pour région [entité spatiale]
[entité spatiale] a pour aspect visuel [boisé | non-boisé]
[entité spatiale] est une partie de [entité spatiale]

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL ➔
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données



3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

3. Structurer, implémenter et tester les modèles de sous-domaine

- Conceptualiser les modèles de sous-domaine
 - Grouper les sujets/objets et prédicats utilisés dans les triplets sémantiques
- Implémenter les modèles de sous-domaine en utilisant OWL
 - À partir de la conceptualisation
- Créer des jeux de données RDF pour chaque sous-domaine
 - À partir des triplets sémantiques
- Faire évoluer les modèles et les jeux de données de chaque sous-domaine de façon itérative
- Tester les modèles et les jeux de données

4. Fusionner, remanier et aligner

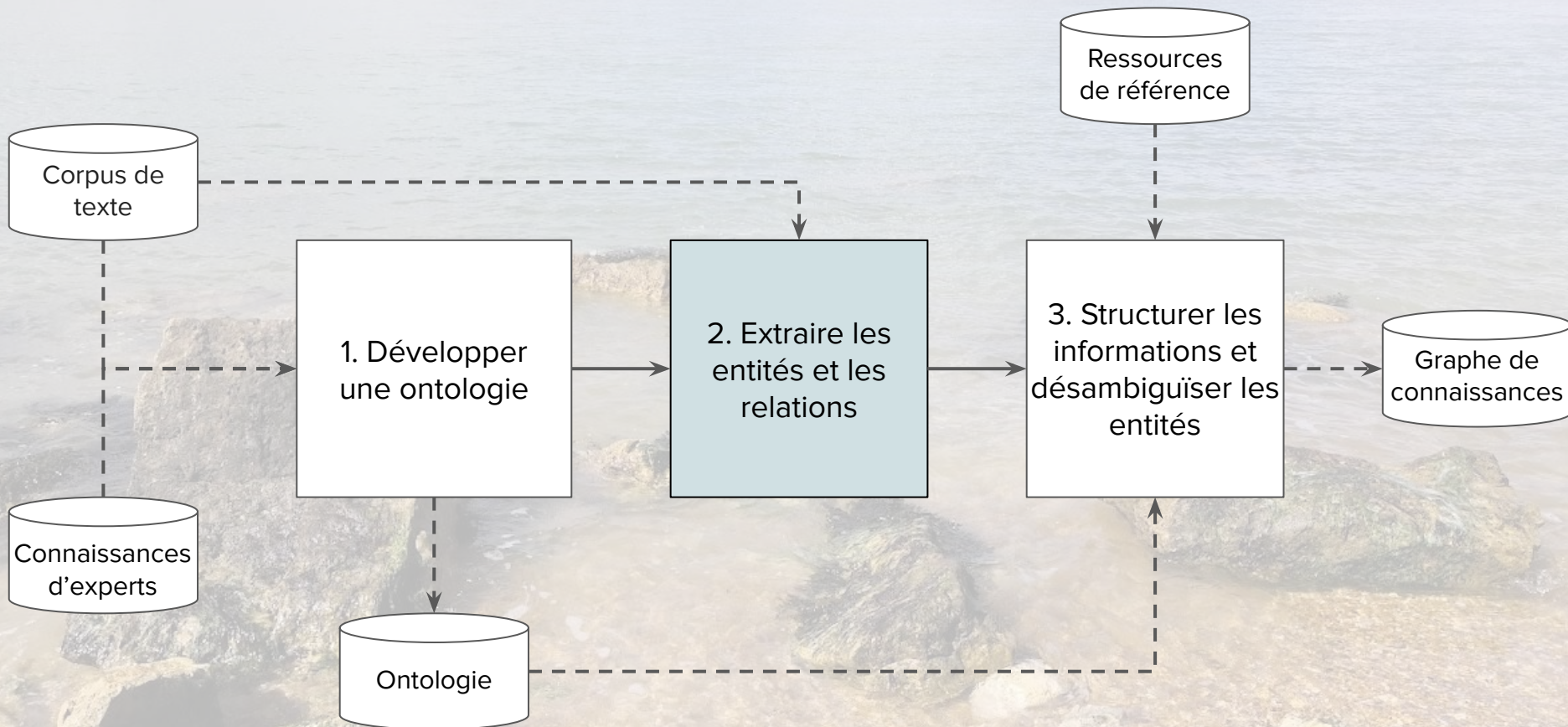
- Fusionner les modèles de sous-domaine afin de créer le modèle complet
- Fusionner les jeux de données de chaque sous-domaine afin de créer le jeu de données exemple pour le modèle complet
- Remanier et aligner le modèle complet
- Tester le modèle complet et le jeu de données exemple

Ontologie ATLANTIS (Rawsthorne et al. 2022)

- 110 classes, 90 propriétés d'objets, 90 propriétés de données
- Évaluée à travers de questions de compétences
- Évaluée à travers de sa réutilisation : *Sailing Directions* (Benoit and Kergus 2022)
- Licence Ouverte Etalab Version 2.0 : github.com/umrlastig/atlantis-ontology

ATONTE étape 2 : extraire les entités et relations spatiales

51



Extraction d'entités plates vs imbriquées

Entités plates

entité spatiale —

Robben Island Lighthouse

Entités imbriquées (Moncla 2015)

geographic feature —

name —

geographic name —

Robben Island Lighthouse

*entité classée
automatiquement*

*indication de sa
position géographique*

Nos besoins

- Extraction d'entités spatiales imbriquées : utile lors de la population de graphes de connaissances
- Extraction de relations spatiales : utile lors de la désambiguïsation des entités
- Nombre inconnu d'entités et de relations par phrase

ATONTE étape 2 : extraire les entités et relations spatiales

54

Princeton **U**niversity **R**elation **E**xtraction system, **PURE** (Zhong and Chen 2021)

- Extraction d'entités plates et génériques et de relations génériques binaires à partir de texte
 - Modèles BERT pré-entraînés
 - Un pour les entités, un pour les relations
 - Pas de partage entre les modèles
 - Entraînés sur un jeu de texte annoté, format JSONL

Notre approche : adaptation de **PURE** (Rawsthorne et al. 2023)

- PURE : extraction d'entités plates et génériques et de relations génériques binaires à partir de texte
- Nous : extraction d'entités spatiales et imbriquées et de relations spatiales binaires à partir de texte
 - Modèles BERT pré-entraînés
 - Un pour les entités, un pour les relations
 - Pas de partage entre les modèles
 - **Entraînés sur un jeu de texte annoté, format JSONL**

Notre approche : schéma d'annotations imbriquées (Rawsthorne et al. 2023)

- Entités : geographic feature, name, geographic name
- Relations : une étiquette pour une paire de mots
 - 16 directions cardinales (N, NE, NNE, etc.)
 - “est au large de”, “est marqué par”, “est un élément de”

Entités spatiales imbriquées

geographic feature —

name —

geographic name —

Relations spatiales

relation dirigée →

Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

est ENE de



Application aux *Instructions nautiques*

- Extraits des PDFs : couverture mondiale
- Outil d'annotation *brat*
- Jeu de données de 101 400 mots, annoté manuellement
 - ~17,000 étiquettes entités
 - ~3,000 étiquettes relations
- Conversion au format JSONL
- Entraînement sur GENCI



Application aux *Instructions nautiques*

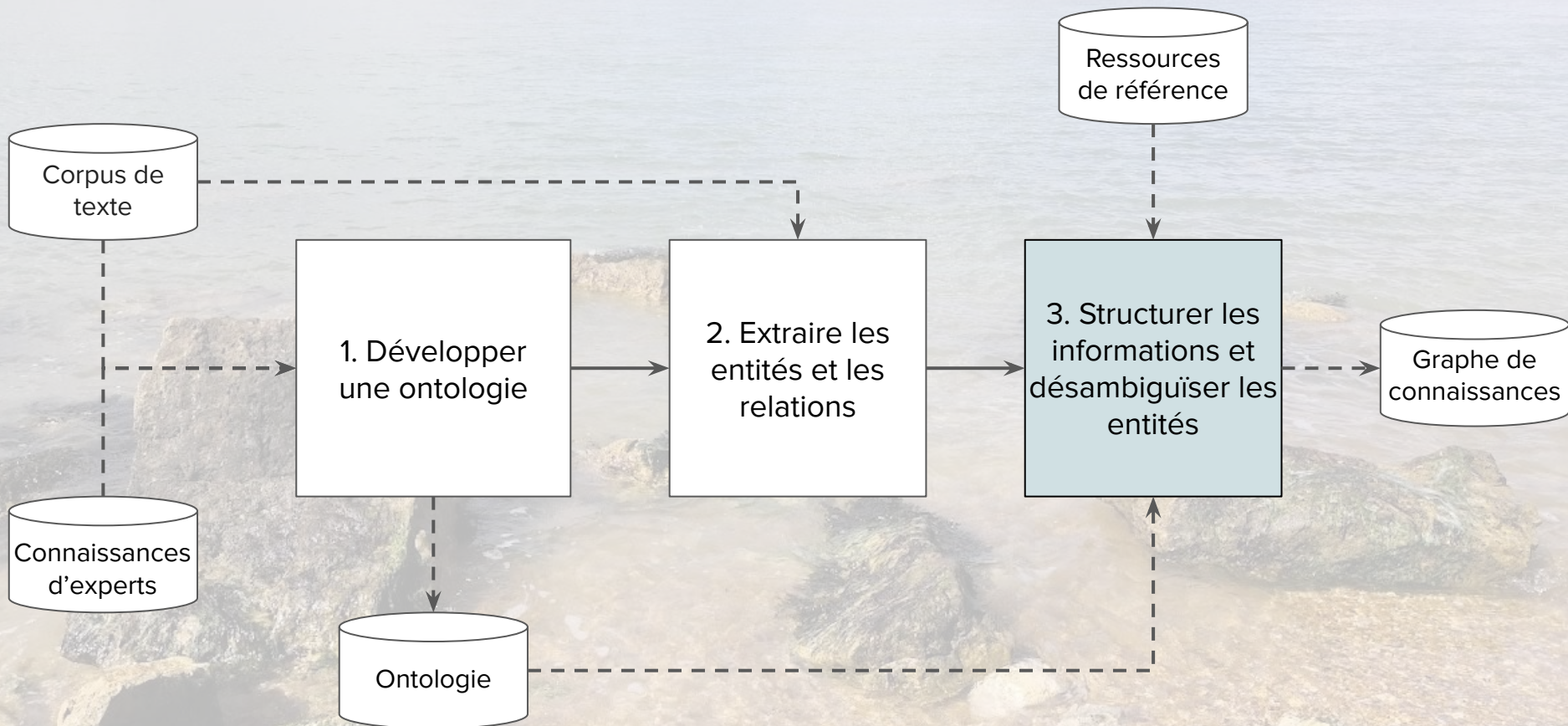
- Modèles BERT : monolingue, multilingue
- Fenêtre de contexte : 0, 50, 100, 150, 200, 248

	Fenêtre de contexte	Modèle BERT	Score F1
Entités	200	multilingue	92.3 ± 0.2
Relations	0	monolingue	64.2 ± 2.2
Relations de bout en bout (entités prédites)	0	monolingue	63.9 ± 2.2

- Jeu de données ATLANTIS de 101,400 mots (Rawsthorne et al. 2023)
 - ~17,000 étiquettes d'entités spatiales imbriquées
 - ~3,000 étiquettes de relations spatiales binaires
 - Formats JSONL et *brat*
 - Licence Ouverte Etalab Version 2.0 : github.com/umrlastig/atlantis-dataset
- Jeu de données TextMine'24 de 66,030 mots (Rawsthorne et al. 2024)
 - ~11,000 étiquettes d'entités spatiales à 1 et 2 niveaux
 - Compétition Kaggle : kaggle.com/competitions/defi-textmine-2024

ATONTE étape 3 : structurer et désambiguïser

60

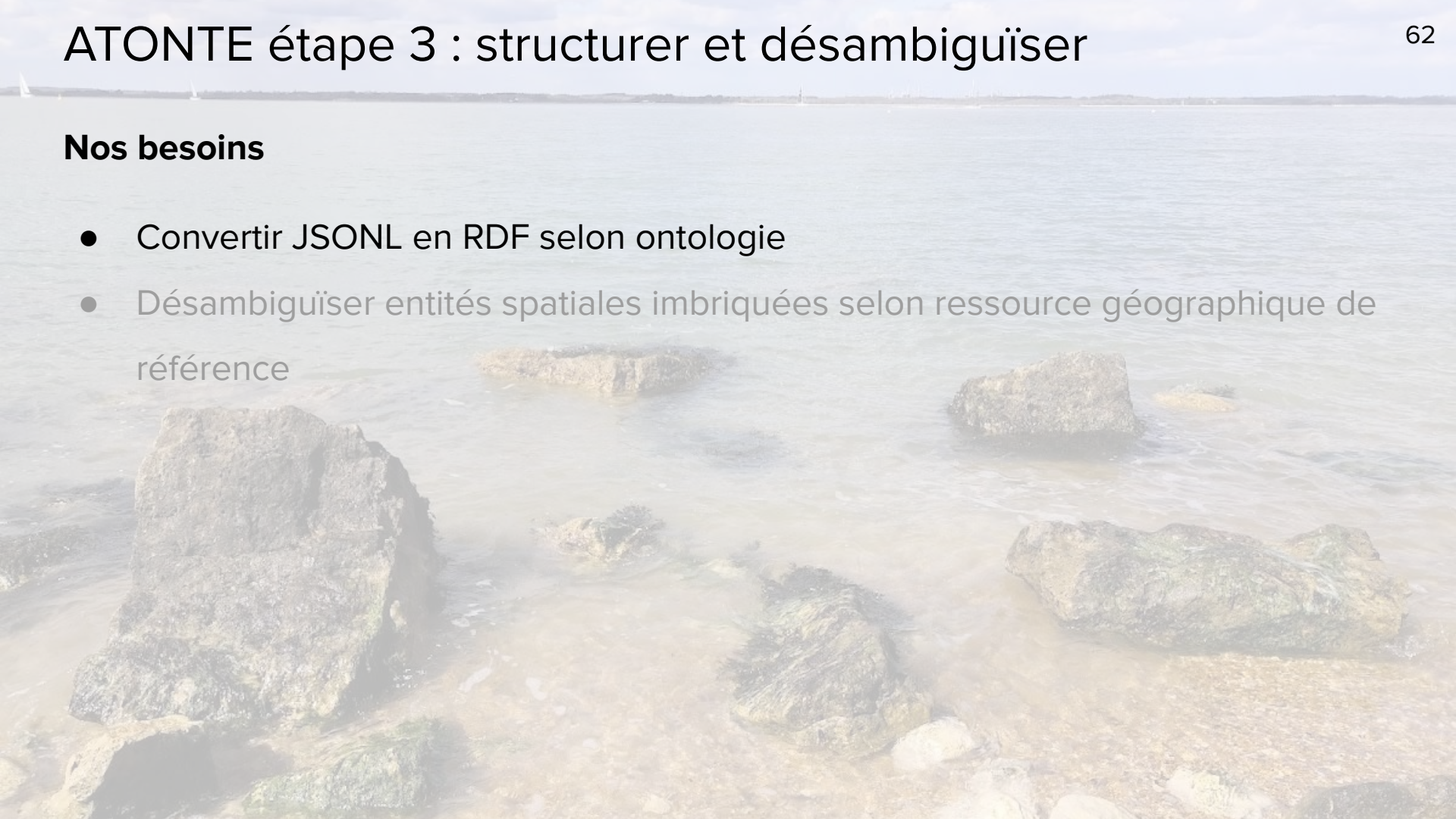


Nos besoins

- Convertir JSONL en RDF selon ontologie
- Désambiguïser entités spatiales imbriquées selon ressource géographique de référence

Nos besoins

- Convertir JSONL en RDF selon ontologie
- Désambiguïser entités spatiales imbriquées selon ressource géographique de référence



Structurer les informations extraites en RDF

- SPARQL-Generate (Lefrançois et al. 2017)
- Extraction d'entités spatiales imbriquées : classées automatiquement
- Extraction de relations spatiales : type dans l'étiquette

Structurer les informations extraites en RDF

- SPARQL-Generate (Lefrançois et al. 2017)
- Extraction d'entités spatiales imbriquées : classées automatiquement
- Extraction de relations spatiales : type dans l'étiquette

Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

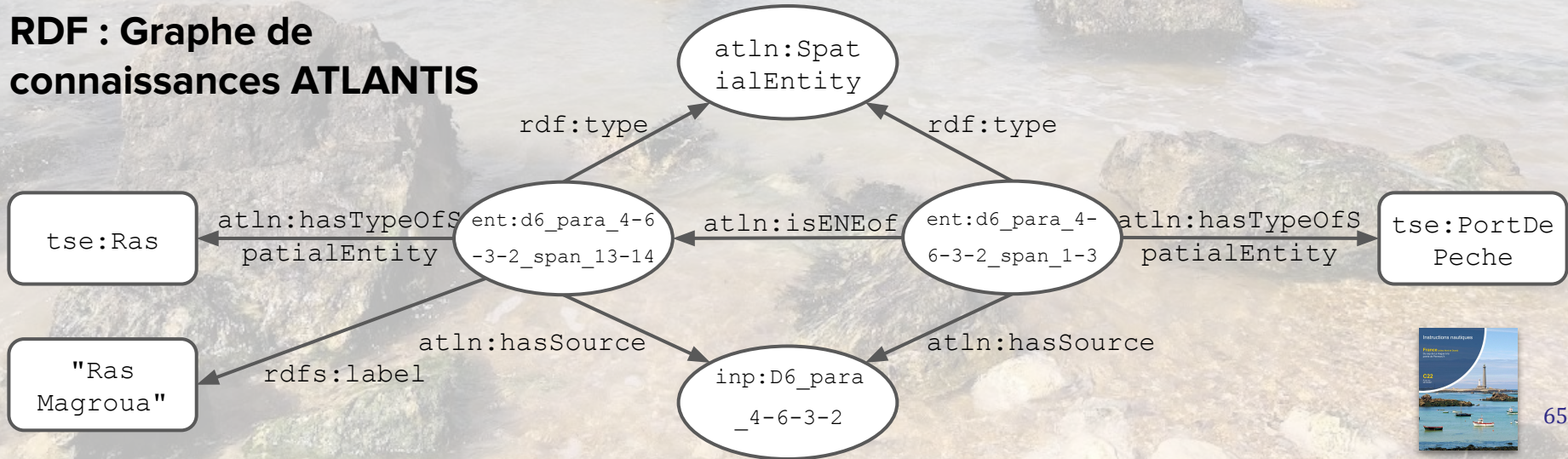
65

JSONL

```
{ "doc_key": "d6_para_4-6-3-2",  
  "dataset": "atlantis",  
  "sentences": [ ["Un", "port", "de", "pêche", "est", "établi", "à", "5,7", "M", "à", "l'",  
"ENE", "de", "Ras", "Magroua", "." ] ],  
  "ner": [ [ [1, 3, "geogFeat"], [13, 13, "geogFeat"], [14, 14, "name"], [13, 14, "geogName"] ] ],  
  "relations": [ [ [1, 3, 13, 14, "isENEof"] ] ] }
```

is ENE of

RDF : Graphe de connaissances ATLANTIS



Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

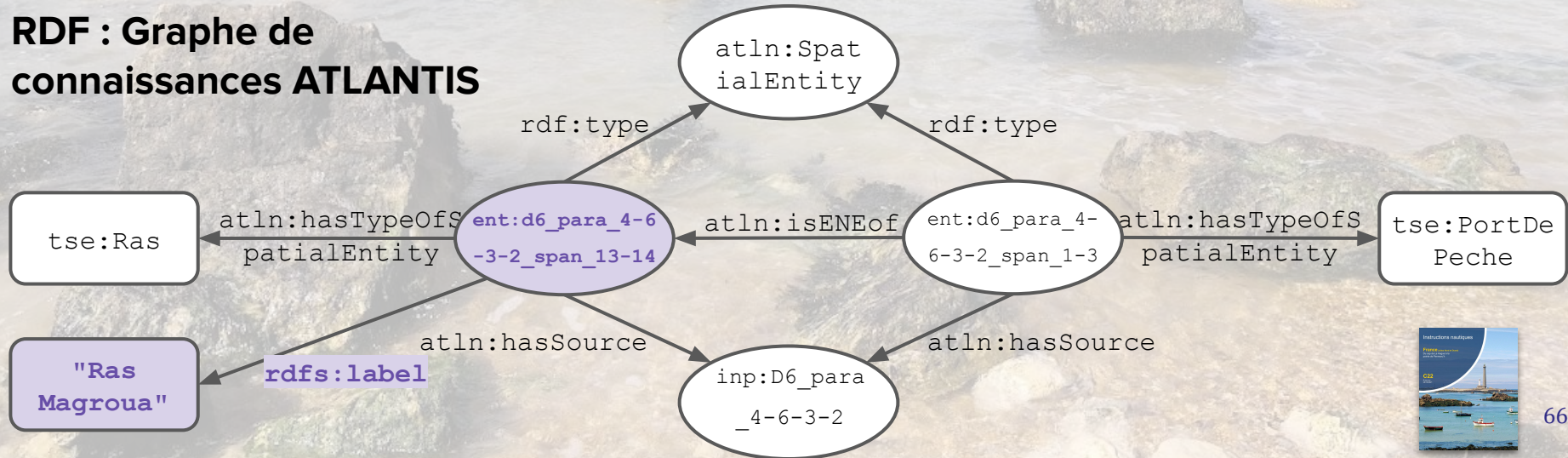
66

JSONL

```
{ "doc_key": "d6_para_4-6-3-2",  
  "dataset": "atlantis",  
  "sentences": [ ["Un", "port", "de", "pêche", "est", "établi", "à", "5,7", "M", "à", "l'",  
"ENE", "de", "Ras", "Magroua", "." ] ],  
  "ner": [ [ [1, 3, "geogFeat"], [13, 13, "geogFeat"], [14, 14, "name"], [13, 14, "geogName"] ] ],  
  "relations": [ [ [1, 3, 13, 14, "isENEof"] ] ] }
```

is ENE of

RDF : Graphe de connaissances ATLANTIS



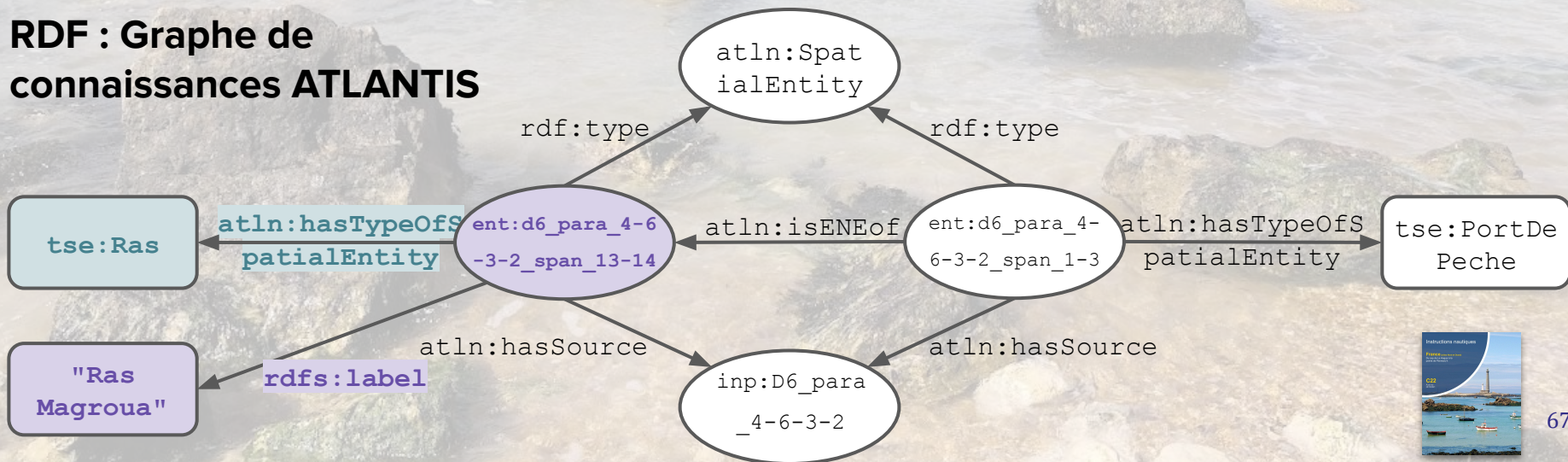
Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

JSONL

```
{ "doc_key": "d6_para_4-6-3-2",
  "dataset": "atlantis",
  "sentences": [ ["Un", "port", "de", "pêche", "est", "établi", "à", "5,7", "M", "à", "l'",
    "ENE", "de", "Ras", "Magroua", "." ] ],
  "ner": [ [ [1, 3, "geogFeat"], [13, 13, "geogFeat"], [14, 14, "name"], [13, 14, "geogName"] ] ],
  "relations": [ [ [1, 3, 13, 14, "isENEof"] ] ] }
```

is ENE of

RDF : Graphe de connaissances ATLANTIS



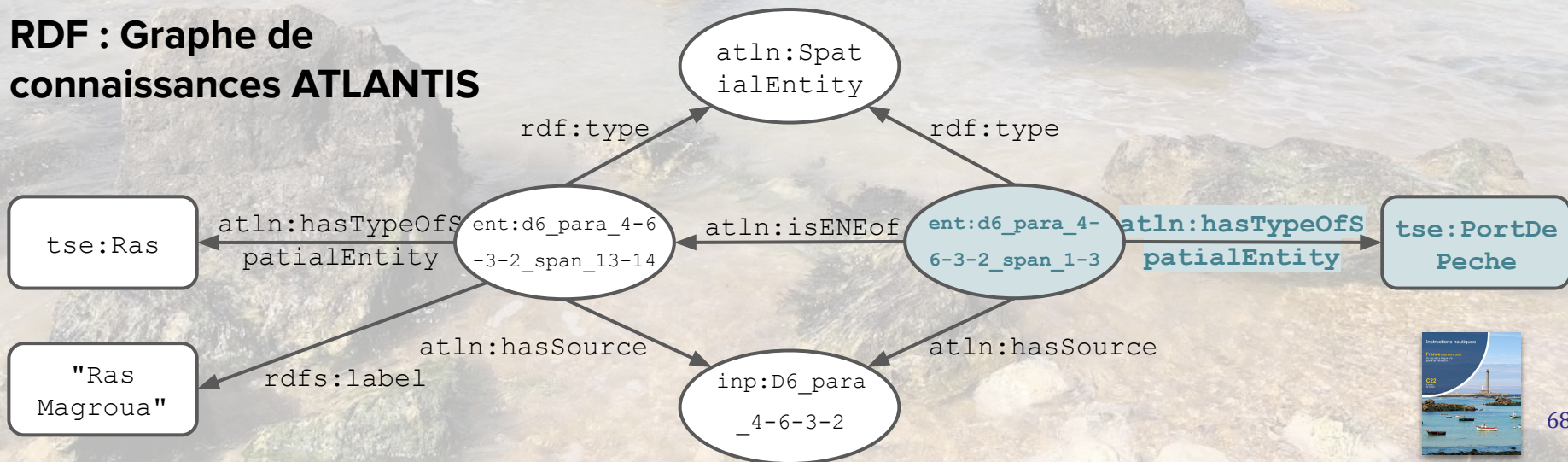
Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

JSONL

```
{ "doc_key": "d6_para_4-6-3-2",
  "dataset": "atlantis",
  "sentences": [ ["Un", "port", "de", "pêche", "est", "établi", "à", "5,7", "M", "à", "l'",
    "ENE", "de", "Ras", "Magroua", "." ] ],
  "ner": [ [ [1, 3, "geogFeat"], [13, 13, "geogFeat"], [14, 14, "name"], [13, 14, "geogName"] ] ],
  "relations": [ [ [1, 3, 13, 14, "isENEof"] ] ] }
```

is ENE of

RDF : Graphe de connaissances ATLANTIS



Structurer les informations extraites en RDF

- SPARQL-Generate (Lefrançois et al. 2017)
- Extraction d'entités spatiales imbriquées : classées automatiquement
- Extraction de relations spatiales : type dans l'étiquette

Un port de pêche est établi à 5,7 M à l'ENE de Ras Magroua .

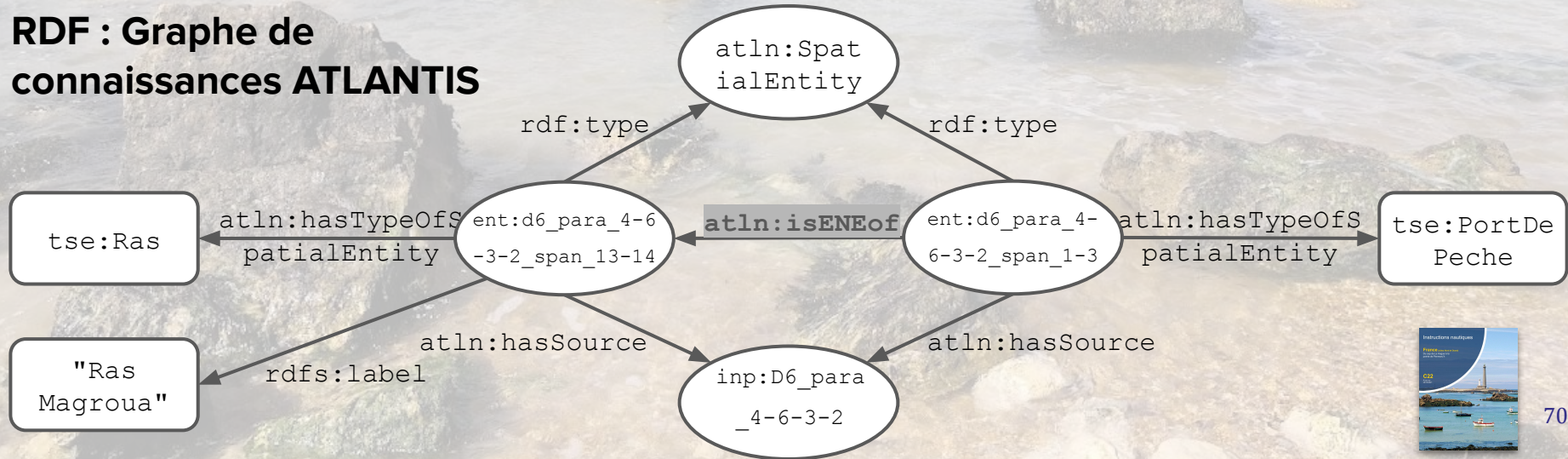
70

JSONL

```
{ "doc_key": "d6_para_4-6-3-2",  
  "dataset": "atlantis",  
  "sentences": [ ["Un", "port", "de", "pêche", "est", "établi", "à", "5,7", "M", "à", "l'",  
"ENE", "de", "Ras", "Magroua", "." ] ],  
  "ner": [ [ [1, 3, "geogFeat"], [13, 13, "geogFeat"], [14, 14, "name"], [13, 14, "geogName"] ] ],  
  "relations": [ [ [1, 3, 13, 14, "isENEof"] ] ] }
```

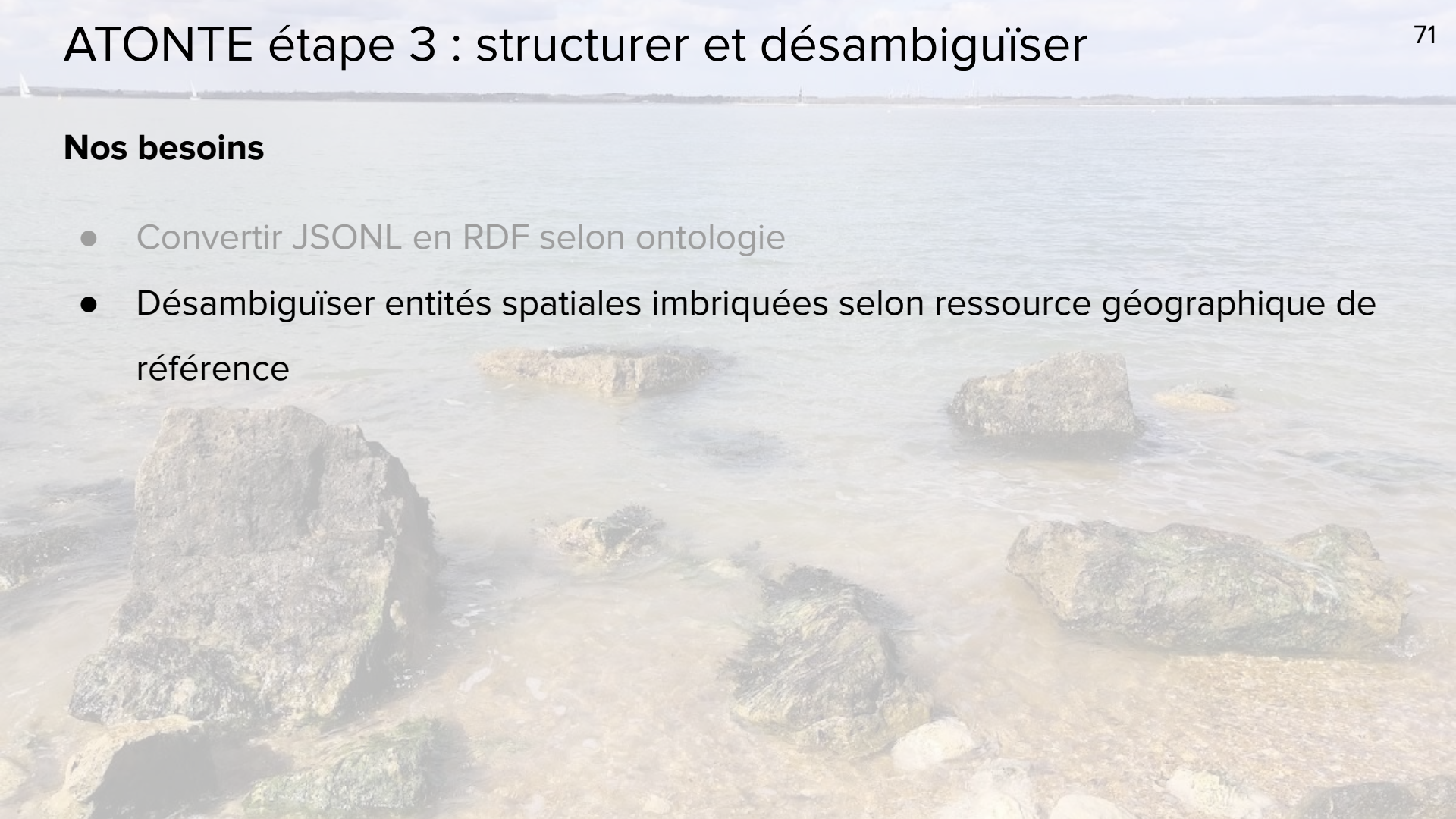
is ENE of

RDF : Graphe de connaissances ATLANTIS



Nos besoins

- Convertir JSONL en RDF selon ontologie
- Désambiguïser entités spatiales imbriquées selon ressource géographique de référence



Nos besoins

- Convertir JSONL en RDF selon ontologie
- Désambiguïser entités spatiales imbriquées selon ressource géographique de référence
 - Lier une entité spatiale à son entrée correspondante dans une base de données géographique de référence
 - ➡ Associer des coordonnées géographiques

Travaux antérieurs (Loynes and Ruiz 2020)

- Désambiguïstation d'entités spatiales plates des *Instructions nautiques*

- En utilisant le full name

Notre approche (Rawsthorne 2024)

- Désambiguïstation d'entités spatiales imbriquées des *Instructions nautiques*

- En utilisant les full name et geographic feature type

- Ressource : BD TOPO® de l'IGN (nom, type, coordonnées géographiques, etc.)
- Sélection de candidates : similarité des chaînes de caractères
- Notation et classement de candidates : similarité des chaînes de caractères ou de distance géographique (position moyenne de l'ensemble des candidats du paragraphe)

Travaux antérieurs (Loynes and Ruiz 2020)

- Terme de recherche dans l'attribut

full name :

- Port de Douarnenez

- Résultats dans le BD TOPO® :

Nom	Note

Notre approche (Rawsthorne 2024)

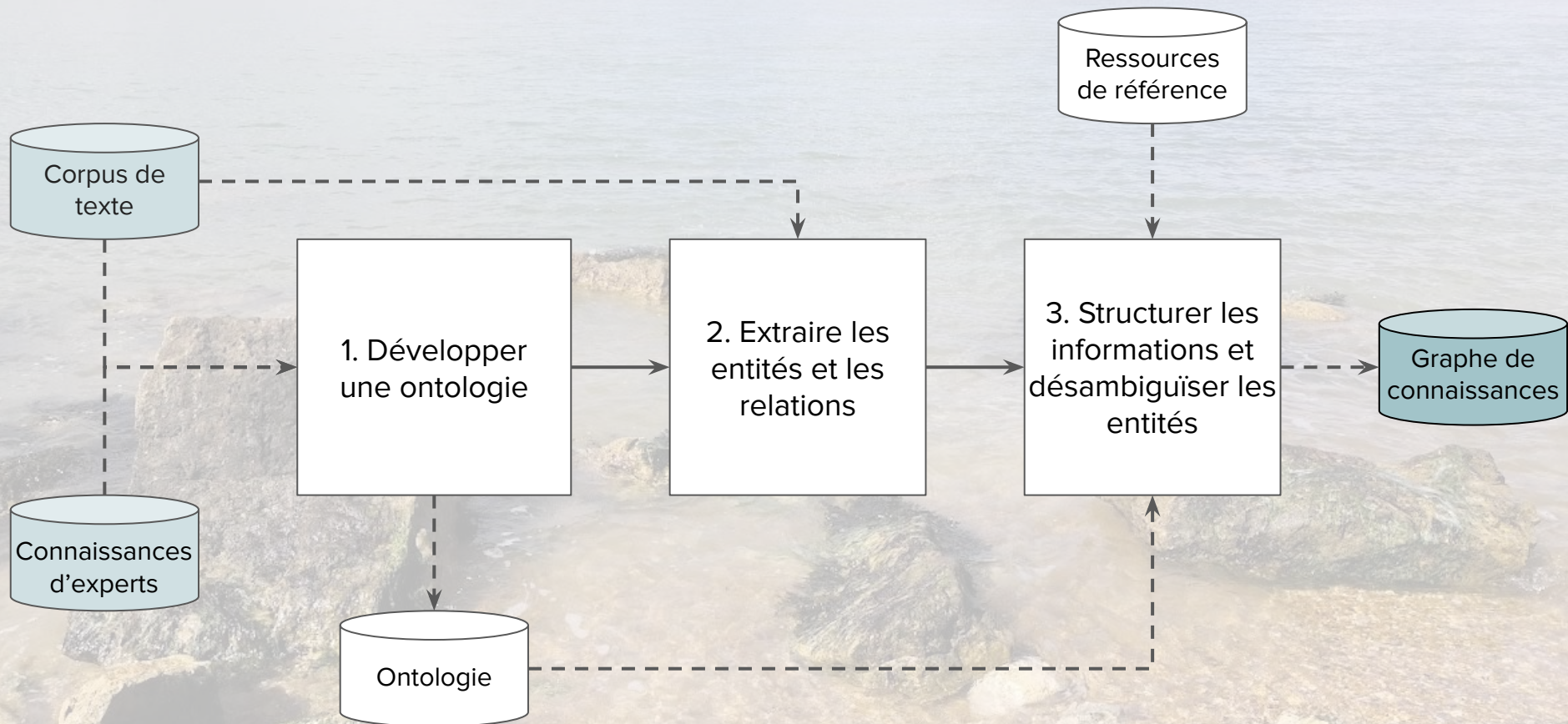
- Termes de recherche dans les attributs

full name et geographic feature type :

- Port de Douarnenez

- Résultats dans le BD TOPO® :

Nom	Type	Note
port de pêche de douarnenez	port	16.31
port-musée de douarnenez	musée	11.41
capitainerie du port de pêche de douarnenez	capitainerie	8.86





Balise

Bouee

Chenal

Espar

FeuASecteur



Lat : 45.8301427156 Lng : -13.4587444144

100 km

(Alla et al. 2022)

Nereus Login: Mdp: [Connexion](#) [Le SHOM](#) [?](#)

☒ Balise
☐ Bouee
☐ Chenal
☐ Espar
☐ FeuAsecteur

Map showing a coastal area with a yellow landmass and blue water. A red buoy is visible in the water. A play button icon is at the bottom center.



(Alla et al. 2022)

Nereus Login: Mdp: [Connexion](#) [Le SHOM](#) [?](#)

Resultats

type

- Tourelle

label

- Men Guen Bras

reference

- C22 page 400

type

- Tourelle

label

- Malvoch

reference

- C22 page 400

Map showing a coastal area with a yellow landmass and blue water. A red buoy is visible in the water. A play button icon is at the bottom center.

type

- Tourelle

label

- Men Guen Bras

reference

- C22 page 400

Map showing a coastal area with a yellow landmass and blue water. A red buoy is visible in the water. A play button icon is at the bottom center.

Des Heulve-Brehat au nord du Four

5.4.2.2 — Ile de Suez au NW (2018) (DTON 071-AT)

25 De nombreux clochers jonchent la côte. Côté de Rouvres (48° 18.46' N — 4° 10.46' W), qui domine l'anne de Kermic, et celui de Plouder (48° 38.54' N — 4° 17.82' W), près d'un chateau d'eau, à 4 M dans la SSE du port de Brignogan-Pagan, sont les plus remarquables. Les clochers de Saint (48° 39.80' N — 4° 05.24' W) et de Cader (48° 35.79' N — 4° 05.17' W) encadrent un chateau d'eau bien visible à 1 M au Sud de l'île de Suez. Le port d'échouage de Moguinec est dominé au SE par un phare (48° 41.12' N — 4° 04.38' W), tour de Suez, haute de 11 m, à sommet vert.

26 L'amélioration intersectorielle n° 2013350-0003 du 18 décembre 2013 (modifié) du préfet maritime de l'Atlantique et du préfet du Finistère-Amélie, sur le rivage de Suez, deux zones de mouillage et d'équipement léger divisées en six secteurs, deux dans l'anne du Pouldu d'une part et quatre au Sud de l'île Verte d'autre part. Leur accès est limité en raison des nombreuses zones qui encadrent leur approche. Ces zones sont situées : 250 mètres d'un quai au moins, jusqu'aux rives de passage. Ces zones sont exploitées du 15 avril au 15 octobre chaque année, la vitesse y est limitée à 3 nœuds.

27 MOULAGES DATTENTE — Ces mouillages, utilisés notamment par les navires se rendant à Roscoff, se trouvent dans l'Est et dans l'Ouest du chenal. Un des meilleurs se trouve au NE de la tourelle « Men Guen Bras », par profondeur de 14 m, fond de sable et coquille de bonne tenue. Prendre garde à la roche couverte de 2.8 m en face à 0.3 M à l'ENE de la tourelle.

28 Par vents d'Est, on peut mouiller partout dans la partie Ouest du chenal jusqu'au Sud de la tourelle « Malvoch », implantée près de l'extrémité de la péninsule de Porz Kernok.

5.4.3. Port de Porz Kernok (île de Batz)

5.4.3.1 Généralités

29 Le port d'échouage de Porz Kernok (48° 44.27' N — 4° 00.83' W), situé sur la côte Sud de l'île de Batz et bien protégé de tous les vents, sert d'abri aux pêcheurs de l'île.

30 Ce port, aux ressources limitées, est relié à Roscoff par un service de vedettes.

Lat : 48.7310639499 Lng : -3.9693260193 500 m

- Méthodologie ATONTE pour la création de graphes de connaissances géospatiaux à partir de sources hétérogènes
 - Développement d'une ontologie à partir de texte et d'experts
 - Extraction d'entités spatiales imbriquées et de relations spatiales à partir de texte
 - Structuration d'information et désambiguïsation d'entités spatiales
- Ontologie ATLANTIS : github.com/umrlastig/atlantis-ontology
- Jeu de données ATLANTIS : github.com/umrlastig/atlantis-dataset
- Jeu de données TextMine'24 : kaggle.com/competitions/defi-textmine-2024



Merci pour votre attention !

Helen Mair RAWSTHORNE

Ingénieure de recherche, Mines Saint-Étienne
helenmair.rawsthorne@emse.fr