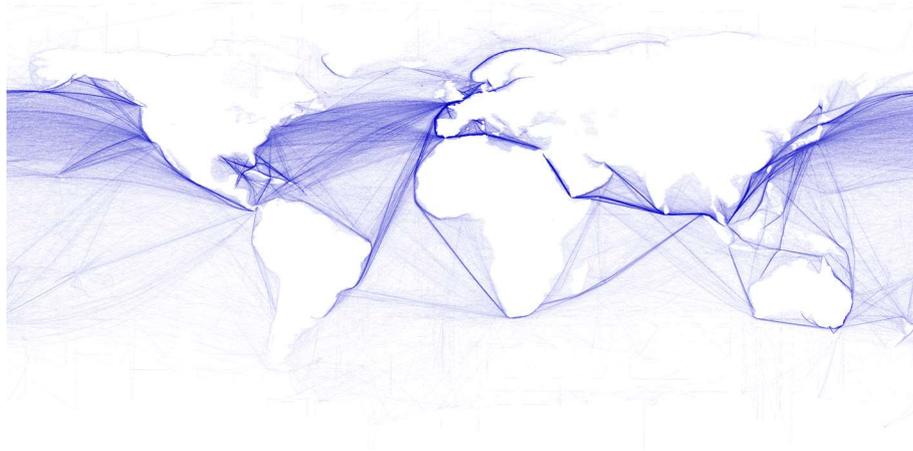


Étude, conception, collecte, curation et évaluation d'un scraping de sites web liés au transport maritime pour améliorer la prédiction du fret de matières premières



Travail de Master réalisé par :
Guy DRUEY

Sous la direction de :
Arnaud GAUDINAT, professeur HES

Carouge, le 17 août 2020

**Master en Sciences de l'information
Haute école de gestion de Genève (HEG-GE)**

Remerciements

Je remercie chaleureusement :

François Can et Arnaud Gaudinat pour leur disponibilité, leur aide et leur confiance ainsi que pour m'avoir donné l'opportunité de travailler sur ce mandat.

David Dällenbach pour sa patience et nos échanges qui ont oscillé entre vallée du désespoir et pente de l'illumination.

Anouk Santos pour sa présence, sa patience et tout le reste.

Ma famille et mes proches pour leur indéfectible soutien.

Tomáš Linhart qui m'a permis de comprendre pourquoi je ne parvenais pas à obtenir des résultats avec Splash (the solution is to disable Private mode !)

Joel Miller qui nous a quitté trop tôt.

Source de l'image de titre :

HENGL, T., 2008. Shipping density (commercial) [en ligne]. [Consulté le 7 août 2020]. Disponible à l'adresse : https://commons.wikimedia.org/wiki/File:Shipping_routes.png

Résumé

Ce travail de master s'inscrit comme étape préalable dans le projet Innosuisse *Precise Intelligence*, ayant comme partenaires la HES-SO Genève et la société de logistique et de courtage maritime RiverLake. Ce projet vise à compléter les données sur les flux du transport maritime de matières premières avec l'aide, notamment, de données présentes sur le web public.

Comme son titre l'indique, cette étude a été réalisée en plusieurs étapes. Tout d'abord, l'examen du domaine du transport maritime de matières premières et de ses différents acteurs a permis la création d'une liste de 100 sources web qui correspond à des ports commerciaux proposant des données sur le web public. De ces différentes sources, un échantillon de onze ports a été retenu. L'analyse de la typologie des données disponibles dans cet échantillon a permis alors la réalisation d'un référentiel normalisé, dans le but d'identifier les données qui seront plus tard récupérées. Enfin, une preuve de concept consistant au *web scraping* de ces différentes sources durant 3 jours, du 25 au 27 juin 2020, a rendu possible la collecte de 3'322 enregistrements comprenant des données sur l'identité et le voyage de 1'472 navires uniques. Ces enregistrements ont été nettoyés et uniformisés dans la mesure du possible avec le logiciel OpenRefine.

Les limites de ce travail sont : 1) l'impossibilité de certifier l'origine des données publiques récoltées ; 2) la qualité des données, notamment en ce qui concerne leur exactitude, n'est pas comparée ni vérifiée ; 3) les données n'indiquent pas, à de rares exceptions, la cargaison transportée ; 4) les données proviennent uniquement des sites web de ports commerciaux.

Malgré ces limites, les perspectives offertes par la méthode employée dans ce travail et les données récupérées indiquent la possibilité de reconstruire la trajectoire de différents navires. Cette conclusion est essentielle pour pouvoir déterminer dans une seconde étape le niveau de complétude des données issues du web public par rapport aux données déjà acquises, grâce à l'AIS notamment.

Les données récupérées et analysées dans ce mémoire sont disponibles à l'adresse suivante : <https://doi.org/10.5281/zenodo.3980515>

Mots clés : Big data, Web scraping, maritime data, data curation, Scrapy, OpenRefine

Table des matières

Remerciements	1
Résumé	2
Liste des tableaux	6
Liste des figures	7
Abréviations	8
1. Introduction	1
1.1 Contexte	1
1.2 Mandat	1
1.3 Objectifs	1
2. Le transport maritime de matières premières	3
2.1.1 Les différents types de fret.....	4
2.1.1.1 Différents bateaux pour différentes marchandises.....	5
2.1.2 Les différents acteurs	6
2.2 Les données disponibles	7
2.2.1 AIS	7
2.2.2 L'intérêt du web	8
2.3 Web mining et données maritimes	9
3. État des lieux des sites web des acteurs du transport maritime	10
3.1 Méthodologie	10
3.1.1 Choix d'un acteur : les ports	10
3.1.1.1 Méthodologie de recherche : les ports maritimes commerciaux.....	10
3.1.1.2 Choix et échantillonnage	11
3.1.2 Un second acteur : les agents maritimes	13
3.2 Synthèse	14
3.2.1 De l'absence de grands ports	14
3.2.2 De l'hétérogénéité des technologies web et de l'information.....	15
3.2.3 Des généralisations selon les pays.....	15
3.2.4 Des difficultés linguistiques.....	15
3.2.5 Des sources inconnues	15
4. Référentiel normalisé	16
4.1 Normaliser les données récupérées	16
4.1.1 Concept de voyage.....	16
4.1.2 Données liées à l'identification.....	16
4.1.3 Données liées au voyage	17
4.1.4 Données techniques liées au bateau	19
4.2 Synthèse	19
5. Preuve de concept	21
5.1 Web scraping & data mining	21

5.2	Méthodologie	21
5.2.1	À la recherche de l'outil idéal	21
5.2.2	Le <i>framework</i> Scrapy	22
5.2.2.1	Outils supplémentaires	23
5.2.2.2	Versions utilisées	23
5.2.3	Considérations juridiques	23
5.2.4	Différents scripts pour différents ports	24
5.3	Synthèse	26
5.3.1	Listes des scripts par ports	27
6.	Évaluation des résultats	28
6.1	État des données récupérées	28
6.1.1	<i>Scraping</i> sur trois jours	28
6.2	Nettoyage des données	29
6.2.1	Élimination des doublons	29
6.2.2	Uniformisation et nettoyage des champs	30
6.2.3	Dépôt des données	31
6.3	Résultats par ports	31
6.3.1	Rotterdam	32
6.3.2	Hambourg	33
6.3.3	Amsterdam	35
6.3.4	Le Havre	36
6.3.5	Dunkerque	38
6.3.6	Klaipeda	40
6.3.7	Niedersachsen	41
6.3.8	Copenhague	42
6.3.9	Aarhus	43
6.3.10	Fredericia	44
6.3.11	Bordeaux	44
6.4	Synthèse	46
6.4.1	Remarques générales	46
6.4.2	Cas particuliers	48
6.4.2.1	Du manque de données d'identification	48
6.4.2.2	Du décryptage d'un voyage	48
7.	Difficultés et recommandations	50
7.1	De la problématique linguistique à la normalisation des données	50
7.2	La recherche de nouvelles données exploitables	50
7.2.1	Veilles envisagées	51
7.3	Le <i>web scraping</i> de données	51
7.3.1	Uniformisation et réconciliation	52
7.4	Prochaines étapes	53
7.4.1	Que faire avec ces données ?	53
7.4.2	Ouvrir la méthode au reste du monde	53

8. Conclusion	54
Bibliographie	55
Annexe 1 : Évolution du trafic maritime mondial	61
Annexe 2 : Catégorisation des navires de commerce.....	62
Annexe 3 : Autorités portuaires nationales identifiées	64
Annexe 4 : Sources web, sites ports commerciaux au 24.04.2020	65
Annexe 5 : Sources web, agents maritimes	87
Annexe 6 : Référentiel normalisé	90
Annexe 7 : Scripts des <i>spiders</i> de Scrapy	92

Liste des tableaux

Tableau 1 : Types de navires dédiés au transport de marchandises.....	5
Tableau 2 : Classement des ports échantillonnés par marchandises transbordées en 2018	13
Tableau 3 : Données liées à l'identification	17
Tableau 4 : Données liées au voyage	18
Tableau 5 : Liste des différents scripts réalisés pour chaque port.....	27
Tableau 6 : Enregistrements récupérés, totaux et moyenne	29
Tableau 7 : Port de Rotterdam, taux de réponse des données récupérées.....	33
Tableau 8 : Port de Hambourg, réconciliation du type de navire	34
Tableau 9 : Port de Hambourg, taux de réponse des données récupérées.....	35
Tableau 10 : Port d'Amsterdam, taux de réponse des données récupérées.....	36
Tableau 11 : Port du Havre, réconciliation du type de navire	37
Tableau 12 : Port du Havre, taux de réponse des données récupérées.....	38
Tableau 13 : Port de Dunkerque, taux de réponse des données récupérées	39
Tableau 14 : Port de Klaipeda, taux de réponse des données récupérées.....	41
Tableau 15 : Ports de Niedersachsen, taux de réponse des données récupérées	42
Tableau 16 : Port de Copenhague, taux de réponse des données récupérées	43
Tableau 17 : Port d'Aarhus, taux de réponse des données récupérées	44
Tableau 18 : Port de Fredericia, taux de réponse des données récupérées.....	44
Tableau 19 : Port de Bordeaux, réconciliation du champ "Type".....	45
Tableau 20 : Port de Bordeaux, taux de réponse des données récupérées	46
Tableau 21 : Synthèse des données récupérées par port	47

Liste des figures

Figure 1 : Évolution du trafic maritime international (en millions de tonnes chargées).....	3
Figure 2 : Trafic maritime international par type de marchandises (en millions de tonnes chargées).....	3
Figure 3 : Trafic maritime mondial par région en 2018 (en pourcentage du tonnage mondial)	4
Figure 4 : Top 20 des ports, transport de marchandises 2013-2018	14
Figure 5 : Quantité d'enregistrements récupérés par port le 25 juin 2020	28
Figure 6 : Quantité de données <i>scrapées</i> du 25 au 27 juin 2020	29
Figure 7 : Totaux des enregistrements récupérés et quantité de doublons	30
Figure 8 : Port de Rotterdam, détail par jour des enregistrements récupérés.....	32
Figure 9 : Port de Hambourg, détail par jour des enregistrements récupérés.....	34
Figure 10 : Port d'Amsterdam, détail des enregistrements récupérés	36
Figure 11 : Port du Havre, détail des enregistrements récupérés.....	37
Figure 12 : Port de Dunkerque, détail des enregistrements récupérés.....	39
Figure 13 : Port de Klaipeda, détail des enregistrements récupérés	40
Figure 14 : Ports de Niedersachsen, détail des enregistrements récupérés.....	41
Figure 15 : Port de Copenhague, détail des enregistrements récupérés	42
Figure 16 : Port d'Aarhus, détail des enregistrements récupérés	43
Figure 17 : Port de Fredericia, détail des enregistrements récupérés	44
Figure 18 : Port de Bordeaux, détail des enregistrements récupérés	45
Figure 19 : <i>Amethyst</i> , un même nom pour deux navires	48
Figure 20 : Le voyage du <i>Bomar Quest</i>	49
Figure 21 : Le voyage du <i>MSC Geneva</i>	49

Abréviations

AIS / SIA	Automatic Identification System / Système Automatique d'Identification
ATA	Actual Time of Arrival
ATD	Actual Time of Departure
CSS	Cascading Style Sheet
DWT	Deadweight Tonnage
ETA	Estimated Time of Arrival
ETD	Estimated Time of Departure
GRT	Gross Registered Tonnage
IMO / OMI	International Maritime Organization / Organisation Maritime Internationale
MMSI	Maritime Mobile Service Identity
TEU / EVP	Twenty-Foot Equivalent Units / Équivalent Vingt Pieds
UCRN	Unique (Port) Call Reference Number
XPATH	XML Path Language

1. Introduction

1.1 Contexte

La gigantesque quantité de données issue du transport maritime permet aujourd'hui d'assurer la sécurité des navires et la transparence du fret. Ces données sont fournies principalement par des outils comme le système d'identification automatique, ou Automatic Identification System (AIS), un système d'échanges automatisés de messages entre navires (Serry et Lévêque 2015), ainsi que par des systèmes de positionnement par satellite toujours plus automatisés et performants.

Cependant, certaines informations, basées uniquement sur les données AIS, comme par exemple la taille des cargaisons, le type de matières premières ou encore l'heure d'arrivée estimée (ETA) sont souvent imprécises, car mal signalées, inconnues ou même parfois falsifiées (Can, Gaudinat et Theodoro 2020).

Pour combler ces lacunes, une des solutions possibles se trouve sur le web public. En effet, un nombre important d'acteurs du trafic maritime comme les ports, les agents maritimes (*shipping agents*) ou encore les affréteurs (*charterers*) semblent fournir des données riches et précises sur leurs sites web, comme par exemple les heures d'arrivées et de départs des navires dans les ports (Can, Gaudinat et Theodoro 2020). Ces données sont cependant souvent proposées de manière hétérogène et non structurée, elles se retrouvent sous des formes et des formats très différents. Pour être exploitées comme sources d'information, il faut donc d'abord identifier ces sources, pour ensuite extraire l'information, notamment grâce aux techniques issues du *web scraping*, puis enfin nettoyer et structurer les données.

1.2 Mandat

Ce travail de master s'inscrit dans le projet Innosuisse «*Precise Intelligence : Big Data Analytics for Comprehensive Global Trade Flow Intelligence* » (Can, Gaudinat et Theodoro 2020) qui associe la HES-SO Genève et la société genevoise de logistique et de courtage maritime Riverlake Shipping SA¹. Ce projet a pour but de compléter les données sur les flux commerciaux des marchandises avec des informations provenant de multiples sources, notamment les données du web public, pour ensuite permettre de travailler sur la prévision du trafic maritime des matières premières.

1.3 Objectifs

Ce travail de Master vise à remplir les objectifs généraux suivants :

- Effectuer un état de l'art du web mining de données maritimes en particulier et de données hétérogènes et distribuées en général.
- Effectuer un état des lieux des sites web des acteurs du transport maritime de matières premières.
- Réaliser un référentiel normalisé pour le transport maritime de matières premières.
- Réaliser une preuve de concept de *web scraping* et de curation de données de sites web des acteurs du secteur.

¹ <https://www.riverlake.ch/en/riverlake-group.php>

- Établir des recommandations pour la généralisation de l'approche.

2. Le transport maritime de matières premières

Le transport par voie maritime est le mode le plus utilisé pour l'acheminement de marchandises : environ 90% du transit commercial mondial se fait aujourd'hui par la marine marchande (International Chamber of Shipping 2020). D'après la revue annuelle 2019 du transport maritime de la Conférence des Nations unies sur le commerce et le développement (2020), le volume de marchandises transportées par voie de mer en 2018 était de 11 milliards de tonnes, soit un nouveau record, avec une croissance estimée à 2,6% pour l'année 2019 qui est, elle, inférieure à la moyenne historique de 3% entre 1970 et 2017.

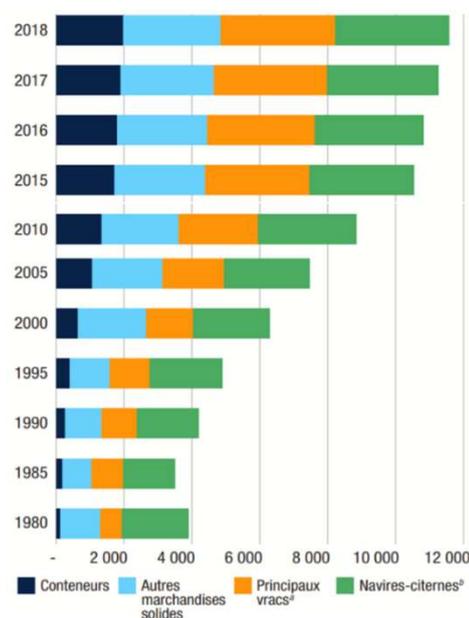
Le ralentissement de cette croissance, qui suit logiquement la même tendance que la croissance du commerce mondial, s'explique par un contexte d'incertitude accrue. Notamment par les tensions commerciales croissantes entre la Chine et les États-Unis, par la géopolitique, avec par exemple les tensions autour du couloir maritime stratégique que représente le détroit d'Ormuz, ou encore par les préoccupations environnementales dues au réchauffement climatique (Conférence des Nations Unies sur le commerce et le développement 2020).

Les figures suivantes sont extraites de l'étude sur les transports maritimes 2019 de la Conférence des Nations Unies sur le commerce et le développement. Elles peuvent être consultées dans leur intégralité dans l'annexe 1 du présent mémoire :

Figure 1 : Évolution du trafic maritime international (en millions de tonnes chargées)

Année	Navires-citernes ^a	Principaux vrac ^b	Autres marchandises solides ^c	Total (toutes catégories)
1980	1 871	608	1 225	3 704
1990	1 755	988	1 265	4 008
2000	2 163	1 186	2 635	5 984
2005	2 422	1 579	3 108	7 109
2010	2 752	2 232	3 423	8 408
2015	2 932	2 930	4 161	10 023
2016	3 058	3 009	4 228	10 295
2017	3 146	3 151	4 419	10 716
2018	3 194	3 210	4 601	11 005

Figure 2 : Trafic maritime international par type de marchandises (en millions de tonnes chargées)



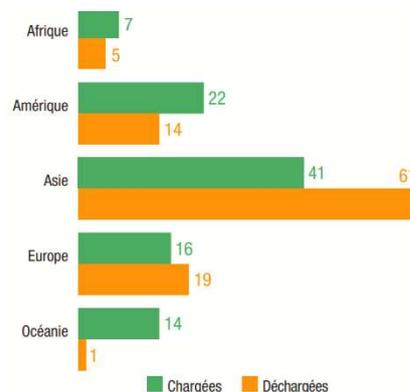
(Adapté de Conférence des Nations Unies sur le commerce et le développement 2020)

Ces deux figures expriment l'évolution du commerce maritime par type de marchandises. Ainsi pour l'année 2018, le transport dans les navires citernes, que ce soit du pétrole brut ou raffiné,

du gaz ou des produits chimiques, a représenté 29% du total du commerce maritime mondial. Les principaux vrac, minéral de fer, céréales et charbon notamment, ont été évalués à quasiment la même proportion, alors que les autres marchandises solides, combinées aux marchandises conteneurisées, s'élevaient quant à elles à 42% du trafic maritime mondial.

Pour finir de contextualiser notre étude, il est important de relever encore la répartition régionale du trafic maritime mondial grâce la figure 3 ci-contre, qui révèle l'importance de l'Asie dans l'import-export : plus de la moitié des marchandises mondiales y sont déchargées. Même si cette figure n'exprime pas la proportion de matières premières dans sa représentation, elle n'en est pas moins très significative des régions qui pourraient proportionnellement fournir le plus de données sur le web.

Figure 3 : Trafic maritime mondial par région en 2018 (en pourcentage du tonnage mondial)



(Conférence des Nations Unies sur le commerce et le développement, 2020)

Nous allons détailler dans la suite de ce chapitre les différents types de fret, ou cargaison, en mettant l'accent sur les matières premières, puis les différents types de navires et les différents acteurs du trafic maritime mondial pour ensuite nous interroger sur les données qui décrivent ce secteur, notamment dans le web public.

2.1.1 Les différents types de fret

Les matières premières que l'on retrouve dans le commerce international peuvent être répertoriées en trois principales catégories : les métaux et minerais, les matières énergétiques et les denrées issues de l'agriculture (Lioudis 2020 ; Trafigura 2020). Ainsi, pour l'Association suisse du négoce de matières premières et du transport maritime (STSA), le transport maritime de matières premières énergétiques implique le charbon, le gaz naturel liquéfié, le pétrole et les produits pétroliers raffinés (Swiss Trading and Shipping Association 2020).

Dans l'univers du commerce maritime on utilisera des terminologies spécifiques au conditionnement des marchandises pour les catégoriser, c'est-à-dire la manière dont elles sont manutentionnées et acheminées (Service des données et études statistiques 2019). Ainsi on considère tout d'abord le vrac, *bulk cargo*, qui désigne les marchandises qui ne sont ni emballées, ni arrimées et transportées dans de grandes quantités. Ce vrac peut être liquide ou solide :

- Le vrac liquide, *liquid bulk*, regroupe les produits directement transportés sous forme liquide dans des navire-citernes (pétroliers, chimiques, etc.). On trouve notamment dans cette catégorie le pétrole brut, les produits pétroliers raffinés, les biocarburants, les huiles végétales, le gaz naturel liquéfié ou encore les produits chimiques à l'état liquide (Ly et Martin 2018).
- Le vrac solide, *solid bulk*, regroupe quant à lui des produits à l'état solide le plus souvent transportés dans des vraquiers. On trouve dans cette catégorie plusieurs types de matières premières, comme le charbon, les minerais, les céréales, le fourrage, les oléagineux, la biomasse, les engrais, ou encore les gravats, le sable, les granulats, etc. (Ly et Martin 2018).

On considère ensuite les marchandises conteneurisées, c'est-à-dire les produits et matières transportés à l'intérieur de conteneurs. Attention, on peut très bien retrouver un produit liquide, comme un hydrocarbure par exemple, conditionné pour être transporté dans un conteneur. La nature même de la marchandise transportée se différencie donc de la manière dont elle est conditionnée (Ly et Martin 2018).

Les marchandises en roulier sont des véhicules de transport chargés dans des navires de commerce. Enfin, les marchandises diverses, *general cargo*, regroupent les cargaisons qui ne se retrouvent pas dans les catégories mentionnées précédemment. On y retrouvera notamment des produits aux formats non conventionnels, généralement lourds et encombrants. Ces deux dernières catégories n'entrant pas dans le registre des matières premières, elles ne nous intéresseront que modérément dans le contexte de ce travail.

2.1.1.1 Différents bateaux pour différentes marchandises

Les évolutions dans l'industrie de la construction navale ont apporté des changements importants dans le transport maritime de marchandises. Ainsi, de plus en plus de navires spécifiques furent construits pour différents types de fret, diminuant ainsi les coûts engendrés par le transport (Swiss trading and shipping association 2020).

Selon la Conférence des Nations Unies sur le commerce et le développement (2020), on dénombrait en 2019 une flotte mondiale de 92'295 navires commerciaux. On peut catégoriser cette flotte suivant la nature du conditionnement des matières transportées, comme présenté dans le tableau 1 :

Tableau 1 : Types de navires dédiés au transport de marchandises

Transport de marchandises solides	
General cargo ships <i>Navires de charge</i> <i>Navires polyvalents</i> <i>Cargos de divers</i> <i>Cargos mixtes</i>	Ils transportent du matériel encombrant, n'entrant pas dans des conteneurs. Cela peut comprendre des marchandises lourdes pour des projets de construction de centrales électriques, installations offshore, etc.
Container ships <i>Porte-conteneurs</i>	Bien plus gros que les navires de charge, les porte-conteneurs sont, comme leur nom l'indique, destinés uniquement au transport de conteneurs. Leur taille s'évalue en quantité de conteneurs de 20 pieds (EVP) qu'ils peuvent arrimer. Les plus gros porte-conteneurs font près de 400 m de long et peuvent transporter jusqu'à 20'000 EVP.
Refrigerated cargo ships <i>Navires réfrigérés</i>	Les navires réfrigérés transportent des denrées périssables (fruits, légumes, poissons, viandes, etc.) à basse température.
Roll-On Roll-Off ships <i>Rouliers (Ro/Ro)</i>	Ces navires peuvent transporter différents types de véhicules sur plusieurs ponts.
Bulk carriers <i>Vraquiers</i>	Les vraquiers transportent des marchandises solides en vrac comme du minerai, du charbon, des céréales, etc.
Transport de marchandises liquides	

<p>Tankers <i>Navire-citernes</i></p>	<p>Les navire-citernes sont utilisés pour le transport de liquides en vrac, que ce soit dans des citernes ou des cuves. On en trouve plusieurs types selon la marchandise :</p> <ul style="list-style-type: none"> - Les pétroliers (<i>crude oil tanker, oil products tanker</i>) pour le pétrole brut ou les produits pétroliers raffinés - Les chimiquiers (<i>chemical tanker</i>) pour les produits chimiques mais aussi pour le vin, le jus de fruit, etc. - Les gaziers (<i>LNG tanker, LPG tanker</i>) pour les gaz liquéfiés, <i>liquefied natural gas</i> et <i>liquefied petroleum gas</i>.
--	---

(Tupper 2013 ; Navire de charge 2020)

De ce tableau, il sort que les navires les plus intéressants pour notre étude, c'est-à-dire ceux qui transportent le plus souvent des matières premières, sont les vraquiers et les navire-citernes. Le détail exhaustif des catégories de navires transportant des marchandises peut être consulté dans l'annexe 2.

2.1.2 Les différents acteurs

Il existe de nombreux intervenants dans le monde du commerce maritime. Afin de déterminer lesquels pourraient s'affirmer comme sources de données potentielles, il est pertinent de réaliser un inventaire de ces différents acteurs. Attention toutefois, certains n'ont pas la même terminologie, ni le même rôle selon les régions linguistiques.

L'armateur (*ship-owner*) : l'armateur arme le navire, c'est-à-dire qu'il a la responsabilité de son entretien, de l'embarquement d'un équipage compétent, du matériel, et de tout ce qui est nécessaire à l'expédition maritime (Durand [sans date]a). Si l'armateur est en général le propriétaire du navire, il peut également en être le « loueur », c'est-à-dire l'affréteur (Armateurs de France 2017). Il porte donc la responsabilité de transporter une marchandise d'un port A à un port B. Les plus grands armateurs mondiaux sont aujourd'hui APM-Maersk, Mediterranean Shg Co et CMA CGM Group (M2R Maritime 2018).

L'affréteur (*charterer*) : l'affréteur est une personne physique ou morale qui loue un navire pour une période donnée ou pour un voyage donné (Armateurs de France 2017).

L'agent maritime (*shipping agent*) / **consignataire de navire** : l'agent maritime est, avec le consignataire, le représentant de l'armateur ou de l'affréteur dans un port lors de l'escale d'un navire. Ils ont pour rôle de faciliter la préparation de l'escale, grâce à leur connaissance du port, du pays et des intervenants sur place. Chacun des deux acteurs s'occupe d'un aspect de cette représentation. L'agent maritime s'occupe de l'aspect commercial, donc de la gestion des marchandises chargées et déchargées et du recrutement du fret, alors que le consignataire gère les aspects techniques, les formalités administratives liées à la cargaison ainsi que les besoins du navire et de l'équipage. Un seul métier peut gérer les deux aspects, on parle alors d'agent consignataire (Durand [sans date]b).

Le courtier maritime (*shipbroker*) : le courtier maritime est une personne physique ou morale qui joue le rôle d'intermédiaire facilitant la négociation entre deux acteurs, que ce soit des

armateurs, des affréteurs, d'un acteur ayant des marchandises à transporter et d'un acteur ayant un navire correspondant à ses besoins. Il s'agit le plus souvent de négociations autour de l'achat, la vente ou la location d'un navire neuf ou d'occasion (Manaadiar 2019).

Le commissionnaire de transport / le transitaire (*freight forwarder*) : Il s'agit d'un intermédiaire commercial qui va se charger de gérer, pour le compte de l'expéditeur ou du destinataire, le transport d'une marchandise d'un point A à un point B. En contact avec l'agent maritime, il va choisir entre différents transporteurs selon la nature de la marchandise (Transport maritime 2020).

Les infrastructures portuaires : les ports de commerce du monde entier ont tendance à remplir de multiples missions et à se spécialiser dans le secteur des vracs liquides ou solides, des marchandises diverses ou encore des hydrocarbures, ce qui en fait des acteurs clés dans le commerce maritime.

2.2 Les données disponibles

Cette section va nous permettre de faire le point sur les données de référence concernant le trafic maritime avec l'AIS, puis de nous concentrer sur ce que peut nous apporter le web public en terme d'information.

2.2.1 AIS

Le système automatique d'identification (AIS) est un dispositif d'échange de données entre navires qui a pour but d'améliorer la sécurité du trafic maritime tout en rendant sa gestion plus efficace (Serry et Lévêque 2015). Développé dans les années 1990, il a été conçu pour être capable d'envoyer et de recevoir automatiquement des informations d'un navire à un autre ou d'un navire aux autorités côtières et réciproquement. Il a été rendu obligatoire par l'Organisation Maritime Internationale (IMO), selon certaines dispositions², depuis le 31 décembre 2004 (Organisation Maritime Internationale 2020a).

Initialement, les données AIS étaient difficiles à collecter, car fortement régionalisées. Les communications étaient limitées à la gamme des ondes radios à très hautes fréquences (VHF) qui ne couvraient que 10 à 20 milles marins³. Ce n'est que depuis 2008 que des satellites sont équipés de récepteurs AIS, qui leur permettent de recevoir des données AIS transmises par les navires du monde entier (Yang, Wu, Wang, Jia et Li 2020).

Les transpondeurs AIS permettent l'envoi et la réception de données multiples, statiques et dynamiques, et en temps réel : on y trouve notamment les informations sur l'identité du navire et sa taille, la position du navire (latitude et longitude) la vitesse, la direction de navigation, ainsi que des données liées au voyage, comme le port de destination et l'heure estimée d'arrivée (Serry et Lévêque 2015 ; Yang, Wu, Wang, Jia et Li 2020).

Les avantages de l'AIS sont variés : le système permet l'amélioration de la sécurité, de la gestion de la flotte, de la voie maritime et de la navigation (Serry et Lévêque 2015). La richesse des données offre également un très grand potentiel dans les multiples développements de l'analyse du trafic maritime. Avec l'amélioration constante de leur qualité et de leur exhaustivité, les applications se basant sur l'exploitation des données AIS se sont étendues

² C'est-à-dire pour tous les navires qui sont engagés dans des voyages internationaux et qui disposent d'un tonnage de plus de 300, ainsi que pour tous les navires à passagers.

³ 18 à 37 kilomètres.

de la sécurité du trafic maritime à de nombreux autres domaines, profitant de la couverture mondiale des données, de leur précision, de leur disponibilité en temps réel et de leur abondance (Yang, Wu, Wang, Jia et Li 2020).

Cependant, tout n'est pas parfait dans les données proposées par l'AIS. En effet, de nombreuses études ont remarqué des erreurs, oublis et même des fraudes. Watson, Holm et Lind (2015) constatent que des imprécisions ou des champs non renseignés en ce qui concerne les heures d'arrivées estimées ont lieu fréquemment. Pour Harati-Mokhtari, Wall, Brooks et Wang (2007), des approximations dans les données concernant l'identification des navires et dans les données de localisation sont remarquées. Enfin, Kazemi, Abghari, Lavesson, Johnson et Ryman (2013) remarquent de nombreuses anomalies dans les données, causées par des erreurs humaines, des mises à jour irrégulières ou encore des erreurs dans les formats. Selon Adland, Jia et Strandenes (2017), il n'existe pas de règlement qui sanctionne les informations frauduleuses entrées manuellement dans le système AIS, comme le tirant d'eau (*draught*) ou la destination par exemple.

Les problèmes de l'AIS sont directement liés à ses avantages. Puisque le système fournit de grandes quantités de données qui sont, à l'origine, gratuites et libres, ces dernières sont difficilement contrôlables. Le système pourrait donc même être mis en cause dans la multiplication des actes de piraterie (Serry et Lévêque 2015).

2.2.2 L'intérêt du web

L'existence de récepteurs AIS satellitaires permet de collecter très facilement les données à l'échelle mondiale et ce, en quasi-temps réel. Ainsi de nombreux sites web commerciaux donnent accès aux bases de données AIS (Yang, Wu, Wang, Jia et Li 2020 ; Kalyvas, Kokkos et Tzouramanis 2017).

On peut mentionner quelques-uns des sites web commerciaux de suivis et de surveillances des navires les plus populaires :

- MarineTraffic⁴, un des sites web les plus connus, développé par l'Université grecque de l'Égée (Kalyvas, Kokkos et Tzouramanis 2017), propose de nombreux services basés sur des algorithmes qui utilisent en majorité les informations proposées en temps réel par les données AIS. Il est également possible d'utiliser leur API qui propose des informations détaillées sur les positions des navires, leurs données, ou encore les prédictions de voyages.
- VesselTracker⁵, qui fournit à ses membres enregistrés des données AIS personnalisées sur les mouvements de navires. Il propose également une couverture en temps réel des principaux ports du monde entier.
- VesselFinder⁶, un autre site web très populaire qui offre des informations similaires aux deux sites précédents, notamment dans le suivi du trafic maritime. Il se distingue par sa proposition de fournir ses services gratuitement à tout partenaire qui lui procurerait des données AIS en temps réel fiable pour compléter son réseau.
- Fleetmon⁷, qui permet le suivi en direct de navires et d'évènements maritimes dans différents ports du monde entier.

⁴ <https://www.marinetraffic.com/>

⁵ <https://www.vesseltracker.com/>

⁶ <https://www.vesselfinder.com/>

⁷ <https://www.fleetmon.com/>

D'autres sites web contiennent des bases de données qui proposent des informations détaillées, historiques et actuelles, sur les navires de commerces qui parcourent les mers :

- La fondation ISSF (International Seafood Sustainability Foundation) propose une base de données gratuite qui répertorie les numéros IMO⁸. Cette dernière est compilée à partir de plusieurs sources, dont le registre d'IHS Markit, une entreprise américaine d'information économique.
- Le portail maritime (payant) d'IHS Markit⁹ fournit des données sur plus de 220'000 navires ainsi qu'un suivi en temps réel basé sur l'AIS.
- La base données Equasis¹⁰ fournit gratuitement à ses membres enregistrés des données diverses concernant les navires et les compagnies qui les possèdent, comme l'historique de propriété, de l'équipage, de l'inspection, etc.
- La base de données du Bureau Veritas¹¹, une société active dans la classification et la surveillance des navires, propose également des informations sur les navires et les compagnies maritimes, tout comme la base de données Maritime-Connector¹².

Bien qu'utiles, notamment pour augmenter et compléter des données brutes, ces sites web généralistes ne sont que des acteurs secondaires du transport maritime. Pour ce travail, on préférera trouver sur le web des sources de données primaires, comme les ports ou les agents maritimes par exemple.

2.3 Web mining et données maritimes

On retrouve de nombreuses études appliquant des méthodes issues du *data mining* aux données AIS. Zhu (2011) utilise ces techniques afin d'explorer de larges jeux de données AIS et ainsi développer des prédictions dans les flux maritimes. De la même manière, Lei (2020) s'intéresse au volume croissant de données AIS pour prévenir les collisions entre navires. D'autres études développent de nouvelles applications basées sur les données fournies par l'AIS comme nous l'avons vu dans le chapitre 2.2.1. Cependant, aucune de ces études n'évoquent l'exploitation de données maritimes basée sur ce qui est disponible librement sur le web. Nous n'avons donc pas trouvé de littérature spécifique au *web scraping* pour ce qui concerne exclusivement les données maritimes, ce qui confirme donc la dimension exploratoire de cette étude.

Si l'on élargit le spectre à des domaines voisins, comme la prédiction logistique, Veldhuis (2015) et Bol Raap (2016) proposent, eux, d'utiliser la technique du *web scraping* en s'intéressant notamment à des solutions logicielles comme Cloudscrape, RapidMiner, WebHarvest ou Mozenda.

Dans un domaine encore plus éloigné, Vargiu et Urru (2012) s'intéressent au *web scraping* dans le domaine de la publicité sur le web. Landers, Brusso, Cavanaugh et Collmus (2016) proposent, eux, une étude sur le *web scraping* en recherche psychologique, développent une méthode en 4 étapes sur la gestion de projet utilisant la technique du *web scraping*, et s'intéressent particulièrement au langage de programmation Python et à l'outil Scrapy.

⁸ <https://iss-foundation.org/knowledge-tools/databases/imo-database/>

⁹ <https://ihsmarkit.com/products/maritime-ship-tracker-ais-live-ship-data-seaweb.html>

¹⁰ <http://www.equasis.org/>

¹¹ <https://marine-offshore.bureauveritas.com/bv-fleet/#/bv-fleet/>

¹² <http://maritime-connector.com/>

3. État des lieux des sites web des acteurs du transport maritime

3.1 Méthodologie

Effectuer un état des lieux des sites web de tous les acteurs du transport maritime n'était pas un objectif réalisable dans le temps imparti pour ce travail. Il a donc fallu choisir les acteurs les plus prometteurs en termes de données disponibles sur le web pour réaliser ensuite un état des lieux le plus complet possible des données pouvant être récupérées.

En terme quantitatif, cet état des lieux suit la ligne des livrables quantifiables du projet *Precise Intelligence* (Can, Gaudinat et Theodoro 2020) en visant un total de 100 sources web accompagnées du détail de la typologie des données accessibles pour du *web scraping*. Il s'agit d'une étape indispensable pour la réalisation du référentiel normalisé que l'on découvrira dans le chapitre suivant.

3.1.1 Choix d'un acteur : les ports

Le choix des ports maritimes comme première source de données n'est pas lié au hasard. Il s'agit, pour les plus importants d'entre eux, d'infrastructures régulées et répertoriées par des organismes d'état, on peut donc relativement facilement les répertorier par pays et obtenir des statistiques concernant leur importance à l'échelle internationale, voire nationale. De plus, il s'agit tout simplement de l'acteur du transport maritime proposant le plus de données sur le web public. Certains ports proposent également dans leurs données des informations sur d'autres acteurs du transport maritime, comme les agents, opérateurs ou affréteurs, ce qui peut aider dans la réalisation de l'état des lieux d'un second acteur.

La région de l'Europe géographique a été privilégiée comme point de départ pour l'état des lieux. Cette approche se justifie principalement par l'aspect exploratoire de ce travail. Pour des raisons de proximité et de linguistique, on préférera commencer à *scrap* des données compréhensibles avant de proposer une généralisation de l'approche aux pays du continent asiatique, dans des sites web de langue russe ou chinoise, et ce malgré l'importance du trafic maritime dans cette région, comme nous l'avons vu au chapitre 2. Précisons encore que les pays sans accès à un littoral n'ont bien évidemment pas été retenus dans cet état des lieux.

3.1.1.1 Méthodologie de recherche : les ports maritimes commerciaux

L'état des lieux des sites web des ports commerciaux maritimes a nécessité une approche méthodologique en trois points, en appliquant une recherche d'information de plus en plus spécifique.

Tout d'abord, une recherche des autorités portuaires nationales, ou association de ports nationaux, des différents pays européens a été réalisée. Ces dernières répertorient en général les principaux ports commerciaux de leur pays. Le tableau de l'annexe 3 répertorie les autorités portuaires nationales identifiées.

Ensuite, pour les quelques pays où l'autorité nationale n'a pas pu être identifiée, la recherche d'information dans des sites web spécialisés, comme *world port source*¹³ ou *searates*¹⁴, a permis de compléter l'état des lieux.

Enfin, étant donné le résultat quantitatif recherché, c'est-à-dire une centaine de source web représentant autant de ports différents, le but n'était pas d'établir un catalogue exhaustif des ports européens. Un échantillonnage s'est donc fait selon trois critères, surtout pour les pays ayant beaucoup d'infrastructures portuaires :

- La disponibilité de données sur leur site web.
- L'importance du trafic commercial, et si possible concernant les matières premières.
- L'exclusion des ports de plaisance et ceux proposant uniquement du transport de passagers et de véhicules (*Roll-On Roll-Off ships*).

Il faut préciser que certains ports ne proposant pas de données sur leur site web ont malgré tout été répertoriés, comme par exemple les grands ports européens d'Anvers ou d'Algeciras. Le but est ici d'indiquer que la disponibilité de données publiques sur le web n'est pas un fait établi et que certains ports, voir certains pays de manière générale, ne divulguent pas les informations concernant les navires qui y font escale.

3.1.1.2 Choix et échantillonnage

De ce premier état des lieux et en accord avec les partenaires du projet *Precise Intelligence*, François Can de Riverlake et Arnaud Gaudinat de la HEG, il a été décidé de sélectionner onze ports pour y effectuer un référentiel normalisé et une preuve de concept que l'on retrouvera dans les chapitres suivants.

Plusieurs critères ont été invoqués dans cette sélection : la présence de données pertinentes répertoriées dans l'état des lieux, l'importance du trafic maritime de matières premières, ainsi que d'autres critères liés aux intérêts de Riverlake ou aux autres développements du projet *Precise Intelligence*. C'est ainsi que la liste suivante a été réalisée :

1. **Hambourg (DE)**

Fondé par Frédéric Barberousse le 7 mai 1189 (Chabot 1952), le port de Hambourg est aujourd'hui le 3^{ème} port européen dans la manutention de marchandises et le plus grand port de commerce d'Allemagne. Il est spécialisé dans le trafic de conteneurs (61%), mais gère également du vrac solide (26%) et du vrac liquide (11%) (Eurostat 2020).

2. **Niedersachsen (DE)**

La société *Niedersachsen Ports*, ou ports de Basse-Saxe en français, exploite 15 ports, dont notamment cinq ports maritimes : Emden, Norden, Wilhelmshaven, Brake et Cuxhaven. Précisons que le port de Wilhelmshaven est le port maritime le plus profond d'Allemagne et qu'il est, après les ports de Hambourg et de Brême, le 3^{ème} port de commerce le plus important du pays. Il gère 80% du transbordement de pétrole brut parmi tous les ports maritimes allemands (Wilhelmshavener Häfen 2020).

3. **Copenhague (DK)**

¹³ <http://www.worldportsource.com/>

¹⁴ <https://www.searates.com/maritime/>

Le port de Copenhague, dont l'origine remonte au Moyen Âge, est le plus grand port danois et l'un des plus importants de la mer Baltique. Il est exploité par la société Copenhague Malmö Port AB (CMP) qui gère également le port de Malmö, les deux ports n'étant séparés que par le détroit d'Øresund (Port of Copenhagen 2020).

4. Aarhus (DK)

Situé dans une des plus vieilles villes danoises, le port en eau profonde d'Aarhus gère plus de 50% du trafic de conteneurs du pays. Il possède plusieurs terminaux spécifiques dont un terminal pétrolier pour les huiles minérales, l'essence et les produits chimiques notamment, et un terminal de vrac pour le grain, le sable et le gravier. Le port d'Aarhus a transbordé 8,4 millions de tonnes de fret en 2017 (Port of Aarhus 2020 ; Aarhus Havn 2020).

5. Fredericia (DK)

Le port de Fredericia est le plus grand port à trafic de conteneurs de la région sud du Danemark avec environ 7.9 millions de tonnes de fret transbordés par année. Il est également l'un des plus grands ports du Danemark pour l'importation et l'exportation de marchandises en vrac solides telles que les céréales et les produits d'alimentation animale. Pour terminer, il traite également du pétrole raffiné et des produits pétroliers bruts (ADP A/S [sans date]).

6. Bordeaux (FR)

Le port de Bordeaux est le 7^{ème} plus grand port maritime français¹⁵. Il dispose de 7 terminaux spécialisés le long d'un estuaire de plus de 100 kilomètres, qui accueillent chaque année environ 1100 navires. Son trafic maritime extérieur est d'environ 6.8 millions de tonnes en 2019, dont plus de 50% est consacré au vrac liquide, et plus précisément aux produits pétroliers (Bordeaux Port Atlantique 2020).

7. Dunkerque (FR)

Alors que ses origines remontent au XI^e siècle, le port de Dunkerque est aujourd'hui le 3^{ème} plus grand port maritime français. Il a traité 51,6 millions de tonnes de fret en 2018, dont principalement des minerais et du charbon (41%), des marchandises diverses (39%) et du vrac liquide (10%) (Dunkerque Port 2019).

8. Le Havre (FR)

Fondé en 1530 (Grand port maritime du Havre 2020), le port du Havre est le 2^{ème} plus grand port maritime français, derrière Marseille, avec près de 65 millions de tonnes de fret en 2018. Il se classe au 8^{ème} rang des ports européens en ce qui concerne le transport de marchandises. Il est spécialisé dans le transport de vrac liquide (61%) et de conteneurs (35%) (Eurostat 2020).

9. Amsterdam (NL)

2^{ème} port des Pays-Bas et 4^{ème} port européen, le port d'Amsterdam a traité 99,5 millions de tonnes de marchandises en 2018, dont principalement du vrac liquide (47%) et solide (44%) (Eurostat 2020).

10. Rotterdam (NL)

Ayant son origine au XII^e siècle, le port de Rotterdam est aujourd'hui le port européen le plus important avec ses 441,5 millions de tonnes de marchandises traitées en 2018, loin devant le port belge d'Anvers (212 mt). Son activité se

¹⁵ Il existe sept grands ports maritimes de métropole en France. Ce statut signifie qu'il s'agit d'établissements publics de l'État chargés de la gestion de leur zone portuaire (Grand port maritime 2020).

retrouve principalement dans la gestion du vrac liquide (47%), des conteneurs (28%) et du vrac solide (17%) (Eurostat 2020).

11. Klaipeda (LT)

Le port de Klaipeda, plus grand port de Lituanie, est un acteur important du commerce dans la mer Baltique avec 46,3 millions de tonnes de marchandises traitées en 2019 (Port of Klaipeda 2020a).

Ces ports peuvent être classés selon leur importance, que l'on définit d'après le critère de la marchandise transbordée. Le tableau 2 permet ainsi de nous rendre compte des différents écarts des ports de notre échantillon :

Tableau 2 : Classement des ports échantillonnés par marchandises transbordées en 2018

	Port	Marchandises transbordées en 2018 (en million de tonnes)	Classement européen 2018 (Eurostat 2020)
1	Rotterdam	441,5 (Eurostat 2020)	1
2	Hambourg	117,6 (Eurostat 2020)	3
3	Amsterdam	99,5 (Eurostat 2020)	4
4	Le Havre	64,9 (Eurostat 2020)	8
5	Dunkerque	51,6 (Dunkerque Port 2019)	-
6	Klaipeda	46,6 (Port of Klaipeda 2020b)	-
7	Niedersachsen	28,3 (Niedersachsen Ports 2020)	-
8	Copenhague	15,1 ¹⁶ (Copenhagen Malmö Port 2020)	-
9	Aarhus	8,5 (Statistics Denmark 2020)	-
10	Fredericia	6,9 (Statistics Denmark 2020)	-
11	Bordeaux	6,8 ¹⁷ (Bordeaux Port Atlantique 2020)	-

3.1.2 Un second acteur : les agents maritimes

Dans l'idée de compléter, ou même d'augmenter les données déjà proposées par l'état des lieux des ports maritimes, la recherche de données proposées par les agents maritimes a été envisagée, avec notamment pour but la récupération des rapports d'agents.

Pour ce faire, les agents maritimes proposés dans les enregistrements des onze ports sélectionnés ont été répertoriés (voir annexe 5). Cette liste non exhaustive a notamment permis de se rendre compte des imprécisions du champ « Agent » : les données révèlent non seulement la présence d'agents, mais aussi d'opérateurs et d'affréteurs. En plus de ce premier constat, il a été établi que les sites web de ces acteurs ne contiennent pas de données publiques qui puissent compléter ou augmenter la typologie des données des ports maritimes déjà répertoriée.

¹⁶ Comprend le port de Malmö en Suède également.

¹⁷ Pour l'année 2019.

3.2 Synthèse

L'état des lieux complet des ports maritimes commerciaux et des données qu'ils mettent à disposition sur le web, effectué à la date du 24 avril 2020, est accessible dans l'annexe 4 de ce travail. La réalisation de cet inventaire a conduit à certaines remarques.

3.2.1 De l'absence de grands ports

La moitié des 20 ports européens les plus importants selon les statistiques de l'Union européenne (Eurostat 2020) ne fournissent pas de données sur le web public. Ainsi, Anvers, Marseille, Algéciras, Barcelone, Gênes, le Pirée, ainsi que la totalité des ports turcs indiquent ne pas proposer de données exploitables dans notre état des lieux.

Figure 4 : Top 20 des ports, transport de marchandises 2013-2018

Top 20 ports handling freight, 2013-2018
(million tonnes)

Rank 2018	Port	*	2013-2017					2018					Change 2018/2017 (%)	Change 2018/2013 (%)			
			2013	2014	2015	2016	2017	By direction			By type of cargo handled (%)						
			Total	Total	Total	Total	Total	Inwards	Outwards	Total	Liquid bulk goods	Dry bulk goods			Large containers	Ro-Ro Mobile units	Other cargo
1	Rotterdam (NL)	=	411.9	418.6	436.9	431.9	433.3	303.2	138.3	441.5	47	17	28	3	5	1.9	7.2
2	Antwerpen (BE)	=	172.0	180.4	190.1	198.7	201.2	109.2	102.8	212.0	35	6	51	3	5	5.4	23.3
3	Hamburg (DE)	=	120.6	126.0	120.2	120.3	118.8	70.5	47.2	117.6	11	26	61	0	1	-1.0	-2.4
4	Amsterdam (NL)	=	92.3	96.3	98.8	96.3	98.5	64.0	35.5	99.5	47	44	1	1	8	1.0	7.7
5	Algéciras (ES)	=	67.6	75.6	79.4	83.4	83.5	48.6	40.1	88.6	36	2	57	1	4	6.2	31.1
6	Marseille (FR) (*)	=	76.2	74.4	77.5	76.4	75.6	55.1	20.6	75.7	60	20	14	3	3	0.1	-0.8
7	Izmit (TR)	=	60.7	58.6	64.2	66.0	72.7	48.3	24.2	72.4	35	31	23	0	11	-0.4	19.4
8	Le Havre (FR)	+1	56.1	57.0	78.1	78.4	70.9	46.5	18.4	64.9	61	2	35	2	0	-8.5	15.7
9	Valencia (ES)	+1	64.4	61.4	62.9	60.0	66.1	28.4	33.6	62.0	3	4	76	3	14	-6.3	-3.8
10	Botas (TR)	-2	53.5	55.0	57.6	58.3	60.1	7.9	52.8	60.7	92	7	0	0	0	1.0	13.6
11	Iskenderun- Hatay (TR)	=	46.0	47.3	49.1	49.3	55.2	38.4	19.0	57.5	11	76	9	0	3	4.2	25.0
12	Trieste (IT)	=	62.6	59.4	59.1	54.4	54.0	45.6	11.8	57.4	57	6	15	12	9	6.2	-8.4
13	Immingham (UK)	+1	39.5	42.0	48.4	50.1	55.3	41.7	14.0	55.6	34	30	4	30	2	0.5	40.8
14	Barcelone (ES)	+3	62.6	59.4	59.1	54.4	55.0	29.3	25.3	54.6	28	8	51	12	1	-0.8	-12.9
15	Aliaga (TR)	-2	34.4	41.2	38.0	39.1	49.7	34.5	18.7	53.3	47	33	16	0	4	7.1	55.0
16	London (UK)	=	43.2	44.5	45.4	50.4	49.9	44.8	8.4	53.2	29	28	26	15	3	6.7	23.1
17	Genova (IT)	-2	40.8	43.4	43.4	45.0	50.7	31.0	20.5	51.6	35	4	42	19	0	1.8	26.3
18	Bremerhaven (DE)	=	54.5	53.6	49.8	52.3	49.3	22.1	29.1	51.2	1	0	89	9	1	3.8	-6.1
19	Piraeus (EL)	+2	40.2	41.4	38.3	41.0	45.2	26.0	25.0	50.9	2	1	88	10	0	12.7	26.7
20	Bergen (NO)	-1	51.8	42.1	43.6	44.7	48.1	9.5	34.8	44.3	92	5	0	0	2	-7.9	-14.4
Total top 20 ports (*)			1 634.1	1 658.3	1 723.4	1 746.2	1 794.4	1 104.4	720.1	1 824.4	39	17	35	5	4	1.7	11.6

Note: (*) column indicates number of positions lost or gained compared to 2017.

(*) 2013-2014: partially estimated by Eurostat.

(*) Total figure for the ports being part of the top 20 ports of the countries reporting data during the reference year concerned.

Source: Eurostat (online data code: mar_mg_aa_pwhd)

eurostat 

(Eurostat 2020)

On peut fournir quelques raisons quant à ces absences. La première étant simplement qu'à la date de l'état des lieux, les données n'étaient pas disponibles ou que la page concernée était en construction. C'est le cas notamment avec le port de Marseille, dont la liste des escales est désormais disponible¹⁸. Une seconde raison de l'absence de données réside dans le fait que certains ports disposent de services numériques externalisés et dédiés, entre autres, à la logistique des mouvements de navires. C'est le cas du port d'Anvers avec son service *Antwerp Port Community System* (Port of Antwerp 2019) et sa plateforme C-point¹⁹. En l'occurrence cette plateforme renvoie de l'information uniquement si l'on connaît déjà le numéro IMO d'un navire ou son nom.

¹⁸ <https://pcs.marseille-port.fr/webix/public/escales>

¹⁹ <https://www.c-point.be/en>

D'autres grands ports qui ne fournissent pas de données sur les escales des navires redirigent l'internaute sur les sites web de leurs terminaux. C'est le cas des ports espagnols d'Algeciras et de Barcelone. Rares sont alors les terminaux fournissant des données sur le trafic maritime, et lorsque c'est le cas, on se retrouve rapidement confronté à la nécessité de créer un compte sur leur site web. C'est le cas par exemple avec APM Terminals pour le port de Barcelone²⁰. Enfin, certains ports nécessitent une authentification pour accéder à l'information. C'est le cas du site web du port de Gênes par exemple²¹.

3.2.2 De l'hétérogénéité des technologies web et de l'information

Cet état des lieux permet de se rendre compte de l'immense diversité en ce qui concerne la typologie des données disponibles ainsi que la manière dont les données sont proposées sur le web. Certains ports fournissent des données très détaillées, principalement sur les caractéristiques techniques des navires, mais également sur des données liées au voyage en cours, alors que d'autres ports se limitent uniquement à l'identification du nom et de l'heure d'arrivée. Les technologies web sont également très variées, on y trouve de l'HTML, du Javascript, parfois couplé avec de l'AJAX, ainsi que des API dédiées.

3.2.3 Des généralisations selon les pays

Si l'on généralise l'accessibilité des données des ports de notre état de lieux aux pays dont ils dépendent, on remarque que certains pays offrent, pour quasiment chacun de leurs ports, des données sur le trafic maritime, alors que d'autres au contraire, ne fournissent pratiquement aucune information. Ainsi, on peut observer que les pays scandinaves (Danemark, Norvège, Suède et Finlande) sont généralement plus ouverts à la mise à disposition de données sur les sites de leurs ports. À l'inverse, des pays comme l'Italie et la Turquie n'offrent quasiment aucune données.

3.2.4 Des difficultés linguistiques

Ce premier état des lieux limité à l'Europe géographique indique déjà une des principales difficultés que l'on pourra rencontrer en généralisant notre approche au reste du monde. La problématique linguistique a été rencontrée dans la recherche d'information sur certains sites de ports grecs et turcs par exemple qui ne proposent pas de version de leur site en anglais.

3.2.5 Des sources inconnues

Une dernière remarque est particulièrement importante : la source des données disponibles sur les sites des ports n'est que très rarement indiquée. Il n'est donc pas impossible que les données puissent provenir de l' AIS, notamment les données dynamiques relatives aux heures d'arrivées et de départs. Dans ce cas, on obtiendrait des données redondantes qui contiendraient les mêmes erreurs et approximations que les données que l'on espère pouvoir compléter et corriger.

²⁰ <https://www.apmterminals.com/barcelona>

²¹ <http://www.pcs-eport.it/eportHomePage/#>

4. Référentiel normalisé

Devant la grande diversité des données proposées par les différentes pages web répertoriées dans l'état des lieux, il paraît impératif de faire l'inventaire de leur typologie afin de cibler ce que l'on va récupérer et d'anticiper les données qui nécessiteront un travail d'uniformisation.

Plusieurs causalités amènent cette diversité. On peut par exemple évoquer les différentes langues des sites web rencontrés. Dans l'échantillon sélectionné, on trouve notamment des données en français, en anglais, mais aussi en allemand, en danois et en lituanien. Cette pluralité de langages implique aussi l'existence de différentes formes d'abréviations, ainsi que de différentes unités de mesure.

4.1 Normaliser les données récupérées

On peut répartir le type de données qu'offrent les différents site web des ports échantillonnés dans trois catégories. Premièrement, les données utiles à l'identification des navires, ensuite les données dynamiques en lien avec le voyage en cours et finalement, les données relatives aux caractéristiques techniques des navires.

4.1.1 Concept de voyage

Il convient de rappeler ici le but du projet *Precise Intelligence* dans lequel s'inscrit ce travail, c'est-à-dire la volonté de compléter et/ou de vérifier des données issues de l'AIS, tout cela dans un but de prédiction de voyages maritimes. On aura donc beaucoup d'attentes à propos des données dites dynamiques, c'est-à-dire celles qui font référence, au moment du *scraping*, au déplacement particulier d'un navire spécifique.

Ainsi, si l'on obtient les informations relatives à la provenance, la destination et aux heures d'arrivée et de départ d'un navire au sein de plusieurs ports, il deviendra alors possible de recréer le voyage d'un navire, du moment que les données qui l'identifient sont suffisantes. C'est pourquoi on se concentrera sur ces deux catégories de données, c'est-à-dire celles liées à l'identification et au voyage, pour créer un référentiel qui servira de norme pour les différents ports de notre échantillon.

4.1.2 Données liées à l'identification

Les données liées à l'identification des navires sont les données statiques suivantes :

- Nom : le nom du navire est la seule donnée à se retrouver, logiquement, dans tous les ports sélectionnés. À noter que cette donnée ne suffit pas à l'identification d'un bateau car plusieurs navires peuvent avoir le même nom.
- Pavillon : le pavillon, que l'on retrouve aussi sous les termes de *Flag*, *Flag code*, ou *Nationality*, permet d'identifier la nationalité du navire, et par conséquent sous quelle juridiction nationale il se retrouve affilié (Mansell 2009). Les données se retrouvent ici sous différentes formes, abrégées ou non, et en différentes langues.
- IMO : le numéro IMO (International Maritime Organization) est un identifiant unique pour les bateaux, obligatoire pour tous les navires depuis le 1^{er} janvier 1996. Il s'agit d'un numéro à 7 chiffres « qui est attribué aux navires de commerce océaniques à propulsion d'une jauge brute égale ou supérieure à 100, au moment de la pose de la

quille [...] » (Organisation Maritime Internationale 2020b). À noter que cet identifiant est conservé si le navire change de pavillon ou de propriétaire.

- **MMSI** : le numéro MMSI (Maritime Mobile Service Identity) est un numéro unique à 9 chiffres utilisé dans les modes de radiocommunication maritime, à la manière d'un numéro de téléphone, comme l'appel sélectif numérique (ASN), ou dans le traçage du système AIS (U.S. Coast Guard Navigation Center 2020). Attention, il a été relevé que plusieurs navires pouvaient porter le même numéro MMSI (Mazzarella et al. 2013).
- **Call sign** : Le call sign d'un navire est un code alphanumérique unique. Il est utilisé, comme le MMSI, dans les radiocommunications avec pour but de faciliter l'identification des navires (Shiptracks 2019).
- **Type** : nous l'avons vu au chapitre 2.1.2.1, il existe de nombreux types de navire. Comme pour le champ Pavillon, les données se retrouvent ici sous différentes formes et terminologies, ainsi que parfois dans différentes langues. Le type du navire a cela d'intéressant qu'il peut aider à l'identification d'un navire et permettre également de déterminer la marchandise transportée, ou tout du moins son type. Néanmoins, nous le verrons dans les résultats, la granularité de l'information peut varier d'un port à l'autre.

De manière générale, les données les plus renseignées dans notre échantillon sont le nom, le pavillon, l'IMO et le type de navire, comme le montre le tableau 3. Ce seront donc les données que l'on sélectionnera pour la partie identification de notre référentiel.

Tableau 3 : Données liées à l'identification

Port	Nom	Pavillon	IMO	MMSI	Call sign	Type
Aarhus (DK)	x					x
Amsterdam (NL)	x	x				
Bordeaux (FR)	x	x				x
Copenhague (DK)	x	x	x		x	x
Dunkerque (FR)	x					
Fredericia (DK)	x					
Hambourg (DE)	x		x			x
Klaipeda (LT)	x	x	x			
Le Havre (FR)	x	x				x
Niedersachsen (DE)	x	x			x	x
Rotterdam (NL)	x	x	x	x	x	

4.1.3 Données liées au voyage

Les données liées au voyage sont générées dynamiquement et concernent le voyage en cours d'un navire spécifique.

- **Statut** : le statut correspond à l'état ou à la position actuelle du navire par rapport au port dont les données sont récupérées. Cette donnée, ayant généralement pour valeur

« Attendu », « À quai » ou « Parti », permet d'aider à l'interprétation des données temporelles d'arrivée et de départ (« ETA », « ETD ») de chaque navire.

- UCRN : le code UCRN (Unique Call Reference Number) est généré par l'autorité portuaire lorsqu'un navire indique son heure probable d'arrivée au port (« ETA ») (Masovic 2019). Dans notre échantillon, cette procédure est effectuée uniquement dans le port de Rotterdam.
- Provenance : la provenance du bateau avant d'arriver dans le port cible ou, parfois, le port cible lui-même lorsque le navire a le statut « Parti ».
- Destination : la destination du bateau après le départ du port cible ou, parfois, le port cible lui-même lorsque le navire a le statut « Attendu ».
- ETA/ATA : l'heure d'arrivée, estimée ou actuelle, dans le port cible.
- ETD/ATD : l'heure de départ, estimée ou actuelle, du port cible.
- Terminal / Quai : ces deux champs déterminent le lieu d'accostage dans le port et peuvent permettre d'aider à l'identification du type de marchandise transportée.
- Marchandise chargée / déchargée : décrit la marchandise présente à bord du navire. Ce champ n'est disponible que pour le port de Bordeaux.
- Agent : détermine l'agent en charge du navire. On remarquera que la valeur de ce champ n'est pas toujours précise et que l'on a parfois affaire à un autre acteur du transport maritime. Cependant il s'agit d'une donnée qui peut avoir son importance car elle peut potentiellement cibler une autre source d'information sur le voyage en cours.
- Armateur : détermine l'armateur en charge du navire et de sa cargaison.
- N°escale : un champ supplémentaire que l'on retrouve dans certains ports.

Tableau 4 : Données liées au voyage

Port	Statut	UCRN	Provenance / Destination	ETA/ATA / ETD/ATD	Terminal	Quai	March. chargée / déchargée	Agent	Armateur	N° escale
Aarhus (DK)	x			x	x	x				
Amsterdam (NL)	x			x	x	x		x		
Bordeaux (FR)			x	x			x	x	x	x
Copenhague (DK)	x		x	x		x		x		
Dunkerque (FR)			x	x		x		x		
Fredericia (DK)	x			x		x		x		
Hambourg (DE)	x			x		x				
Klaipeda (LT)				x		x		x		
Le Havre (FR)			x	x		x		x		x
Niedersachsen (DE)	x		x	x	x	x				
Rotterdam (NL)		x		x		x		x		

Le tableau 4 représente le récapitulatif des données liées au voyage selon les ports de notre échantillon. Les colonnes grisées représentent les données qui sont conservées dans le référentiel normalisé.

Pour les ports qui ne proposent pas l'information sur le statut du navire, il est possible de l'ajouter manuellement selon la page web d'où provient l'enregistrement. Par exemple, le port du Havre propose trois pages qui contiennent tous les enregistrements des navires : arrivées, départs et à quai, mais les enregistrements présents sur ses pages ne renseignent pas la donnée « Statut ». Il est donc possible de rajouter manuellement une valeur dans le champ « Statut » pour chaque enregistrement selon la page depuis laquelle l'enregistrement a été récupéré.

Bien que seul le port de Bordeaux propose pour ses enregistrements des champs concernant la marchandise chargée et déchargée, ces informations sont intégrées dans le référentiel normalisé étant donné leur importance.

Les données Quai et Terminal sont sujettes à ambiguïté dans certains ports, tout comme la donnée « Agent », qui fait parfois référence à d'autres acteurs du commerce maritime, comme les opérateurs (*operator*) ou les courtiers (*broker*). Dans la mesure du possible, on essaiera de clarifier ces ambiguïtés dans notre preuve de concept.

4.1.4 Données techniques liées au bateau

Enfin, les données faisant référence aux caractéristiques techniques des navires sont disponibles dans certains gros ports, notamment Amsterdam, Rotterdam, Bordeaux et Hambourg. On peut les répertorier ainsi :

- Port d'attache, Port of Registry
- Jauge brute/nette (*Gross tonnage/Nett tonnage*)
- Année de construction (*Year of build, Built*)
- Port en lourd (*Deadweight, DWT*)
- EVP (*TEU Container, Container*)
- Longueur (*Length*)
- Largeur (*Breadth, Width*)
- Tirant d'eau avant, arrière (*Draft, Draught*)
- Puissance du bateau (*Engine power*)
- Autres infos (*Speed, etc.*)

Ces données liées aux caractéristiques techniques des navires ne sont pas les plus pertinentes pour ce mandat, elles ne sont donc pas retenues dans le référentiel. N'oublions pas qu'étant donné leur nature statique, elles peuvent être récupérées ultérieurement et compléter les informations déjà récupérées si nécessaire.

4.2 Synthèse

On peut difficilement considérer une seule donnée comme étant suffisante pour identifier un navire, notamment lorsqu'il s'agit que du nom. Ainsi, pour les ports de Dunkerque et de Fredericia, il est d'ores et déjà possible d'affirmer que les données liées à l'identification seront insuffisantes. L'information capitale pour identifier un navire reste le numéro IMO, qui est

considéré comme un identifiant unique. Malheureusement, seul 4 ports sur les 11 de notre échantillon proposent cette donnée (Copenhague, Hambourg, Klaipeda et Rotterdam). Pour les 5 ports restants, l'identification se fera grâce au croisement de deux ou trois données, comme le nom, le pavillon et le type. On ne peut cependant garantir qu'il sera impossible de confondre deux navires malgré ces informations.

Concernant les données liées au voyage, on remarque que, logiquement, les données les plus importantes, c'est-à-dire celles qui concernent les arrivées et les départs des navires, sont présentes dans chacun des onze ports de notre échantillon. La provenance et la destination sont quant à elles disponibles dans cinq ports. Ce sont principalement ces informations qui permettront de recréer les trajectoires des navires.

Le référentiel normalisé est visible dans l'annexe 6. En plus des données liées à l'identification et au voyage des navires, on a rajouté deux données dites « artificielles » : d'une part, la date de récupération de l'enregistrement (*date scraping*), qui peut permettre, lorsque l'on a plusieurs enregistrements provenant d'un même port et qui concernent un même navire, de renseigner si certains enregistrements sont obsolètes. Et d'autre part, le port d'où vient l'enregistrement, ce qui permet de rajouter une donnée précieuse sur le voyage des navires.

5. Preuve de concept

Notre preuve de concept s'est déroulée sur trois jours, du 25 au 27 juin 2020, avec un *scraping* par jour sur chacune des pages web sélectionnées. Ce chapitre décrit les outils et les technologies utilisées pour cette opération, ainsi que les spécificités induites par chacun des onze ports sélectionnés dans notre échantillon.

5.1 Web scraping & data mining

Le *web scraping* se définit comme l'action ou le processus qui permet de collecter et d'extraire automatiquement, via des scripts ou des programmes, des données en provenance du web, et notamment de pages HTML (Malik et Rizvi 2011 ; Vargiu et Urru 2012). Généralement, cela implique également de formater les données réunies dans un fichier, de format CSV par exemple.

Souvent affilié au *data mining*, le *web scraping* se différencie car il n'implique pas d'analyser les données. Réciproquement, le *data mining* n'implique pas l'extraction de données mais se concentre uniquement sur les processus complexes de l'analyse de grands jeux de données, pouvant impliquer algorithmes et *machine learning*. On peut donc considérer que les techniques de collecte de données peuvent être utilisées pour créer les jeux de données qui seront plus tard utilisés dans du *data mining* (Parsehub 2020).

On distinguera également le *web scraping* du *web crawling* : le premier consiste à extraire automatiquement des données spécifiques présentes sur une page web. Le second consiste à naviguer sur une URL et à indexer tout ce qu'il s'y trouve dans un but de recherche d'information (Patel 2020).

Dans notre cas, le *web scraping* consistera à récupérer, sur chaque page web des onze ports échantillonnés, les données définies dans notre référentiel normalisé, et ce pour chaque enregistrement de navire disponible.

5.2 Méthodologie

5.2.1 À la recherche de l'outil idéal

Il existe d'innombrables outils et solutions permettant d'effectuer du *web scraping*. On peut citer Scraper API²², ScrapeSimple²³, ou encore Octoparse²⁴. Ces solutions « clés en main » sont toutefois payantes et n'autorisent pas une grande liberté d'action. Il était donc plus intéressant et plus adapté aux besoins de ce mandat de trouver une solution gratuite, accessible, personnalisable, et de préférence open source.

Le choix s'est dans un premier temps dirigé sur le langage de programmation Python, considéré actuellement par de nombreuses sources comme le langage de référence lorsqu'il s'agit de *web scraping*, et ce pour plusieurs raisons (Chapman 2019 ; Koshy 2020 ; X-Byte Enterprise Crawling 2019) :

- Python est un langage simple et naturel utilisé par une large communauté.

²² <https://www.scraperaapi.com/>

²³ <https://www.scrapesimple.com/>

²⁴ <https://www.octoparse.com/>

- Python propose de nombreux *frameworks* et *libraries* dédiés au *web scraping*.
- Python fonctionne sur tous les principaux systèmes d'exploitation.

Le langage Python propose plusieurs solutions spécifiques pour le *web scraping*, on retiendra notamment les outils suivants :

- **Scrapy**, un *framework* Python collaboratif et open source, dont le but est de permettre l'extraction rapide et simple de données depuis des sites web (Scrapy 2020).
- **BeautifulSoup**, une *library* Python qui permet d'analyser des documents HTML et XML grâce à des arbres syntaxiques, *parse trees*, qui facilitent l'extraction de données depuis le HTML (Mehta et al. 2020).
- **Selenium**, un *framework* Python qui permet d'interagir et de naviguer sur des pages web comme le ferait un véritable utilisateur (Kite 2020 ; Tech with Tim 2020).

Chacune de ces solutions possède ses avantages et ses inconvénients (Kite 2020). Pour les besoins de ce travail, le choix s'est porté sur le *framework* Scrapy pour les raisons suivantes :

- Scrapy est simultanément un *web scraper* et un *web crawler*, il permet à la fois d'analyser des pages web et de récupérer des données (Myers et McGuffee 2015).
- Scrapy, en tant qu'outil *Open Source*, possède une large communauté sur Github²⁵, Twitter²⁶ et sur StackOverflow²⁷, ce qui facilite grandement la résolution de problèmes lors de la rédaction de scripts.
- Scrapy, grâce à l'outil *Splash*, peut récupérer des données sur des sites web fonctionnant avec du JavaScript (Mehta et al. 2020).
- Scrapy propose également de nombreux services payants sur *scrapinghub*²⁸, la plateforme professionnelle de Scrapy, avec notamment un cloud pour y déposer et gérer ses scripts, le *scrapy cloud*²⁹.

5.2.2 Le *framework* Scrapy

Contrairement à BeautifulSoup qui est une *library* Python, c'est-à-dire une collection de fonctions et de méthodes qui permettent d'effectuer de nombreuses actions sans que l'on écrive soi-même de sempiternelles lignes de code ; Scrapy est un *application framework*, c'est-à-dire un « squelette » de programme qui permet l'écriture de *spiders*, à savoir des scripts qui sont constitués de classes Python qui vont être exécutées pour parcourir différentes pages web afin d'y extraire des données (Scrapy 2020b).

Pour chacun des onze ports de notre échantillon, un projet Scrapy a été créé. Ce dernier comprend les fichiers nécessaires au *web scraping* avec un répertoire contenant un ou plusieurs scripts, nommés *spiders*, selon le nombre de pages web qui doivent être *scrapées*. La liste des scripts créés est disponible dans le tableau 5 du chapitre 5.3.1.

²⁵ <https://github.com/scrapy/scrapy>

²⁶ <https://twitter.com/ScrapyProject>

²⁷ <https://stackoverflow.com/tags/scrapy/info>

²⁸ <https://www.scrapinghub.com/>

²⁹ <https://www.scrapinghub.com/scrapy-cloud/>

5.2.2.1 Outils supplémentaires

D'autres outils alliés à Scrapy ont été utilisés dans ce travail et méritent d'être mentionnés :

- Scrapy dispose de sa propre interface en ligne de commande, baptisée **Scrapy shell**³⁰, qui permet de tester et de debugger son code très rapidement sans devoir lancer son script.
- Pour pouvoir analyser et récupérer les données provenant de pages web réalisées en Javascript, on a utilisé l'outil **Splash**³¹. Cet outil nécessite la modification des paramètres du projet en cours (Scrapinghub 2017) ainsi que l'installation d'un logiciel tiers, *Docker*³².
- L'étude de la page web s'est faite avec l'inspecteur de code du navigateur Chrome. Des extensions comme **SelectorGadget**³³ sont particulièrement utiles pour analyser les chemins XPATH ou les sélecteurs CSS des données à récupérer.
- **Dateparser**³⁴ est une *library* Open Source pour analyser les champs de type date et les uniformiser selon le format désiré (Dowinton 2015). On a pris le parti d'utiliser cette fonction dans le code de nos scripts pour uniformiser les champs dates avant l'exploitation des résultats dans un logiciel de curation de données. Malheureusement, une fois la phase de *scraping* terminée, des erreurs ont été constatées dans les champs « ETA » et « ETD » de certains enregistrements récupérés. En effet, *Dateparser* intervertissait parfois le mois et le jour d'une même date, obtenant par exemple des dates affichant le 07/01/2020 au lieu du 01/07/2020. Ces erreurs ont été corrigées et des avertissements ont été rajoutés en commentaire dans le code des scripts concernés.
- L'utilisation d'**expressions régulières**³⁵ a parfois été nécessaire pour extraire uniquement les données pertinentes dans certaines balises HTML.

5.2.2.2 Versions utilisées

Pour notre preuve de concept, la distribution libre et open source Anaconda (v2019.10)³⁶ a été utilisée pour son interface en ligne de commande Anaconda Prompt. Les versions des éléments suivants ont été utilisées :

- Dateparser 0.7.5
- Python 3.7.4
- Scrapy 1.8.0
- Scrapy-splash 0.7.2
- OpenRefine 3.3

5.2.3 Considérations juridiques

Il n'est pas étonnant à notre époque, où les données sont de plus en plus considérées comme le pétrole de l'ère digitale (Bhageshpur 2019), que la question légale à propos du *web scraping*

³⁰ <https://docs.scrapy.org/en/latest/topics/shell.html>

³¹ <https://splash.readthedocs.io/en/stable/>

³² <https://www.docker.com/>

³³ <https://selectorgadget.com/>

³⁴ <https://dateparser.readthedocs.io/>

³⁵ <https://docs.python.org/fr/3/library/re.html>

³⁶ <https://docs.anaconda.com/anaconda/>

puisse se poser. Il existe en effet plusieurs points de vue contradictoires sur la légalité ou l'illégalité des pratiques de *web crawling* et de *web scraping*.

Par exemple, pour de nombreux sites web commerciaux qui proposent des comparaisons de prestations entre concurrents, la pratique du *web scraping* est considérée comme une manière rapide, simple et légale de récupérer des données (Mehta et al. 2020).

D'autres acteurs du web sont néanmoins beaucoup plus sensibles sur la récupération de ce qu'ils considèrent comme leur propriété. Citons à titre d'exemple le procès qui a opposé le réseau social *LinkedIn* à la société d'analyse de données *hiQ Labs*. Une affaire où *LinkedIn* a demandé à *hiQ Labs* de cesser sa pratique de récolte de données provenant de ses profils publics. *HiQ Labs*, dont la source de données principale était le réseau social, n'a eu d'autre choix que d'attaquer la société détenue par Microsoft en justice, et a eu gain de cause, les données de *LinkedIn* étant considérées comme disponibles publiquement.

Le 9 septembre 2019, la justice américaine a par ailleurs statué que le *web scraping* de données publiques sur le web ne violait pas le *Computer Fraud and Abuse Act* (CFAA), une loi du gouvernement fédéral américain mise en vigueur en 1986 et qui concerne la sécurité des systèmes d'information (Computer Fraud and Abuse Act 2020). Par une telle décision, le tribunal a d'une certaine manière légalisé la pratique de *web scraping* de données publiques, considérant le *web scraper* à la manière d'un internaute depuis son navigateur internet (Parsers 2020).

Sous nos latitudes et pour ce qui nous concerne, les données ne représentant pas un intérêt économique particulier, nous nous contenterons de vérifier les instructions du fichier *robots.txt* présent à la racine de chaque site web. Ce dernier indique les règles que doivent suivre les robots d'indexation (*crawlers*), notamment les éventuelles sections du site web qui ne doivent pas être indexées et par conséquent, *scrapées* (Mehta et al. 2020).

5.2.4 Différents scripts pour différents ports

Cette section passe en revue les différents sites web des ports de notre échantillon et présente les pages qui seront *scrapées* ainsi que les principales spécificités des futurs *spiders*.

Aarhus : La page web qui nous intéresse sur le site du port d'Aarhus est la page « Vessels in port », dans la section « Ship calls »³⁷. Les navires à quai y sont classés par terminaux. Si la dernière date et heure de mise à jour est affichée, la source des données n'est quant à elle pas précisée. Le port d'Aarhus a nécessité la création d'un seul script.

Amsterdam : Les arrivées et départs du port d'Amsterdam sont toutes disponibles sur la page web *Arrivals & Departures*³⁸. Une page générée avec du Javascript, ce qui a impliqué l'utilisation de *Splash*. Le code du script a nécessité l'utilisation d'expressions régulières pour nettoyer le champ « Statut ». De plus, des structures conditionnelles ont été nécessaires suivant le statut du navire pour déterminer correctement les données liées aux heures d'arrivées et de départs. Précisons encore que ni la dernière date de mise à jour, ni la source des données ne sont indiquées.

³⁷ https://www.aarushavn.dk/en/ship_calls/vessels_in_port/index.htm

³⁸ <https://myport.portofamsterdam.com/arrivals/>

Bordeaux : Le site web du port de Bordeaux propose également une page avec des données dynamiques affichant les navires à quai et attendus³⁹. En plus de l'utilisation de *Splash*, la particularité du *scraping* a été qu'il n'a pas été possible de déterminer le statut spécifique des navires. Les données sont indiquées comme étant transmises « en temps réel par la Capitainerie du port de Bordeaux ». La date de dernière mise à jour est également fournie.

Copenhague : Comme vu dans le chapitre 3.1.1.2, le site web du port de Copenhague est affilié au port de Malmö. Cependant, il est possible d'avoir accès aux données spécifique à un seul port, et même d'obtenir les appels de navire concernant un type en particulier, en l'occurrence les pétroliers⁴⁰. Un script a donc été codé pour récupérer les données de cette liste en particulier. La source des données ainsi que la date de dernière mise à jour ne sont pas indiquées.

Dunkerque : Le site du port de Dunkerque propose trois pages contenant des données relatives à ses activités commerciales. Les navires présents⁴¹, les navires attendus en entrée/transit⁴² et les navires attendus en sortie⁴³. Trois scripts ont donc été réalisés. La date et l'heure de mise à jour sont indiquées sur chacune des pages. La source des données n'est quant à elle pas mentionnée.

Fredericia : Le site web de la compagnie d'ADP A/S gère à la fois le port de Fredericia et de Nyborg au Danemark. Une page propose spécifiquement la liste des arrivées dans le port de Fredericia⁴⁴, ce qui a permis la création d'une seul script. L'origine des données n'est pas indiquée contrairement à la date et heure de dernière mise à jour.

Hambourg : Le site web du port de Hambourg propose plusieurs fonctionnalités intéressantes, notamment la possibilité de rechercher des navires par catégorie ou par comparaison de caractéristiques. Nous nous sommes cependant intéressés à deux autres pages : les navires attendus lors des 2 à 3 prochains jours⁴⁵ (*expected vessels*) et les mouvements de navires⁴⁶ (*ship movements*).

Ces deux pages utilisent une technique web qui a amené une problématique particulière pour notre opération de collecte de données : *l'infinite scrolling*. Il s'agit d'un procédé permettant de charger le contenu d'une page web au fur et à mesure que l'internaute fait défiler la page vers le bas, évitant ainsi le besoin de créer de multiples pages (Loranger 2014). Ces pages comprennent une limite de 20 enregistrements. À chaque fois que l'on défile jusqu'en bas de la page, 20 nouveaux enregistrements sont chargés, et ainsi de suite. Contrairement à la page des navires attendus, qui comprend un nombre relativement faible de navires en prévision, il est possible de faire défiler la page concernant les mouvements de navire quasiment indéfiniment et de récupérer ainsi la totalité de la base de données du site de Hambourg. Au 22 juillet 2020, il était donc possible de récupérer 58204 enregistrements de navires, le plus ancien datant du 19 juin 2017.

³⁹ <https://www.bordeaux-port.fr/fr/navires-quais-attendus>

⁴⁰ <http://www.cmport.com/ships-in-port/oil/copenhagen>

⁴¹ <http://www.dunkerque-port.fr/fr/activites-commerciales/navires-presents-port-dunkerque.html>

⁴² <http://www.dunkerque-port.fr/fr/activites-commerciales/navires-entree-transit.html>

⁴³ <http://www.dunkerque-port.fr/fr/activites-commerciales/navires-sortie.html>

⁴⁴ <https://www.adp-as.dk/en/maritime-services/calls/>

⁴⁵ <https://www.hafen-hamburg.de/en/vessels/eta>

⁴⁶ <https://www.hafen-hamburg.de/en/vessels/ship-movements>

Il a été finalement décidé de limiter dans nos scripts le nombre de défilement de pages à 20, ce qui permet à chaque opération de *scraping* de récolter un maximum de 420 enregistrements.

Précisons encore que pour le site web de Hambourg, il n'y a d'indications ni pour la date de mise à jour, ni pour la source des données.

Klaipeda : Le site web du port de Klaipeda propose également plusieurs listes intéressantes avec notamment : *list of vessels within the port*⁴⁷, *planned calls*⁴⁸, *last calls*⁴⁹ et *last departures*⁵⁰. Ces 4 pages seront *scrapées* par autant de scripts différents.

On remarque que Klaipeda pose la problématique de la linguistique puisque certaines données sont en lituanien. À noter encore que le site n'indique pas la source des données mises à disposition.

Le Havre : Le site web du port du Havre propose un « guichet électronique de gestion des escales »⁵¹ (Haropa 2019). Ce dernier offre des informations sur les mouvements de navires avec trois onglets : arrivées, départs et navires à quai. Il a été décidé de récupérer les données provenant uniquement des deux premiers onglets, les navires à quai étant d'une part essentiellement des navires spéciaux (pousseurs, automoteurs, drague, etc.) et d'autre part, il y avait un fort risque de redondance avec les deux premiers onglets. Deux scripts ont donc été créés.

Niedersachsen : Le site web des ports de Basse-Saxe dispose d'une seule page proposant la liste des arrivées et départs des navires dans les différents ports gérés par l'entreprise⁵², ce qui a permis la création d'un unique script pour la collecte de données. Si la source des données n'est pas indiquée, on trouve par contre la date et l'heure de la dernière mise à jour. À noter que cette page est mise à jour toutes les quatre heures.

Rotterdam : Le site du port de Rotterdam permet un accès direct au fichier JSON⁵³ comprenant tous les enregistrements affichés sur la page *Arrivals and departures of vessels* du site web⁵⁴. Nous avons ainsi créé un script pour chaque membre JSON : *Present*, *Departed* et *Expected*. L'heure et date de dernière mise à jour et la source des données ne sont pas indiquées.

5.3 Synthèse

Chacun des scripts a été codé de manière à récupérer les données mises en exergue dans le référentiel normalisé du chapitre 4.

Quelques éléments ont été généralisés pour chacun des scripts afin de procéder à une première forme d'uniformisation des données :

⁴⁷ <https://www.portofklaipeda.lt/list-of-vessels-within-the-port>

⁴⁸ <https://www.portofklaipeda.lt/planned-calls>

⁴⁹ <https://www.portofklaipeda.lt/last-calls>

⁵⁰ <https://www.portofklaipeda.lt/last-departures>

⁵¹ <https://www.havre-port.com/>

⁵² <https://www.nports.de/en/ship-arrivals-and-departures/>

⁵³ https://www.portofrotterdam.com/ship_visits/all

⁵⁴ <https://www.portofrotterdam.com/en/shipping/operational-information/nautical-information/arrivals-and-departures-of-vessels>

- Les données du champ « Statut » ont été traduites, uniformisées et augmentées, lorsqu'elles n'étaient pas présentes, pour avoir autant que possible les choix « À quai », « Attendu » ou « Parti ».
- La date a été uniformisée au format YYYY-MM-DD HH:SS grâce à la bibliothèque Python *dateparser*. Et comme nous l'avons mentionné dans le chapitre 5.2.2.1, des erreurs ont été constatées.

5.3.1 Listes des scripts par ports

Le tableau suivant résume les différents scripts créés pour chacun des ports en précisant le nom de la page web *scrapée*. Le code de chacun des scripts est disponible dans l'annexe 7 du présent mémoire.

Tableau 5 : Liste des différents scripts réalisés pour chaque port

Port	Script(s) (pages web scrapées)
Aarhus (DK)	- Vessels in port
Amsterdam (NL)	- Arrivals and departures
Bordeaux (FR)	- Navires attendus
Copenhague (DK)	- Port calls oil
Dunkerque (FR)	- Navires attendus - Navires partants - Navires présents
Fredericia (DK)	- Port calls
Hambourg (DE)	- Ships expected - Ships movements
Klaipeda (LT)	- Last calls - Last departures - Planned calls - Vessels within port
Le Havre (FR)	- Arrivées - Départs
Niedersachsen (DE)	- Arrivals and departures
Rotterdam (NL)	- Departed ships - Expected ships - Present ships

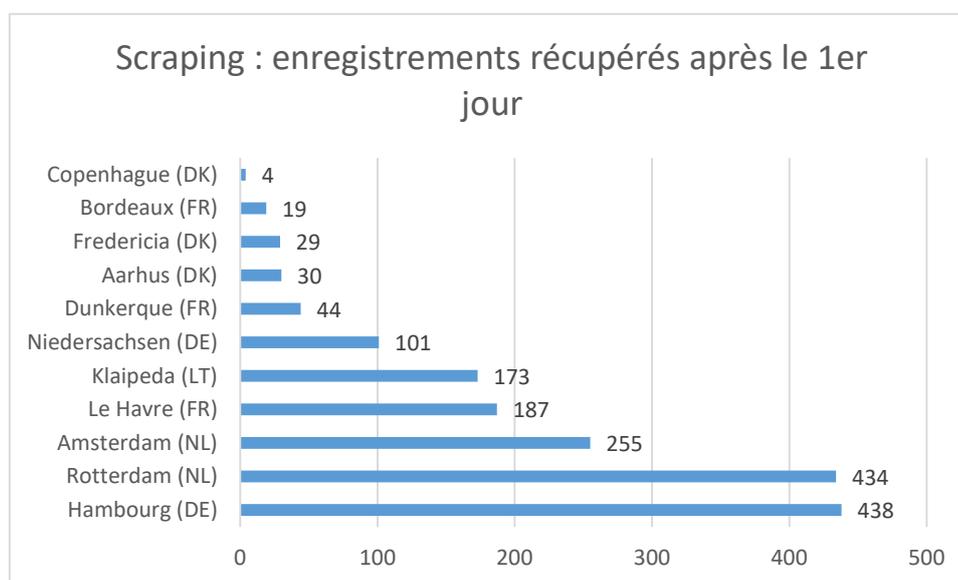
6. Évaluation des résultats

La preuve de concept s'est déroulée sur trois jours, du 25 au 27 juin 2020, avec le lancement de tous les scripts de *web scraping* une fois par jour. Pour chacun des scripts, les enregistrements retournés ont été compilés dans un fichier au format .csv.

6.1 État des données récupérées

Après le premier jour de *scraping*, il a été possible de réaliser un bilan quantitatif des enregistrements récupérés. Il convient de préciser que chaque enregistrement comprend les données relatives à un navire, données déterminées par le référentiel normalisé décrit au chapitre 4.

Figure 5 : Quantité d'enregistrements récupérés par port le 25 juin 2020

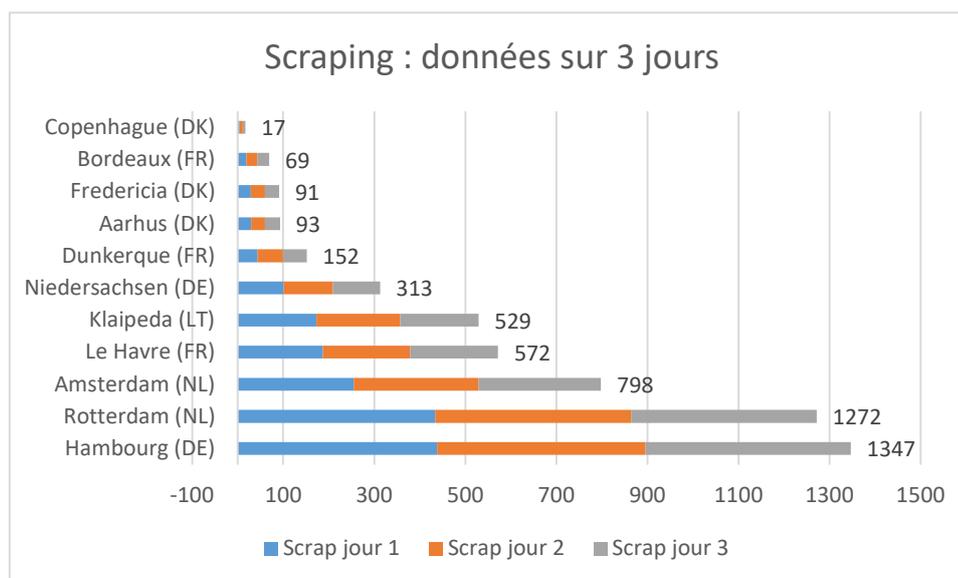


Les deux extrêmes affichés dans la figure 5 sont en l'occurrence des cas particuliers : comme nous l'avons vu dans le chapitre 5.2.4, il est possible de *scaper* la totalité de la base de données des mouvements de navires présente sur le site web du port de Hambourg, avec une limite fixée dans ce travail à 420 enregistrements par jour de *scraping*. Concernant le port de Copenhague, rappelons que la collecte de données se fait sur la page répertoriant uniquement les pétroliers, il est donc normal de trouver un nombre très restreint d'enregistrements récupérés. On obtient donc 1714 enregistrements pour ce premier jour avec une quantité d'enregistrements récupérés proportionnelle à la « taille » du port évoquée dans le tableau 2 du chapitre 3.1.1.2. Attention, il ne s'agit pas pour autant de 1714 navires différents, beaucoup d'entre eux étant présents dans les différentes pages web *scrapées* d'un même port.

6.1.1 Scraping sur trois jours

Rien de particulier à signaler sur les deux jours suivants, la récupération des enregistrements du 26 et 27 juin est quantitativement très proche de celle effectuée le 25 juin comme démontré dans la figure 6. Concernant ces données brutes, l'ordre quantitatif entre les différents ports et également respecté. Les trois plus gros port, Hambourg, Rotterdam et Amsterdam, sont, comme attendu, les plus prolifiques en termes de données. On peut également remarquer que les ports danois de notre échantillon offrent sensiblement moins d'enregistrements que les ports français.

Figure 6 : Quantité de données *scrapées* du 25 au 27 juin 2020



On a donc récupéré un total de 5253 enregistrements de navires, ce qui correspond à une moyenne de 1751 enregistrements récupérés par jour. Par rapport à cette moyenne, les écarts quotidiens sont très faibles, la récupération des données s'est donc faite de manière régulière durant les trois jours de la preuve de concept :

Tableau 6 : Enregistrements récupérés, totaux et moyenne

	Jour 1	Jour 2	Jour 3	Moyenne
Totaux	1714	1794	1745	1751
Écart moyenne	-2.1%	+2.5%	-0.3%	

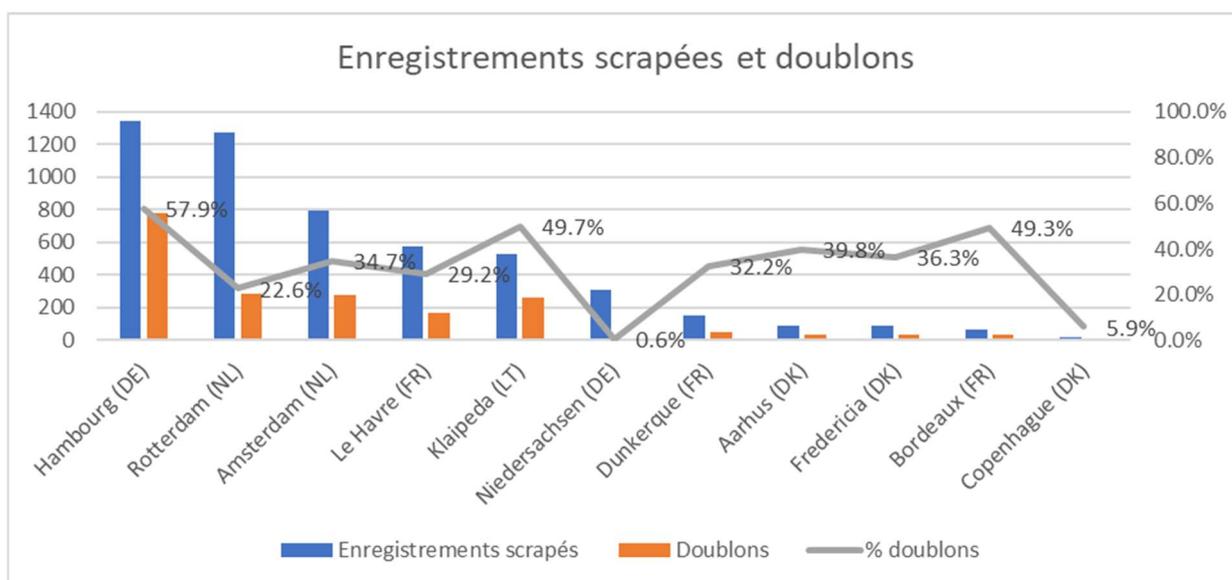
6.2 Nettoyage des données

Une fois les enregistrements bruts récupérés, il a fallu les nettoyer pour permettre l'analyse de nos résultats. L'élimination des doublons a d'abord permis de rendre compte de la véritable quantité de données récupérées puis, grâce au logiciel OpenRefine, nous avons pu uniformiser certains champs, pour que les valeurs récupérées soient le plus possible sous la même forme et dans la même langue.

6.2.1 Élimination des doublons

L'opération d'élimination des doublons a permis la suppression des enregistrements récupérés lors des jours 2 et 3 qui étaient déjà présents lors du *scraping* du jour 1.

Figure 7 : Totaux des enregistrements récupérés et quantité de doublons



La figure 7 représente la quantité d'enregistrements véritablement utilisable au terme des 3 jours de *scraping*, après la suppression des doublons. Ainsi on remarque qu'il a fallu supprimer, dans la plupart des cas, entre 20% et 50% des données brutes récupérées. On peut fournir les explications suivantes pour les cas extrêmes :

Nous l'avons vu, avec le port d'Hambourg il a été possible de récupérer chaque jour un nombre important et défini d'enregistrements dans la base de données. Ainsi il est normal que de nombreux enregistrements se retrouvent à double au bout des trois jours qu'ont duré notre preuve de concept. Concernant les ports de Niedersachsen et de Copenhague, qui ont très peu de doublons, on peut fournir deux hypothèses qui devront être vérifiées dans l'analyse des résultats. D'une part, il est possible que les enregistrements soient mis à jour régulièrement, ce qui veut dire que l'on risque de se retrouver avec plusieurs enregistrements pour le même navire, mais avec des données temporelles (champs « ETA » et « ETD ») différentes. D'autre part, il est aussi possible que le laps de temps entre chaque opération de *scraping* soit trop long, et que par conséquent les enregistrements récupérés chaque jour concernent uniquement de nouveaux navires. Ainsi, il y aurait un risque de rater certaines informations, notamment les mises à jour les plus récentes sur le voyage de certains navires.

6.2.2 Uniformisation et nettoyage des champs

Pour rappel, certains champs ont déjà été uniformisés dans les scripts :

- Le champ « **Statut** », que l'on a traduit directement dans le script afin d'avoir autant que possible les valeurs uniformes *Attendu*, *À quai* et *Partant*.
- Les champs impliquant des dates « **ETA** », « **ETD** », afin d'avoir toujours le même format *YYYY-MM-DD HH:MM*.

Toutefois, l'uniformisation des champs impliquant des abréviations, des langues étrangères ou simplement une typologie libre à chaque port nécessite un outil de préparation de données capable d'effectuer des micro-transformations.

Ainsi, le logiciel libre de nettoyage et de mise en forme de données OpenRefine⁵⁵ a été retenu. Initialement développé Metaweb, puis par Google sous le nom de Google Refine, OpenRefine est un logiciel open source et gratuit (Saby 2018). Bien que limité et plutôt lent dans ses performances sur de gros jeux de données, il convient parfaitement dans le cadre de notre étude.

L'uniformisation des champs est rendue possible grâce au service de réconciliation sur OpenRefine. Ce service propose d'uniformiser chaque valeur à l'aide d'une liste de candidats issue d'une source de données normalisées, comme la base de données *wikidata*⁵⁶. Cela va nous permettre notamment de regrouper des valeurs similaires, mais proposées originellement dans des langues ou des typologies différentes. Pour les ports qui possèdent les informations adéquates dans leurs enregistrements, nous avons procédé aux réconciliations suivantes :

- Pavillon -> entité *country*⁵⁷ de wikidata
- Type de bateau -> entité *ship type*⁵⁸ de wikidata
- Provenance et destination -> entité *port*⁵⁹ ou *city*⁶⁰ de wikidata

Attention cependant, une réconciliation sera effectuée uniquement s'il est possible de garder le même niveau de précision que pour la valeur originale. Ainsi, dans des cas où il y a un risque de perte d'information, dû à des homonymes par exemple, la réconciliation n'est pas appliquée et la valeur du champ est laissée telle quelle.

6.2.3 Dépôt des données

Les données brutes, les données nettoyées, c'est-à-dire uniformisées et réconciliées, ainsi que les projets Scrapy dédiés à chaque port peuvent être consultés sur le dépôt de données Zenodo à l'adresse suivante : <https://doi.org/10.5281/zenodo.3980515>.

6.3 Résultats par ports

Pour chaque port, nous récapitulons la quantité d'enregistrements récupérés au total et pour chaque page web *scrapée* durant les trois jours de notre preuve de concept, en observant également la quantité de navires uniques comptabilisée dans ces enregistrements. Dans un second temps, nous analysons le taux de réponse, c'est-à-dire la quantité et la proportion de données récupérées dans les champs proposés par le référentiel de chaque port.

Précisons également que nous ne nous préoccupons pas de la qualité des données, c'est-à-dire que nous ne déterminerons pas dans ce travail le niveau d'exactitude des informations récupérées sur les sites des ports échantillonnés. Cependant, dans l'étape d'uniformisation et de réconciliation des données brutes, nous avons tout de même pris le parti d'effacer chaque champ rempli avec une mention inconnue⁶¹, afin de les considérer comme s'ils n'étaient simplement pas renseignés. Cette opération est toutefois mentionnée dans les résultats lorsqu'elle a eu lieu.

⁵⁵ <https://openrefine.org/>

⁵⁶ https://www.wikidata.org/wiki/Wikidata:Main_Page

⁵⁷ <https://www.wikidata.org/wiki/Q6256>

⁵⁸ <https://www.wikidata.org/wiki/Q2235308>

⁵⁹ <https://www.wikidata.org/wiki/Q44782>

⁶⁰ <https://www.wikidata.org/wiki/Q515>

⁶¹ C'est-à-dire pour des valeurs comme « Unknown », « Inconnu », etc.

Pour plus d'intérêt de lecture et de clarté, nous proposons ces résultats selon la taille du port, comme proposé dans le tableau 2 au chapitre 3.1.1.2.

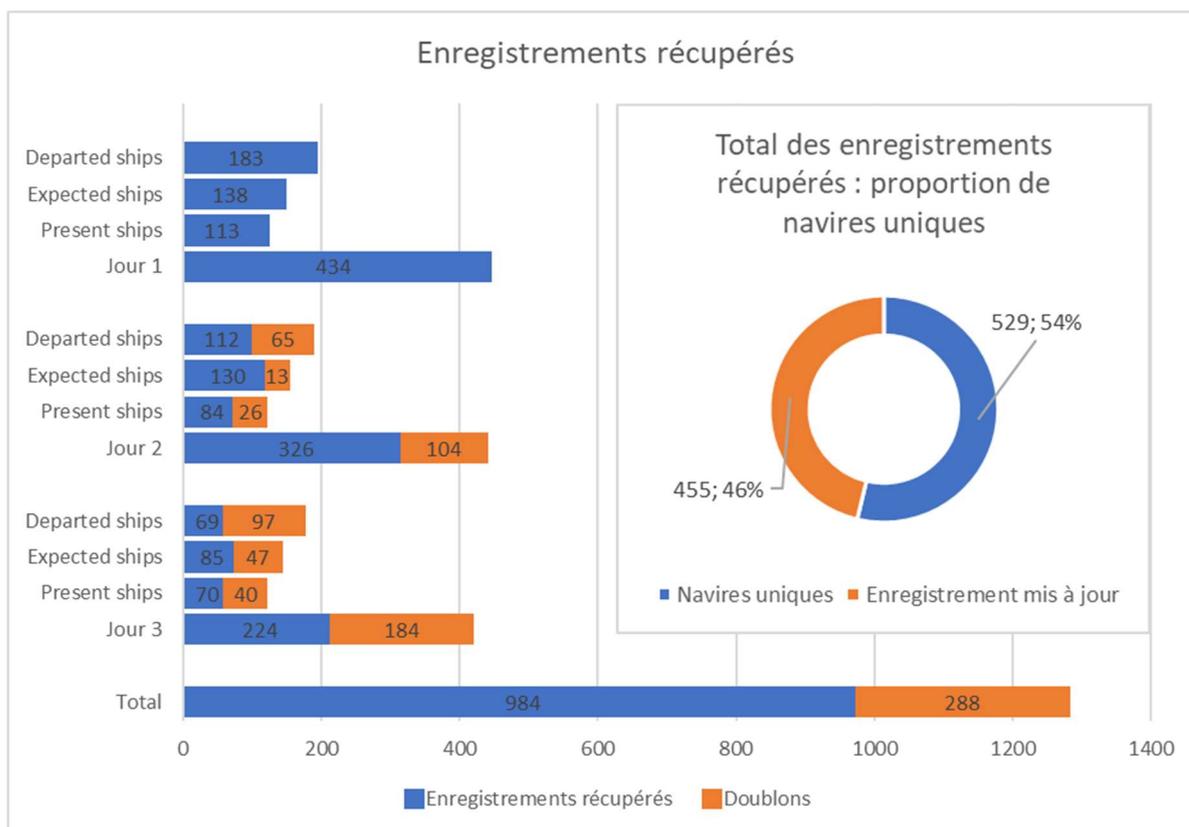
6.3.1 Rotterdam

Port européen le plus important, Rotterdam proposait trois catégories d'enregistrements : *Departed ships* (navires partis), *Expected ships* (navires attendus), et *Present ships* (navires à quai). C'est en suivant ces catégories que le champ « Statut » des enregistrements a été rempli.

Avec une moyenne de 424 enregistrements de navires récupérés par jour, le port de Rotterdam remplit bien son statut de port le plus important. Si l'on excepte le cas particulier du port d'Hambourg, il est largement devant le prochain port de notre échantillon, Amsterdam, qui lui propose 266 enregistrements par jour en moyenne.

Au terme du 3^{ème} jour de *scraping*, le port de Rotterdam proposait encore près de 55% d'enregistrements non doublonnés, nous aurions donc pu encore obtenir de nombreuses données si la preuve de concept avait duré quelques jours de plus.

Figure 8 : Port de Rotterdam, détail par jour des enregistrements récupérés



Sur les 984 enregistrements récupérés, 54% représentent des navires uniques. La proportion des enregistrements de navires mis à jour est relativement élevée étant donné qu'un navire peut être présent au fil des jours dans les trois pages *scrapées*.

Les données brutes ont subi ensuite plusieurs transformations sous OpenRefine :

- Les valeurs du champ « Nom », entièrement en majuscules, ont été modifiées selon la transformation *To titlecase*.

- Les valeurs du champ « Pavillon », composées d'abréviations de 3 lettres, ont été réconciliées pour obtenir des noms de pays complet.
- Les champs renseignés avec les valeurs « Unknown » ont été effacés : cela correspond à 73 enregistrements pour le champ « ETD » et 3 enregistrements pour le champ « Pavillon ».

On obtient un taux de réponse moyen de 86,5%. Comme le montre le tableau 7, les champs du référentiel sont bien remplis : concernant les données d'identification du navire, on obtient presque un sans-faute, il n'y a qu'une valeur de pavillon inconnue. Les données liées au voyage sont quant à elles plus éparées, notamment en ce qui concerne l'heure d'arrivée prévue. Les navires ayant le statut « À quai » (*present ships*) et « Parti » (*departed ships*) ne renseignent notamment pas ce champ.

Tableau 7 : Port de Rotterdam, taux de réponse des données récupérées

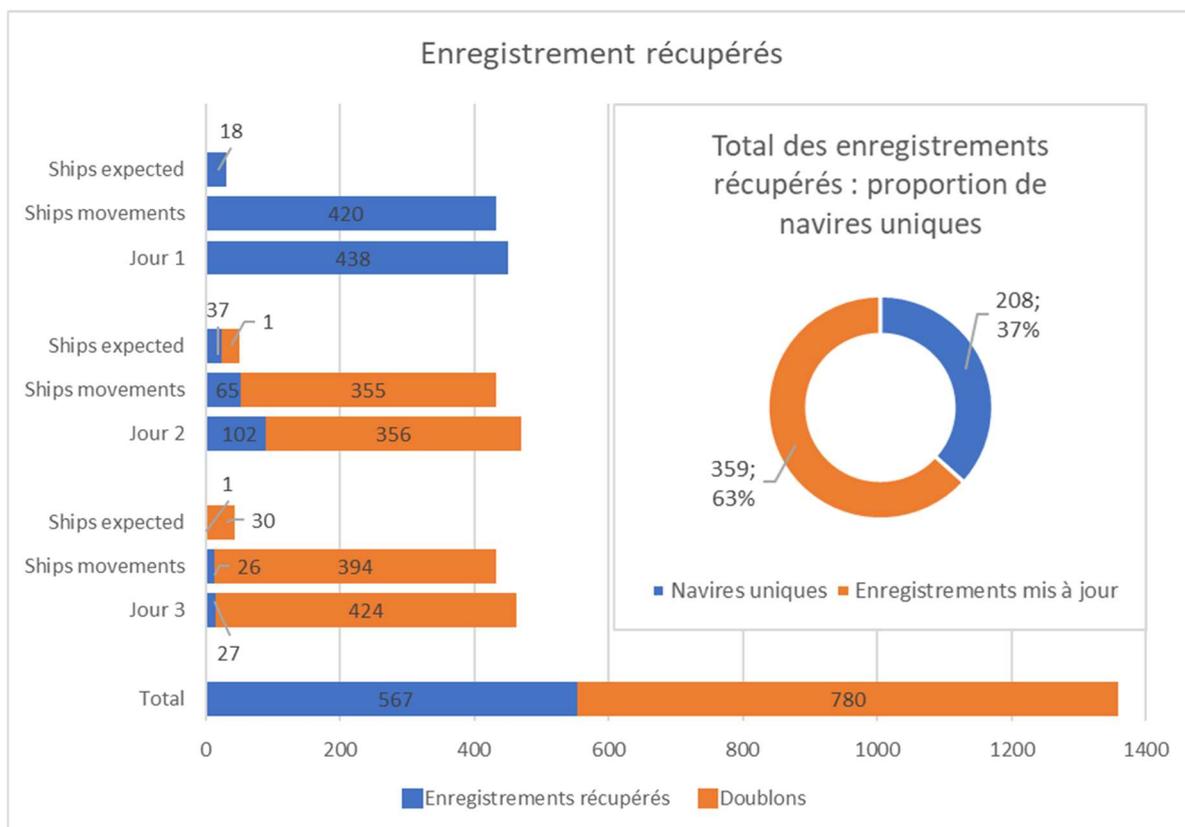
Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	Marchandise déchargée	Marchandise chargée	Agent
984	984	981	984	-	-	-	353	557	-	984	-	-	980
100.0%	100.0%	99.7%	100.0%	-	-	-	35.9%	56.6%	-	100.0%	-	-	99.6%

6.3.2 Hambourg

Second port le plus important de notre échantillon, Hambourg est, nous l'avons vu au chapitre 5.2.4, un cas particulier avec son nombre fixe de 420 enregistrements de mouvements de navire récupérés chaque jour. Par conséquent, il n'est pas étonnant de trouver un grand nombre de doublons dans nos enregistrements. Au final, le port de Hambourg nous a fourni 567 enregistrements pour 208 navires uniques, ce qui représente une des plus faibles proportions (37%) parmi les ports de notre échantillon.

Précisons que la page *ships expected* comprenait les enregistrements ayant le statut « Attendu » et la page *ships movements*, les enregistrements ayant le statut « Parti » et « À quai ».

Figure 9 : Port de Hambourg, détail par jour des enregistrements récupérés



La figure 9 nous fait remarquer que, pour un port de cette envergure, la quantité de nouveaux enregistrement récupérés après le 1^{er} jour, une fois les doublons supprimés, paraît très faible. Rappelons cependant que le port de Hambourg enregistre environ 8'000 appels de navires par année (Port of Hamburg 2020), par conséquent les 27 nouveaux enregistrements récupérés le 3^{ème} jour sont de proportion correcte. Toutefois ce qui est plus étonnant est la proportion inverse d'enregistrements récupérés et de doublons dans les navires attendus (*ships expected*) lors des jours 2 et 3.

Les données brutes ont subi les transformations suivantes sous OpenRefine :

- Les valeurs du champ « Nom » ont été modifiées selon la transformation *To titlecase*.
- Deux valeurs du champ « Quai » sous la forme '-' ont été effacées.
- Le type de navire étant en allemand, il a été réconcilié selon le tableau 8.

Tableau 8 : Port de Hambourg, réconciliation du type de navire

Valeur originale	Valeur après réconciliation
RoRo/ConRo schiff	roll-on/roll-off ship
Bagger	dredger
Containerschiff	container ship
Fahrzeugtransporter	car carrier
Feederschiff	feeder ship

Kreuzfahrtschiff	cruise ship
Kühlschiff	reefer ship
Massengutschiff	bulk carrier
Mehrzweckschiff	multi-purpose vessel
Museumsschiff	museum ship
Tankschiff	tanker ship

Avec un taux de réponse de 72,8%, les données renseignées par le port de Hambourg dans les champs de son référentiel sont plutôt médiocres. Alors que les données d'identification sont complètes, les données concernant le voyage sont quant à elles, très limitées. On obtient ainsi moins de 50% de données disponibles pour le champ « ETD ». La date d'arrivée « ETA » reste logiquement indisponible pour les navires ayant le statut « Parti ». Le quai est renseigné pour la quasi-totalité des navires attendus (à 96,4%), mais reste vide pour les navires déjà présents dans le port (statut « À quai »).

Tableau 9 : Port de Hambourg, taux de réponse des données récupérées

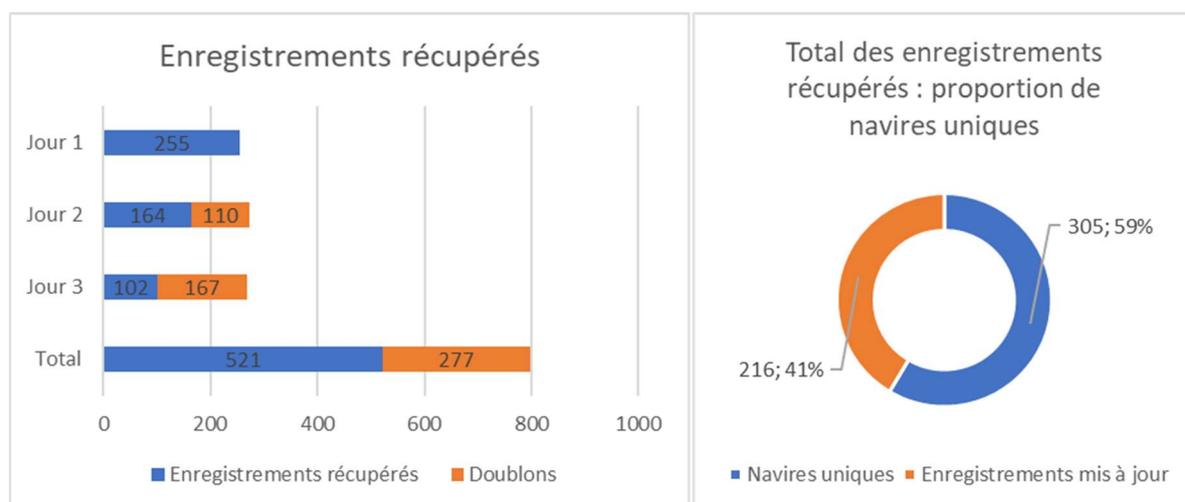
Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	Marchandise déchargée	Marchandise chargée	Agent
567	567	-	567	567	-	-	309	258	-	54	-	-	-
100.0%	100.0%	-	100.0%	100.0%	-	-	54.5%	45.5%	-	9.5%	-	-	-

6.3.3 Amsterdam

3^{ème} port le plus important de notre échantillon, le port d'Amsterdam nous a fourni 521 enregistrements comprenant 305 entrées de navires uniques. Il s'agit du premier port proposant une seule page web pour tous ses enregistrements. On y trouve les statuts suivants : « À quai », « Partant », « Attendu » et « En déplacement ».

Par rapport au port de Hambourg qui est d'importance similaire (voir tableau 2), le nombre d'enregistrements récupérés est dans le même ordre de grandeur. On remarque néanmoins qu'au bout du 3^{ème} jour de *scraping*, on récupère encore beaucoup d'information à Amsterdam, 102 enregistrements contre 27 pour Hambourg.

Figure 10 : Port d'Amsterdam, détail des enregistrements récupérés



Les données brutes ont subi les transformations suivantes sous OpenRefine :

- Les valeurs du champ « Nom », entièrement en majuscules, ont été modifiées selon la transformation *To titlecase*.
- Les valeurs du champ « Pavillon », composées d'abréviations de 3 lettres, ont été réconciliées pour obtenir des noms de pays complet.

Le taux de réponse affiche une moyenne de 81.9%, ce qui implique que de nombreux champs ne sont pas toujours renseignés. On peut expliquer une partie de ces absences par le fait que les navires ayant le statut « Partant » n'ont pas d'ETA. Plus étonnant par contre, aucun navire dont le statut est « À quai » n'a son champ « ETD » rempli.

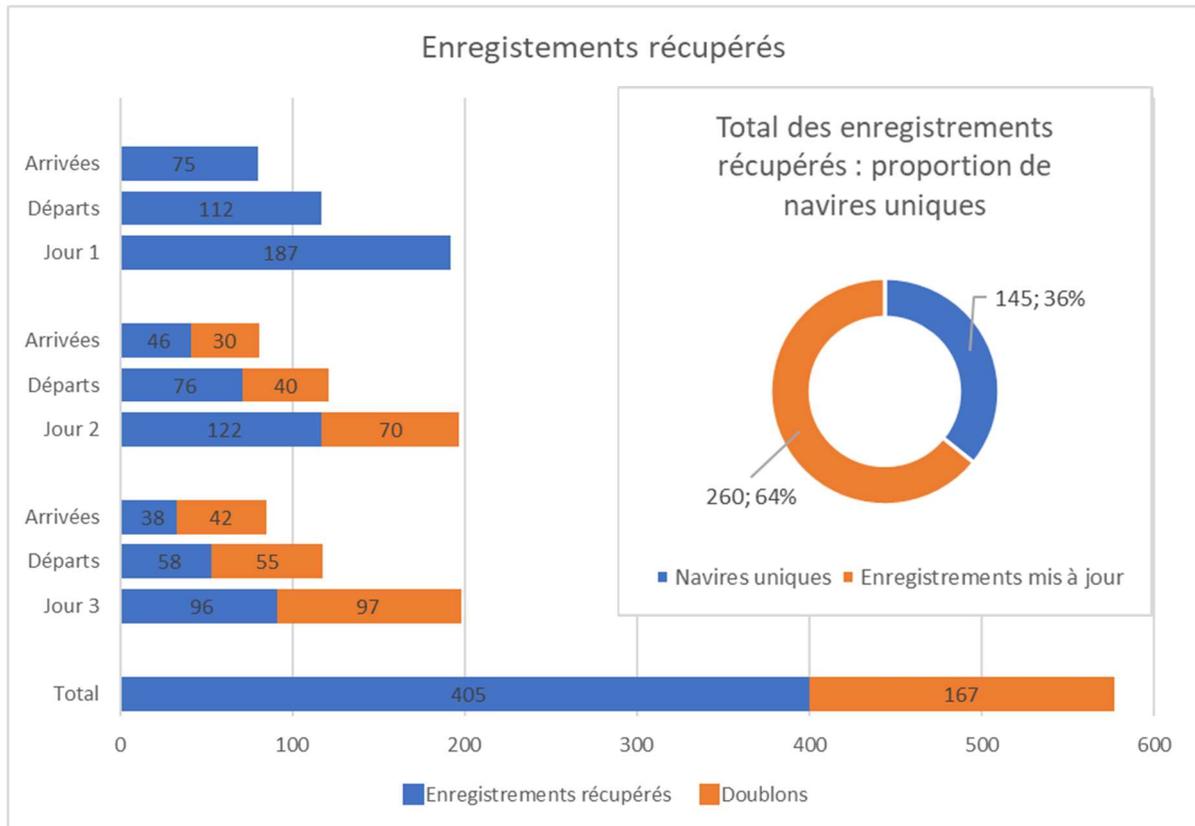
Tableau 10 : Port d'Amsterdam, taux de réponse des données récupérées

Taux de réponse													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
521	521	406	-	-	-	-	349	431	237	521	-	-	428
100.0%	100.0%	77.9%	-	-	-	-	67.0%	82.7%	45.5%	100.0%	-	-	82.1%

6.3.4 Le Havre

Le grand port maritime du Havre nous a fourni 405 enregistrements, pour 145 navires uniques, sur les trois jours de notre preuve de concept, soit une des proportions les plus faibles de notre échantillon. Les enregistrements possèdent soit le statut « Attendu », soit le statut « Parti », selon le script et la page web dont ils sont issus.

Figure 11 : Port du Havre, détail des enregistrements récupérés



Les données brutes ont été réconciliées et nettoyées comme suit :

- Les valeurs du champ « Nom », entièrement en majuscules, ont été modifiées selon la transformation *To titlecase*.
- Suppression des valeurs inconnues (2 pour le champ « IMO », 68 pour le champ « Destination »)
- Les valeurs des champs « Pavillon », « Provenance » et « Destination » ont été réconciliées.
- Le type de navire a été réconcilié selon le tableau 11

Tableau 11 : Port du Havre, réconciliation du type de navire

Valeur originale	Valeur après réconciliation
AUTOMOTEUR	automoteur (non réconcilié)
BARGE	barge
CHEM.OIL CARRIER	chemical oil carrier (non réconcilié)
CHEM. TANK	chemical tanker
CONTAINER CARRIER	container ship
CRUDE OIL TANKER	crude oil tanker (non réconcilié)
GENERAL CARGO	cargo ship
LIQUEFIED GAS TANKER	gas carrier

PASSENGER RO/RO	passenger ro/ro (non réconcilié)
PRODUCT TANKER	product tanker (non réconcilié)
PUSHER TUG	pusher
RO/RO	roll-on/roll-off ship
SUCTION HOPPER DREDGER	suction hopper dredger (non réconcilié)
TRAILING SUCTION HOPPER DREDGER	trailing suction hopper dredger
TUG	tugboat
VEHICLE CARRIER	car carrier
YACHT	yacht

Le port du Havre propose un taux de réponse plutôt bon avec 88,4%. Si l'on excepte le champ « Terminal », ce taux monte même à 96,5%. Il s'agit d'un des ports proposant le plus d'informations intéressantes, surtout en ce qui concerne les données liées au voyage.

Tableau 12 : Port du Havre, taux de réponse des données récupérées

Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
405	405	405	335	404	404	337	405	405	30	405	-	-	-
100.0%	100.0%	100.0%	82.7%	99.8%	99.8%	83.2%	100.0%	100.0%	7.4%	100.0%	-	-	-

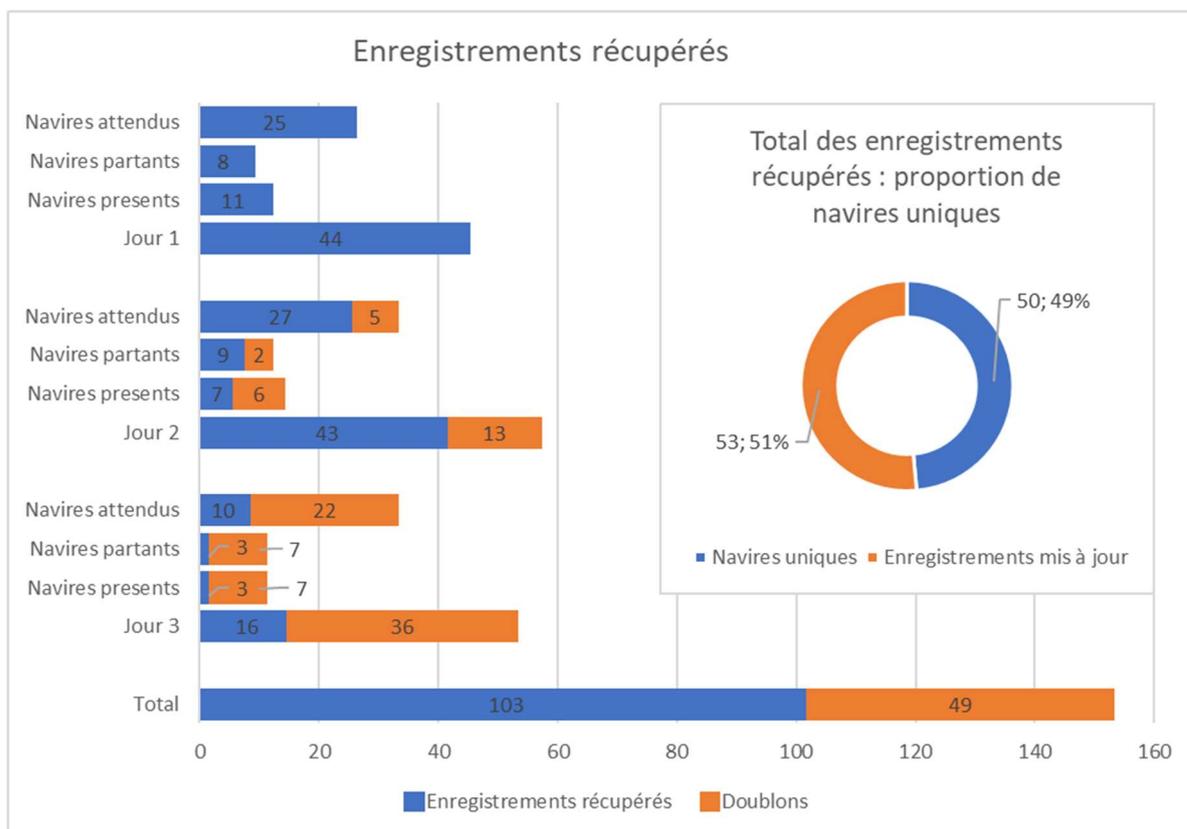
6.3.5 Dunkerque

Le grand port maritime de Dunkerque nous a fourni 103 enregistrements, pour 50 navires uniques, sur les trois jours de *scraping*. Les enregistrements possèdent les statuts « À quai », « Attendu » et « Partant » selon le script et la page dont ils sont issus.

Possédant pourtant une quantité de marchandises transbordées annuelle assez proche du port du Havre, le nombre d'enregistrements récupérés pour le port de Dunkerque est près de quatre fois moindre.

Remarquons également que plus de 60% des enregistrements récupérés proviennent des navires attendus. La proportion d'enregistrements correspondant à des mises à jour de navires est assez élevée, ce qui est normal étant donné la présence potentielle d'un même navire dans les trois pages web *scrapées*.

Figure 12 : Port de Dunkerque, détail des enregistrements récupérés



Les données brutes ont été uniformisées et reconciliées de la manière suivante :

- Les valeurs du champ « Nom », entièrement en majuscules, ont été modifiées selon la transformation *To titlecase*.
- 21 valeurs inconnues ont été effacées dans le champ « Destination ».
- Les champs « Provenance » et « Destination » ont été reconciliés.

Avec un taux de réponse moyen de 82.5%, on remarque que les données dynamiques liées au voyage sont rarement complètes ce qui correspond toutefois à une certaine logique. Pour les champs « Provenance » et « ETA », ce sont les enregistrements des navires partants qui ne sont pas renseignés. Pour le champ « ETD », il s'agit des enregistrements des navires attendus.

Fait inquiétant, on peut remarquer que l'hétérogénéité des données dans le champ « Quai » amène des valeurs se voulant être visiblement identiques à être écrites différemment (abréviations, casse et orthographe différentes).

Tableau 13 : Port de Dunkerque, taux de réponse des données récupérées

Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	Marchandise déchargée	Marchandise chargée	Agent
103	103	-	-	-	83	82	83	40	-	83	-	-	103
100.0%	100.0%	-	-	-	80.6%	79.6%	80.6%	38.8%	-	80.6%	-	-	100.0%

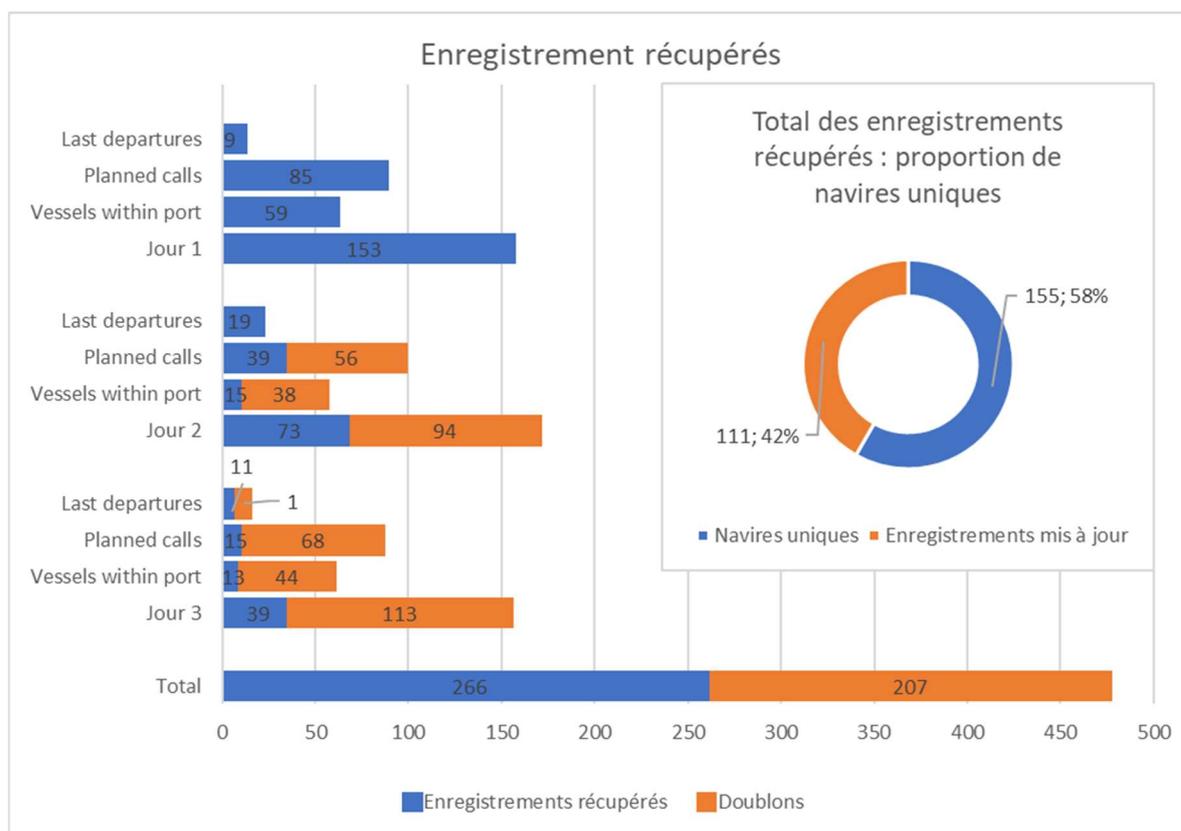
6.3.6 Klaipeda

Bien que quatre scripts aient été utilisés pour récupérer les données du port de Klaipeda, nous avons remarqué que les données présentes sur la page *Last calls* étaient incluses, à une exception près, dans la page *Vessels within port*. Par conséquent, les résultats de la collecte de données de la page *Last calls* ne seront pas représentés graphiquement, l'unique enregistrement étant rajouté au total final.

Les statuts des enregistrements ont été défini comme suit : « Parti » pour les données récupérées de la page *Last departures*, « Attendu » pour la page *Planned calls* et « À quai » pour la page *Vessels within port*.

Avec 266 enregistrements proposés pour 155 navires, Klaipeda fait mieux que Dunkerque et, malgré les quatre pages *scrapées*, propose une proportion de navires uniques dans ses enregistrements de près de 60%.

Figure 13 : Port de Klaipeda, détail des enregistrements récupérés



Les données brutes ont été uniformisées et reconciliées de la manière suivante :

- Les valeurs du champ « Nom », entièrement en majuscules, ont été modifiées selon la transformation *To titlecase*.
- Le champ « Pavillon », en lituanien, a été reconcilié.

Le taux de réponse moyen du port de Klaipeda est de 84,2%. En ce qui concerne les données d'identification, 17 navires (sur 155) n'ont pas le champ « IMO » rempli. Pour les données liées au voyage, les 39 enregistrements ayant le statut « Parti », provenant de la page *Last departures*, ne renseignent pas l'ETA. À l'inverse, ce sont les seuls à donner l'information « ETD ».

Tableau 14 : Port de Klaipeda, taux de réponse des données récupérées

Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	Marchandise déchargée	Marchandise chargée	Agent
266	266	266	249	-	-	-	227	39	-	213	-	-	266
100.0%	100.0%	100.0%	93.6%	-	-	-	85.3%	14.7%	-	80.1%	-	-	100.0%

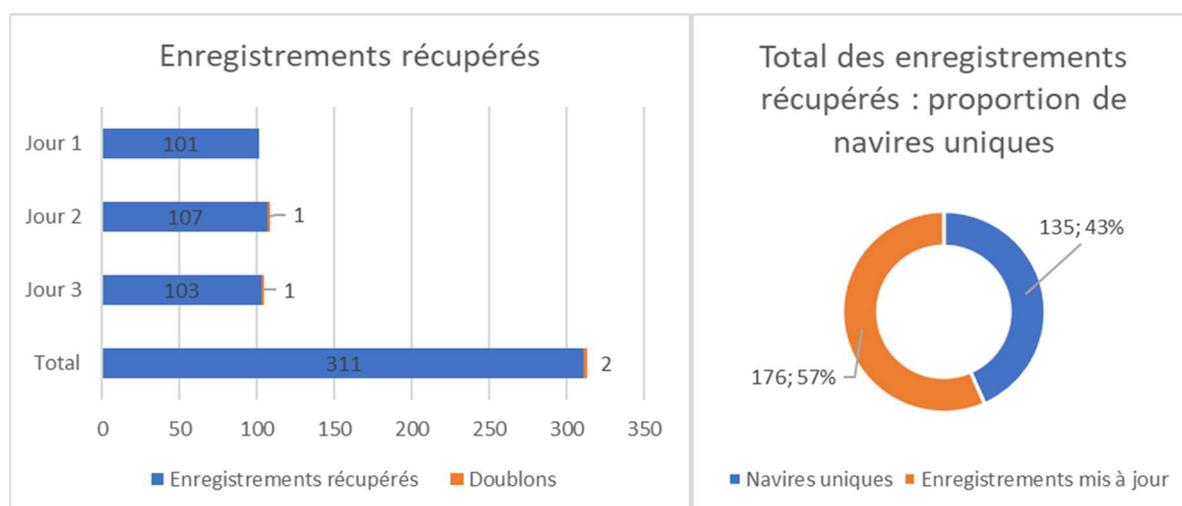
Remarquons encore l'existence de données « extrêmes » provenant de la page *Vessels within ports*. Des navires ont comme valeur dans leur champ « ETA » des dates remontant jusqu'en 2013.

6.3.7 Niedersachsen

Le site web des ports de Niedersachsen propose 311 enregistrements pour 135 navires uniques, ce qui est plus que les ports de Dunkerque et de Klaipeda, qui sont pourtant de tailles plus importantes. La particularité du *scraping* du site de Niedersachsen est le très faible nombre de doublons récupérés sur les jours deux et trois⁶². Toutefois, on obtient un nombre conséquent d'enregistrements mis à jour, surtout pour un unique script utilisé.

Le champ « Statut » a été directement traduit dans le code du script pour obtenir les valeurs suivantes : « Parti » (*abgefahren*), « À quai » (*angekommen*) et « Attendu » (*avisiert*).

Figure 14 : Ports de Niedersachsen, détail des enregistrements récupérés



Les données brutes ont été uniformisées et reconciliées de la manière suivante :

- Les valeurs du champ « Nom », de casse variable, ont été modifiées selon la transformation *To titlecase*.
- Les données inconnues (*Unbekannt*) dans les champs « Provenance » (47 enregistrements) et « Destination » (95 enregistrements) ont été effacées.
- Le champ « Pavillon », composé d'abréviations de deux lettres, a été reconcilié pour obtenir des noms de pays complets, tout comme les champs « Provenance » et « Destination ».

⁶² La page *scrapée* est mise à jour toutes les 4 heures, ce qui explique la faible quantité de doublons.

Le champ relatif au type de bateau que propose Niedersachsen est très générique et donc peu pertinent. Avec les 3 choix, *Fischerei*, *Seeschiff* et *Bes. Fahrzeug*, ces données ont été laissées telles quelles sans passer par une réconciliation. Remarquons encore que les enregistrements récupérés ont pour terminal un des cinq ports suivants : Brake, Cuxhaven, Emden, Stade-Bützfleth et Wilhelmshaven. Enfin les données relatives au champ « Quai » sont incorrectes à cause d'une erreur de code : on peut constater dans les enregistrements récupérés qu'elles sont identiques pour chaque jour de *scraping*.

Le taux de réponse suite au *scraping* est très bon, avec 95,4% de champs remplis. Toutefois, il est important de préciser, comme vu plus haut, que les champs « Type » et « Quai » sont inexploitable pour des raisons qui leurs sont propres.

Tableau 15 : Ports de Niedersachsen, taux de réponse des données récupérées

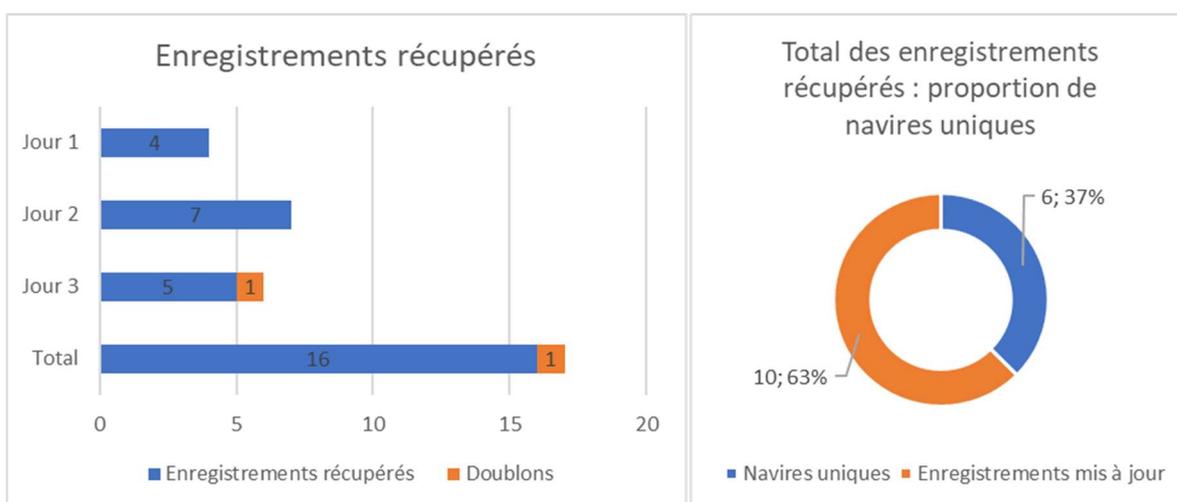
Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
311	311	311	-	311	264	216	311	311	311	311	-	-	-
100.0%	100.0%	100.0%	-	100.0%	84.9%	69.5%	100.0%	100.0%	100.0%	100.0%	-	-	-

6.3.8 Copenhague

L'opération de *scraping* du port de Copenhague fut particulière étant donné que nous avons sélectionné uniquement les enregistrements des *tankers*. Ainsi, la quantité de données extraites ne peut être comparée avec celle des autres ports de notre échantillon.

Bien que peu de doublons ont été signalés, une grande proportion d'enregistrements mis à jour est mentionnée. Cela signifie donc que les données des navires ont été régulièrement mises à jour et que nous n'avons pas manqué d'enregistrements entre deux opérations de *scraping*.

Figure 15 : Port de Copenhague, détail des enregistrements récupérés



Les données brutes ont été réconciliées de la manière suivante :

- Le champ « Pavillon », en danois, a été réconcilié, tout comme les champs « Type » (*tanker* passant à *tanker ship*), « Provenance » et « Destination ».

Avec un taux de réponse moyen de plus de 97% dans un référentiel quasiment complet, le port de Copenhague, bien que proposant que peu d'enregistrements, est un des ports de notre échantillon qui offre le plus de données par navire.

Tableau 16 : Port de Copenhague, taux de réponse des données récupérées

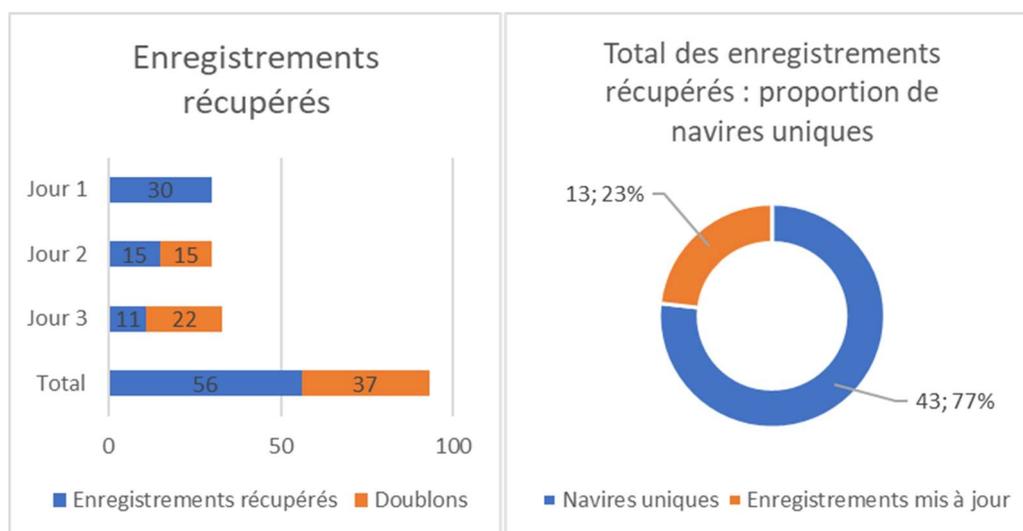
Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
16	16	16	16	16	16	12	16	16	16	16	-	-	-
100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	75.0%	100.0%	100.0%	100.0%	100.0%	-	-	-

6.3.9 Aarhus

Le site web du port danois d'Aarhus nous a fourni 56 enregistrements, dont 43 représentent des navires uniques. On obtient le plus faible taux d'enregistrements mis à jour avec seulement 23%.

Le champ « Statut » a été directement traduit dans le code du script pour obtenir les valeurs suivantes : « Attendu » (*Expected*) et « À quai » (*Arrived*).

Figure 16 : Port d'Aarhus, détail des enregistrements récupérés



Les données brutes ont été uniformisées et reconciliées de la manière suivante :

- Les valeurs du champ « Nom » ont été modifiées selon la transformation *To titlecase*.
- Les valeurs inconnues du champ « Type » ont été effacées (18 enregistrements).
- Le champ « Type » a ensuite été reconcilié.

Malgré un référentiel peu rempli (on ne retrouve que deux données liées à l'identification) le taux de réponse moyen est bon avec plus de 95% de champs renseignés.

Tableau 17 : Port d'Aarhus, taux de réponse des données récupérées

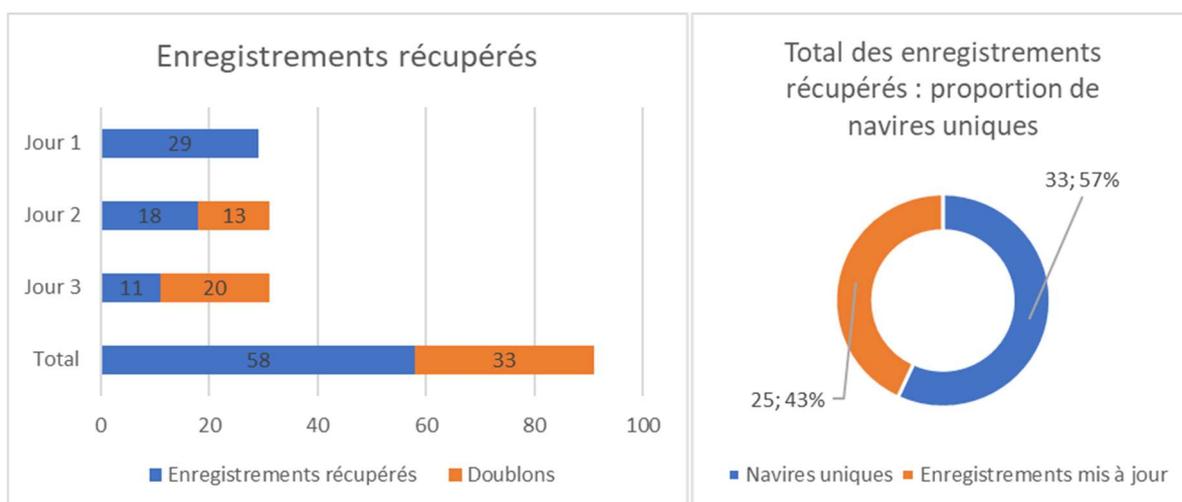
Taux de réponse													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
56	56	-	-	38	-	-	56	56	56	56	-	-	-
100.0%	100.0%	-	-	67.9%	-	-	100.0%	100.0%	100.0%	100.0%	-	-	-

6.3.10 Fredericia

D'importance similaire au port d'Aarhus, le port de Fredericia nous a fourni une quantité de données similaire avec 58 enregistrements pour 33 navires uniques.

Le champ « Statut » a été directement traduit dans le code du script pour obtenir les valeurs suivantes : « Attendu » (*Expected*) et « À quai » (*Arrived*). Précisons également que les données du champ « Agent » proviennent de la colonne *Broker* de la page *scrapée*.

Figure 17 : Port de Fredericia, détail des enregistrements récupérés



La seule uniformisation des données brutes a été la transformation *To titlecase* des valeurs du champ « Nom ».

Fredericia est le port de notre échantillon fournissant le moins de données d'après le référentiel normalisé. En contrepartie, son taux de réponse est le meilleur de notre échantillon avec une moyenne à 99,7%.

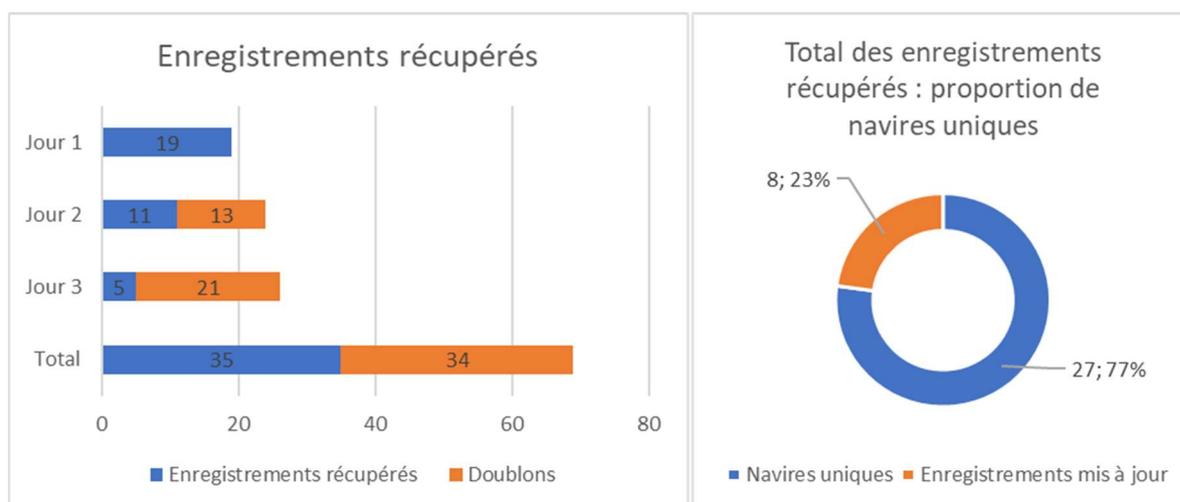
Tableau 18 : Port de Fredericia, taux de réponse des données récupérées

Taux de réponses													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	Marchandise déchargée	Marchandise chargée	Agent
58	58	-	-	-	-	-	58	58	-	58	-	-	57
100.0%	100.0%	-	-	-	-	-	100.0%	100.0%	-	100.0%	-	-	98.3%

6.3.11 Bordeaux

Plus « petit » port de notre échantillon, le port de Bordeaux est également le port proposant le moins d'enregistrements (si l'on excepte le cas particulier de Copenhague), soit 35 enregistrements pour 27 navires uniques.

Figure 18 : Port de Bordeaux, détail des enregistrements récupérés



Le champ « Statut » est renseigné de manière à mentionner pour chaque enregistrement les deux types de statut disponible (« Attendu/À quai »), étant donné qu'aucun champ ne permettait la différenciation du statut lors du *scraping*.

Les données brutes ont été uniformisées et reconciliées de la manière suivante :

- Les valeurs du champ « Nom », de casse variable, ont été modifiées selon la transformation *To titlecase*.
- Les valeurs des champs concernant les marchandises ont été nettoyées.
- Les champs « Provenance » et « Type » ont été reconciliés, ce dernier selon le tableau suivant :

Tableau 19 : Port de Bordeaux, réconciliation du champ "Type"

Valeur originale	Valeur après réconciliation
CHIMIQUIER	chemical tanker
DRAGUE	dredger
P.C. INTEGRAL	container ship
PAQUEBOT	passenger ship
PETRO.DLE COQUE	oil tanker
CARGO > 500 TJB	cargo ship
NAV. CITERNE	tanker ship
BARGE	barge
SCIENTIFIQUE	research vessel
GAZ LIQUEFIÉS	gas carrier
VRAQUIER	bulk carrier

Bien que ne proposant que peu d'enregistrements, le port de Bordeaux était l'un des ports les plus prometteurs en ce qui concerne la diversité des données liées au voyage qu'il proposait. Il était notamment le seul à proposer des informations concernant le détail des marchandises transportées. Malheureusement les données récupérées sont très incomplètes, le taux de

réponse moyen n'est que de 68.6%. De plus, les champs renseignant le pavillon et la sortie prévue (ETD) ne sont jamais remplis.

Tableau 20 : Port de Bordeaux, taux de réponse des données récupérées

Taux de réponse													
Statut	Nom	Pavillon	IMO	Type	Prov.	Dest.	ETA	ETD	Terminal	Quai	March. déchargée	March. chargée	Agent
35	35	0	-	35	35	-	35	0	-	-	23	7	35
100.0%	100.0%	0.0%	-	100.0%	100.0%	-	100.0%	0.0%	-	-	65.7%	20.0%	100.0%

6.4 Synthèse

Il est important de rappeler que, dans notre analyse, nous nous sommes limités principalement à l'aspect quantitatif des données. Nous n'avons pas vérifié ou corrigé l'aspect qualitatif, si l'on excepte la présence de valeurs inconnues ou nulles qui ont été effacées. Il est donc impossible de certifier l'exactitude des données *scrapées*, particulièrement en ce qui concerne les données liées au voyage, comme les dates et heures d'arrivée et de départ des navires. Et comme nous l'avons déjà vu, la source des données n'est pratiquement jamais mentionnée, ce qui nous empêche d'avoir la certitude que nous n'avons pas affaire à des données issues de l'AIS.

6.4.1 Remarques générales

Au total 3'322 enregistrements ont été récoltés, représentant des données pour 1'636 navires, soit un ratio de quasiment 50%. Cela implique que plus de la moitié des enregistrements récoltés sont constitués en fait de mises à jour sur la situation d'un navire. Le tableau 21 de la page suivante nous permet d'émettre plusieurs constats :

L'ordre de la quantité d'enregistrements récupérés selon l'importance du port, en ce qui concerne la quantité de marchandises transbordées, est pratiquement toujours respecté. Seul le port de Dunkerque semble être très limité dans le nombre d'enregistrements qu'il propose sur son site web, et ce par rapport aux autres ports de son acabit.

Au niveau des différentes informations que proposent les sites web, on remarque que la quantité de données proposées par rapport au référentiel normalisé reste très médiocre. Ainsi, on n'atteint jamais plus de 80% de champs renseignés, avec une moyenne à 61%. Les données liées au voyage des navires sont les moins représentées par rapport au référentiel normalisé (59%). Cela signifie qu'en moyenne, on ne compte que 6 champs liés au voyage des navires de notre référentiel normalisé sur 10 qui sont renseignés.

Si l'on va plus loin et que l'on s'intéresse aux données véritablement présentes dans ces champs renseignés, on obtient une moyenne du taux de réponse de 86,6%. À noter que les ports danois sont tous à plus de 95% de taux de réponse. Les données liées à l'identification des navires sont renseignées à près de 94%. Les données liées aux voyages sont quant à elles moins bien remplies avec une moyenne à 82,7%. On explique cette différence car il y a tout simplement beaucoup plus de champs potentiels pour le voyage du navire (10 contre 4 pour l'identification).

Finalement, non seulement le référentiel normalisé n'est pas totalement rempli, mais en plus certains champs proposés ne sont pas complètement renseignés et même dans certains cas, comme le pavillon et la prévision de date de départ pour le port de Bordeaux, jamais remplis.

Tableau 21 : Synthèse des données récupérées par port

	Port	Marchandises transbordées en 2018 (en million de tonnes) ⁶³	Pages web scrapées / scripts	Proportion des champs renseignés selon le référentiel normalisé ⁶⁴			Taux de réponse ⁶⁵			Enregistrements récupérés	Navires uniques (proportion par rapport au total des enregistrements)
				Données liées à l'identité du navire	Données liées au voyage du navire	Total	Données liées à l'identité du navire	Données liées au voyage du navire	Total		
1	Rotterdam	441,5	3	75% (3)	50% (5)	57% (8)	99,9%	78,4%	86,5%	984	529 (54%)
2	Hambourg	117,6	2	75% (3)	40% (4)	50% (7)	100%	52,4%	72,8%	567	208 (37%)
3	Amsterdam	99,5	1	50% (2)	60% (6)	57% (8)	89%	79,6%	81,9%	521	305 (59%)
4	Le Havre	64,9	2	100% (4)	70% (7)	79% (11)	95,6%	84,3%	88,4%	405	145 (36%)
5	Dunkerque	51,6	3	25% (1)	70% (7)	57% (8)	100%	80%	82,5%	103	50 (49%)
6	Klaipeda	46,6	4 (3) ⁶⁶	75% (3)	50% (5)	57% (8)	97,9%	76%	84,2%	266	155 (58%)
7	Niedersachsen	28,3	1	75% (3)	70% (7)	71% (10)	100%	93,5%	95,4%	311	135 (43%)
8	Copenhague	15,1 ⁶⁷	1	100% (4)	70% (7)	79% (11)	100%	96,4%	97,7%	16	6 (37%)
9	Aarhus	8,5	1	50% (2)	50% (5)	50% (7)	84%	100%	95,4%	56	43 (77%)
10	Fredericia	6,9	1	25% (1)	50% (5)	43% (6)	100%	99,7%	99,7%	58	33 (57%)
11	Bordeaux	6,8 ⁶⁸	1	75% (3)	70% (7)	71% (10)	66,7%	69,4%	68,6%	35	27 (77%)
				66%	59%	61%	93,9%	82,7%	86,6%	3322	1636 (49%)

⁶³ Sources : voir tableau 2, p. 10.

⁶⁴ Proportion de champs proposé par chaque site web par rapport au référentiel normalisé, avec entre parenthèses leur nombre effectif. Le nombre total de champs du référentiel normalisé est de 14 (4 concernant l'identité du navire et 10 concernant son voyage).

⁶⁵ Proportion de données effectivement présentes dans les champs du référentiel de chaque port.

⁶⁶ Voir chapitre 6.3.6.

⁶⁷ Comprend le port de Malmö en Suède également.

⁶⁸ Pour l'année 2019.

Bien entendu, les données obtenues ont toujours une certaine valeur, notamment si l'on prend en compte les données dynamiques des heures d'arrivée et de départ qui peuvent être, lors d'une étape ultérieure, comparées aux données de l'AIS. Cependant, on garde l'impression que les données fournies par les ports gagneraient à être complétées par celles provenant d'autres acteurs du transport maritime, notamment dans les champs qui ont très souvent fait défaut dans cette preuve de concept, comme tout ce qui concerne le détail des marchandises transportées.

6.4.2 Cas particuliers

Une fois les données fusionnées en seul fichier, le nombre de navires uniques atteint 1472, cela signifie donc que certains navires apparaissent dans les enregistrements de plusieurs ports. Ceci amène à quelques observations particulières.

6.4.2.1 Du manque de données d'identification

L'importance d'avoir le numéro IMO d'un navire est déterminante dans son identification. La capture suivante, avec des enregistrements récupérés du port de Hambourg et du port de Klaipeda, le démontre avec deux navires ayant le même nom, *Amethyst*, l'un étant un *tanker ship* au pavillon chypriote⁶⁹, l'autre un *general cargo ship* au pavillon russe⁷⁰.

Figure 19 : *Amethyst*, un même nom pour deux navires

Date scraping	Port source	Statut	Nom	Pavillon	IMO	Type	ETA	ETD	Quai	Agent
2020-06-25	Hambourg (DE)	Parti	Amethyst		9246918	tanker ship		2020-06-18 14:10		
2020-06-25	Hambourg (DE)	À quai	Amethyst		9246918	tanker ship	2020-06-17 21:10			
2020-06-25	Klaipeda (LT)	À quai	Amethyst	Russia	9549607		2020-06-22 07:07		15	UAB EUROGA

6.4.2.2 Du décryptage d'un voyage

Les deux exemples suivants illustrent les possibilités de reconstitution de voyage grâce aux données qui ont été récupérées dans notre preuve de concept.

Le Bomar Quest est un *tanker ship* au pavillon des Îles Marshall⁷¹. Nous pouvons déterminer un extrait de son parcours grâce aux enregistrements issus du *web scraping* présentés dans la figure 20 de la page suivante. Le navire apparaît d'abord à Rotterdam comme étant à quai et partant le 25 juin à 23h (enregistrement n°3109). Il est ensuite déclaré parti le 25 à 23h59 (n°2570) puis finalement le 26 juin à 16h04 (n°2696). Étant donné que ce dernier enregistrement est plus récent, on aura tendance à l'interpréter comme une mise à jour de l'heure effective du départ du port.

Du côté de Niedersachsen, on attend le Bomar Quest pour le 27 juin à 17h avant qu'il ne reparte pour Tjeldbergodden le 28 juin à 5h du matin (n°2227 et 2320). Bien que son type soit inconnu, le fait que son agent soit précisé, *North Sea Tankers BV*, nous fait présumer à juste titre qu'il s'agit d'un *tanker ship*.

⁶⁹ <https://www.vesselfinder.com/fr/vessels/AMETHYST-IMO-9246918-MMSI-209223000>

⁷⁰ <https://www.vesselfinder.com/fr/vessels/AMETHYST-IMO-9549607-MMSI-273358930>

⁷¹ <https://www.vesselfinder.com/fr/vessels/BOMAR-QUEST-IMO-9272735-MMSI-538002583>

Figure 20 : Le voyage du *Bomar Quest*

	Date scraping	Port source	Statut	Nom	Pavillon	IMO	Type	Provenance	Destination	ETA	ETD	Quali	Agent
3109.	2020-06-25	Rotterdam (NL)	À quai	Bomar Quest	Marshall Islands	9272735					2020-06-25 23:00	CALANDK EVOS MIDDEN	North Sea Tankers BV
2227.	2020-06-26	Niedersachsen (DE)	Attendu	Bomar Quest	Marshall Islands		Seeschiff	Rotterdam	Tjeldbergodden	2020-06-27 17:00	2020-06-28 05:00	Nordkai Ost	
2570.	2020-06-26	Rotterdam (NL)	Parti	Bomar Quest	Marshall Islands	9272735					2020-06-25 23:59	CALANDK EVOS MIDDEN	North Sea Tankers BV
2320.	2020-06-27	Niedersachsen (DE)	Attendu	Bomar Quest	Marshall Islands		Seeschiff	Rotterdam	Tjeldbergodden	2020-06-27 17:00	2020-06-28 05:00	Hafenkaje	
2696.	2020-06-27	Rotterdam (NL)	Parti	Bomar Quest	Marshall Islands	9272735					2020-06-26 16:04	3E PET VOPAK 3	North Sea Tankers BV

À noter que l'identification pour le port de Niedersachsen étant sans le numéro IMO, on présume dans ce cas qu'il s'agit du même navire que pour le port de Rotterdam grâce au champ indiquant la provenance.

D'autres voyages sont plus complexes à interpréter, notamment parce que certaines données fournissent malgré elles du bruit par leur obsolescence. C'est le cas pour le *MSC Geneva* qui est annoncé dans notre preuve de concept par 3 ports différents pour 8 enregistrements au total, comme le montre la figure 18.

Les enregistrements proviennent des ports du Havre, de Hambourg et de Rotterdam. Les données du port du Havre nous indiquent que le navire vient d'Anvers et se dirige vers Limassol. Les enregistrements *scrapés* le 25 juin nous informent que le navire est d'abord attendu à Hambourg le 25 juin à 15h (enregistrement n°791), puis à Rotterdam le 27 juin à 5h14 (n°2835), puis enfin au Havre le 30 juin à 19h30, qu'il quittera le 1^{er} juillet à 14h (n°1651). On sait grâce au *scraping* des jours suivants, qu'il est arrivé au port de Hambourg le 25 juin à 13h40 (n°1326), qu'il en est parti le 26 juin à 12h30 (n°1277), et qu'il est pour l'instant attendu au port de Rotterdam, d'après la dernière mise à jour, pour le 27 juin à 11h22 (n°2977).

Figure 21 : Le voyage du *MSC Geneva*

	Date scraping	Port source	Statut	Nom	Pavillon	IMO	Type	Provenance	Destination	ETA	ETD	Quali	Agent
791.	2020-06-25	Hambourg (DE)	Attendu	Msc Geneva		9320427	container ship			2020-06-25 15:00		Eurogate-CTH	
1651.	2020-06-25	Le Havre (FR)	Attendu	Msc Geneva	Portugal	9320427	container ship	Antwerp	Limassol	2020-06-30 19:30	2020-07-01 14:00	QUAI DU HAVRE	
2835.	2020-06-25	Rotterdam (NL)	Attendu	Msc Geneva	Portugal	9320427				2020-06-27 05:14		EUROPAH ECT DDN	MSC Mediterranean Shipping Co
1818.	2020-06-25	Le Havre (FR)	Parti	Msc Geneva	Portugal	9320427	container ship	Antwerp	Limassol	2020-06-30 19:30	2020-07-01 14:00	QUAI DU HAVRE	
1326.	2020-06-26	Hambourg (DE)	À quai	Msc Geneva		9320427	container ship			2020-06-25 13:40			
2903.	2020-06-26	Rotterdam (NL)	Attendu	Msc Geneva	Portugal	9320427				2020-06-27 13:14		EUROPAH ECT DDN	MSC Mediterranean Shipping Co
1277.	2020-06-26	Hambourg (DE)	Parti	Msc Geneva		9320427	container ship				2020-06-26 12:30		
2977.	2020-06-27	Rotterdam (NL)	Attendu	Msc Geneva	Portugal	9320427				2020-06-27 11:22		EUROPAH ECT DDN	MSC Mediterranean Shipping 🔗

Après avoir déterminé ce parcours, on se rend compte que les enregistrements n°2835 et 2903 sont des mises à jour obsolètes du port de Rotterdam qui ne sont plus utiles. L'enregistrement n°791 du port de Hambourg est également obsolète et le n°1818 du port du Havre n'apporte aucune information supplémentaire. Dans ce cas spécifique, il n'y a donc que 4 enregistrements sur 8 qui nous sont vraiment utiles.

Remarquons encore que grâce aux enregistrements des trois ports, il est possible de compléter les données du navire et de se rapprocher de l'idéal du référentiel normalisé.

Pour conclure, et au vu de ces trajectoires déchiffrées, on peut imaginer qu'avec beaucoup plus de ports et de jours de *scraping*, il serait relativement facile d'obtenir les coordonnées temporelles des voyages des navires qui nous intéressent.

7. Difficultés et recommandations

Avant d'évoquer les principales difficultés rencontrées lors des diverses étapes de la réalisation de ce travail de master, on peut commencer par mentionner sa complexité due à son aspect pluridisciplinaire. L'appréhension du domaine du transport maritime de matières premières, puis la recherche systématique de sources web a été menée selon les principes de base des sciences de l'information, alors que la preuve de concept nécessita, elle, l'usage de compétences en programmation Python et en curation de données. Mais c'est finalement cette complexité qui aura rendu ce travail passionnant, un travail qui se rapproche d'ailleurs de la réalisation d'un projet de par son aspect très pratique.

7.1 De la problématique du langage à la normalisation des données

Nous l'avons vu, malgré la restriction de notre étude à l'Europe géographique, la problématique liée à la langue utilisée sur certains sites web pour trouver des données pertinentes a déjà été rencontrée, notamment dans les sites de ports turcs ou grecs. Bien entendu, cette problématique va se généraliser si la démarche de ce travail est étendue au reste du monde, et notamment aux acteurs du continent asiatique qui représentent la plus grande activité du transport maritime commercial.

Au-delà de la recherche d'information, on rencontre également cette problématique liée au langage au niveau des données que l'on récupère. Nous avons rencontré ce cas ici, avec des données en lituanien ou en allemand par exemple, qu'il a fallu réconcilier avec OpenRefine afin de rendre les données de chaque port uniformes. Cette difficulté implique qu'au-delà de la création d'un référentiel normalisé, qui détermine les différents champs que l'on veut récupérer sur chaque site web comme nous l'avons réalisé ici, une deuxième étape serait de définir des normes d'uniformisation, basées sur une langue de référence ou sur des abréviations standardisées, pour les valeurs comprises dans les champs récupérés. Au final, cela permettrait d'avoir un référentiel normalisé non seulement au niveau des champs récupérés, mais également au niveau des valeurs des données elles-mêmes.

7.2 La recherche de nouvelles données exploitables

Certains pays ne partagent de manière générale aucune information sur les sites web de leurs ports commerciaux. Nous avons rencontré cette problématique pour des pays comme l'Italie ou la Turquie. Et d'autres acteurs ne partagent tout simplement pas d'information par souci de confidentialité, ainsi il nous a été impossible de trouver des sources web pour obtenir des rapports d'agents maritimes par exemple. Pressé par le temps imparti à ce travail, qui nécessitait de définir rapidement des sources en vue de réaliser une preuve de concept, il n'a pas été possible d'effectuer un état des lieux complet des données disponibles sur le web pour chacun des acteurs du transport maritime de matières premières. Fort de l'expérience acquise au terme de ce travail, on peut considérer que la cartographie complète de la présence sur le web de ces différents acteurs, et par conséquent plus seulement ceux actifs dans l'Europe géographique mais également dans le monde entier, ainsi que l'expression des données qu'ils partagent publiquement, pourraient être l'objet d'un travail de master à lui tout seul. Il serait intéressant, dans un tel travail, de déterminer si les tendances que nous avons pu observer en Europe, notamment en ce qui concerne la faible quantité de données rendues publiques sur le web, se vérifie à l'échelle internationale. Cependant avant d'aller aussi loin, on peut

commencer par se concentrer sur les ports commerciaux du monde entier, comme expliqué au chapitre 7.4.2.

7.2.1 Veilles envisagées

Lorsqu'il s'agit de rechercher de nouvelles sources de données, ou de se tenir au courant de l'évolution de la disponibilité de données déjà inventoriées, comme nous l'avons fait avec les sites web de ports commerciaux, l'usage de techniques de veille est rapidement envisagé.

Une veille sur les sites web des ports n'offrant à l'heure actuelle pas de données sur les escales des navires peut être envisagée. En effet, plusieurs sites web de ports ont évolué pendant notre étude, comme par exemple le North Sea Port⁷² qui propose désormais une section complète concernant les mouvements de navires⁷³. Dans notre échantillon, le port d'Amsterdam a notamment vu la page concernant les arrivées et départs de son site web évoluer.

Une seconde veille exploratoire pour trouver d'autres acteurs du transport maritime exposant des données sur le web public serait une excellente initiative. Pour cela il faudrait étudier plus profondément les structures de ces acteurs pour découvrir notamment les instances qui font référence.

7.3 Le web *scraping* de données

La diversité au niveau du code des scripts réalisés, rien que pour un échantillon de onze sites web, est déjà conséquente. Chaque site web de port rencontré possédait son architecture et sa technologie propres, ce qui ne pouvait permettre un simple copier/coller entre scripts. De plus, des technologies web, comme le Javascript, impliquent, au-delà même du code Python, d'utiliser des moyens supplémentaires pour parvenir à nos fins, notamment selon le système d'exploitation sur lequel le *web scraping* est réalisé.

Force est de constater que pour chacun des ports rencontrés dans notre échantillon, la création des scripts a nécessité l'inspection attentive de la page web concernée. L'architecture du code HTML dans ces différents sites est souvent complexe, pas toujours claire ou cohérente, et cela implique de se familiariser avec les notions permettant de localiser les données à récupérer avec l'aide des sélecteurs CSS ou XPATH. Ainsi l'exploitation d'une large boîte à outils a été des plus satisfaisante. On ne peut que recommander d'utiliser le *framework* Scrapy et de profiter de la communauté de développeurs qui le fait vivre. En effet, la quasi-totalité des problèmes que nous avons rencontrés ont trouvé leur solution sur le site web de questions/réponses StackOverflow. De plus, nous recommandons de consulter les mises à jour de Scrapy, qui fournissent des fonctionnalités supplémentaires pouvant s'avérer très utiles.

Enfin, il est évident que le code des différents scripts que nous avons réalisés n'est probablement pas le plus propre ou le plus efficace possible, il aurait par exemple peut-être été possible de réaliser un seul script par port afin de récupérer les enregistrements de toutes les pages web concernées. Si un budget est acquis, l'utilisation de *Scrapy Cloud*⁷⁴ peut

⁷² Il s'agit de la fusion des ports néerlandais de Flessingue et de Terneuzen avec le port belge de Gand.

⁷³ <https://en.northseaport.com/enigma>

⁷⁴ <https://www.scrapinghub.com/scrapy-cloud/>

également être un plus : ce dernier permet, entre autres, de planifier l'exécution de scripts, ce qui peut s'avérer pratique lorsque l'on se retrouve avec plusieurs dizaines de scripts à lancer au même moment.

7.3.1 Uniformisation et réconciliation

Nous avons essayé d'uniformiser déjà quelques valeurs dans le code même des scripts, comme le statut du navire et les dates. Suite aux erreurs obtenues avec le module *dateparser* et la multiplication des types de statut obtenus sur les différents sites web, on se rend compte qu'il aurait peut-être été préférable et plus simple de procéder à des opérations d'uniformisation seulement au terme de l'opération de *web scraping*, c'est-à-dire au moment du traitement des données brutes avec OpenRefine.

En plus de cette volonté d'uniformisation directement dans le code, nous avons également rajouté manuellement un champ pour mentionner le nom du port d'où l'enregistrement est tiré ainsi qu'un champ pour indiquer la date même du *web scraping*. Concernant cette dernière, il nous semblerait presque plus pertinent et plus précis de récupérer, lorsque c'est possible, la date de dernière mise à jour de la page web *scrapée*. Cela permettrait en effet de repérer bien plus rapidement les enregistrements à double lors d'une opération de collecte de données s'étalant sur plusieurs jours.

Le logiciel OpenRefine est pratique jusqu'à un certain point. Dans notre cas, avec 3322 enregistrements, ou *records*, il a fonctionné de manière parfaitement convenable. Cependant, en cas de quantité de données plus importantes, d'autres outils pour le nettoyage des données, comme Dataiku⁷⁵, pourraient être plus appropriés.

Nous avons pratiqué la réconciliation dans OpenRefine avec *wikidata*, qui nous a permis d'utiliser les entités de pays, de ville et de type de bateau pour uniformiser les données que nous avons à disposition. À noter que *wikidata* possède également une entité « IMO Number »⁷⁶, que nous n'avons pas utilisée, les résultats des réconciliations testées étant des plus mitigés. Cependant, nous n'avons pas trouvé d'autres sources de données normalisées qui pourraient être plus adéquates pour les caractéristiques des navires. Précisons également que lors de nos réconciliations, nous n'avons pas cherché à exprimer les données selon une norme ou un standard de codification, comme pour le pavillon notamment, simplement parce qu'aucune norme ne prévalait dans les données proposées originellement par les onze ports de notre échantillon.

Enfin, nous n'avons pas poussé la réflexion sur l'enrichissement des données, notamment pour celles qui concernent l'identification des navires. Il pourrait effectivement être intéressant d'essayer de récupérer des données depuis des APIs, comme celles des sites web *MarineTraffic* ou *VesselFinder*, ou depuis d'autres bases de données web, comme *Equasis*, pour ainsi compléter les données liées à l'identification des navires de certains ports qui sont parfois très limitées.

⁷⁵ <https://www.dataiku.com/>

⁷⁶ <https://www.wikidata.org/wiki/Q1970938>

7.4 Prochaines étapes

7.4.1 Que faire avec ces données ?

Plus de la moitié des enregistrements que l'on a récupéré sont en fait des mises à jour sur la situation d'un navire, notamment sur les données de son voyage. Cela veut donc dire qu'une grande partie de ces enregistrements est donc déjà obsolète au profit de l'enregistrement *scrapé* le plus récent. Cependant, on ne peut pas exclure la possibilité que plusieurs enregistrements concernant un même navire évoquent deux voyages différents, et dans ce cas-là, les deux enregistrements se compléteraient sans que l'un ne rende l'autre dépassé. Cela implique qu'il ne suffit pas de supprimer les enregistrements qui paraissent obsolètes, mais qu'il est nécessaire d'évaluer les enregistrements les uns après les autres, tout cela en cas de nettoyage plus approfondi des données récoltées.

On peut imaginer deux prochaines étapes pour les données que nous avons récoltées. Tout d'abord, une évaluation de leur qualité en ce qui concerne l'exactitude des informations qu'elles contiennent. Ensuite une comparaison avec les données issues de l'AIS afin de découvrir quel est leur niveau de similarité et de complétude. Pour cela, on peut imaginer, à la manière du décryptage d'un voyage vu en chapitre 6.4.2.2, de réaliser pour un échantillon de navires la reconstruction d'une séquence de leurs déplacements, puis de comparer cette séquence issue des données du *web scraping* avec la même séquence de voyage, celle-ci issue de données provenant de l'AIS.

7.4.2 Ouvrir la méthode au reste du monde

D'après le classement des principaux ports mondiaux de l'American Association of Port Authorities (2016), il n'y a que 5 ports de notre échantillon que l'on retrouve dans le classement des 100 ports les plus importants en ce qui concerne le volume de marchandises. Il va sans dire qu'en restant en Europe, nous n'avons pu qu'effleurer l'univers du trafic maritime de matières premières. Les ports des continents asiatiques et américains devraient fournir de nombreuses autres données pertinentes. La méthodologie utilisée pour la réalisation de l'état des lieux des sites web de ports maritimes est d'ailleurs recommandée à plus grande échelle, c'est-à-dire en cherchant d'abord les autorités nationales qui répertorient le plus souvent les structures portuaires les plus importantes, puis en complétant l'échantillon avec les informations données sur les sites web spécialisés.

8. Conclusion

Avec 3'322 enregistrements récoltés, représentant des données relatives à 1'472 navires différents, cette étude exploratoire sur l'application de la technique du *web scraping* à des données publiques du transport maritime disponibles sur le web a été un succès. Non seulement grâce à l'aspect quantitatif des données récoltées, mais aussi car ces dernières permettent de reconstruire des séquences de voyage. Ainsi, la méthode décrite dans ce travail, si elle n'est bien entendu pas parfaite, permet dans tous les cas d'affirmer, à l'heure où les données ont une valeur si particulière, qu'il y a un véritable potentiel à récupérer ce qu'offrent certains acteurs sur le web public.

Bien entendu, il reste des zones d'ombres qu'on ne peut négliger. Premièrement, les sources de ces données publiques. Le fait qu'elles soient le plus souvent non mentionnées indique qu'elles peuvent très bien provenir de l'AIS. Il reste également à établir le niveau de qualité de ces données, notamment en ce qui concerne leur exactitude. Ensuite, le manque d'information contenue dans ces données, notamment en ce qui concerne la cargaison transportée, implique qu'il est difficile de cerner les enregistrements qui se rapportent uniquement aux navires affectés par le transport de matières premières. Enfin, le manque d'information provenant d'autres acteurs du transport de matières premières, comme les agents maritimes par exemple, est à déplorer car ils auraient eu sans nul doute de nombreuses données intéressantes à partager.

Pour conclure, mentionnons que cette étude a été effectuée en pleine pandémie de Covid-19. Même si le coronavirus n'a pas empêché la réalisation de ce travail de master, les navires commerciaux ont bien continué à naviguer et je ne suis pas tombé malade, il sera sans doute intéressant d'analyser dans quelques mois les effets et les conséquences qu'un tel phénomène mondial a eu sur le trafic maritime de matières premières.

Bibliographie

AARHUS HAVN, 2020. Terminals. *aarhushavn.dk* [en ligne]. 2020. [Consulté le 10 juillet 2020]. Disponible à l'adresse : <https://www.aarhushavn.dk/en/terminals/index.htm>

ADLAND, Roar, JIA, Haiying et STRANDENES, Siri P., 2017. Are AIS-based trade volume estimates reliable? The case of crude oil exports. *Maritime Policy & Management*. 4 juillet 2017. Vol. 44, n° 5, p. 657-665. DOI [10.1080/03088839.2017.1309470](https://doi.org/10.1080/03088839.2017.1309470)

ADP A/S, [sans date]. The Port of Fredericia is centrally located at the European transport corridor - ADP A/S. *adp-as.dk* [en ligne]. [Consulté le 10 juillet 2020]. Disponible à l'adresse : <https://www.adp-as.dk/en/ports/port-of-fredericia/>

AMERICAN ASSOCIATION OF PORT AUTHORITIES, 2016. World Port Ranking 2016. *Aapa-ports.org* [en ligne]. 2016. [Consulté le 6 août 2020]. Disponible à l'adresse : <http://aapa.files.cms-plus.com/Statistics/WORLD%20PORT%20RANKINGS%202016.xlsx>

ARMATEURS DE FRANCE, 2017. Armateurs, Affréteurs, Chargeurs. *armateursdefrance.org* [en ligne]. 2017. [Consulté le 14 juillet]. Disponible à l'adresse : http://www.armateursdefrance.org/sites/default/files/decryptages/fiche_armateurs_affreteur_chargeur_adf_2017.pdf

BHAGESHPUR, Kiran, 2019. Data Is The New Oil -- And That's A Good Thing. *Forbes* [en ligne]. 15 novembre 2019. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://www.forbes.com/sites/forbestechcouncil/2019/11/15/data-is-the-new-oil-and-thats-a-good-thing/>

BOL RAAP, Wouter, 2016. *The design of a common data model for generic synchromodal cargo-tracking in logistics using web scraping and big & open data* [en ligne]. University of Twente. [Consulté le 4 août 2020]. Disponible à l'adresse : http://essay.utwente.nl/69690/1/BolRaap_MA_EEMCS.pdf

BORDEAUX PORT ATLANTIQUE, 2020. Chiffres clés | Port de Bordeaux. *bordeaux-port.fr* [en ligne]. 2020. [Consulté le 10 juillet 2020]. Disponible à l'adresse : <https://www.bordeaux-port.fr/fr/le-port-de-bordeaux/chiffres-cles>

CAN, François, GAUDINAT, Arnaud, et THEODORO, Douglas, 2020. *PreciseIntelligence : Big Data Analytics for Comprehensive Global Trade Flow Intelligence* [Innosuisse - online application].

CHABOT, Georges, 1952. Hambourg. Le port et la ville. *L'Information Géographique* [en ligne]. 1952. Vol. 16, n° 4, p. 127-136. [Consulté le 9 juillet 2020]. Disponible à l'adresse : https://www.persee.fr/doc/ingeo_0020-0093_1952_num_16_4_1174

CHAPMAN, Rachael, 2019. 5 Best web scraping languages for 2020 | By Limeproxies. *LimeProxies.com* [en ligne]. 20 décembre 2019. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://limeproxies.com/blog/5-best-web-scraping-languages/>

Computer Fraud and Abuse Act. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 6 juin 2020 à 16:00. [Consulté le 4 juillet 2020]. Disponible à l'adresse : https://en.wikipedia.org/wiki/Computer_Fraud_and_Abuse_Act

CONFERENCE DES NATIONS UNIES SUR LE COMMERCE ET LE DEVELOPPEMENT, 2020. Étude sur les transports maritimes 2019. *unctad.org* [en ligne]. Juin 2020. [Consulté le

4 août 2020]. ISBN 978-92-1-004303-8. Disponible à l'adresse : https://unctad.org/fr/PublicationsLibrary/rmt2019_fr.pdf

COPENHAGEN MALMÖ PORT, 2020. Facts & Figures - Copenhagen Malmö Port. *Cmport.com* [en ligne]. 6 août 2020. [Consulté le 22 juillet 2020]. Disponible à l'adresse : <http://www.cmport.com/corporate/finance/facts-and-figures-54>

DOWINTON, Richard, 2015. Parse Natural Language Dates with Dateparser. *blog.scrapinghub.com* [en ligne]. 9 novembre 2015. [Consulté le 23 juillet 2020]. Disponible à l'adresse : <https://blog.scrapinghub.com/2015/11/09/parse-natural-language-dates-with-dateparser>

DUNKERQUE PORT, 2019. *Rapport d'activité 2018* [en ligne]. [Consulté le 10 juillet 2020]. Disponible à l'adresse : <http://www.dunkerque-port.fr/index.php?cmpref=61757&lang=fr&module=media&action=Display>

DURAND, Michel, [sans date]a. L'armateur. *L'Antenne - Les transports et la logistique au quotidien* [en ligne]. [Consulté le 14 juillet 2020]. Disponible à l'adresse : https://www.lantenne.com/L-armateur_a14021.html

DURAND, Michel, [sans date]b. Fret maritime pratique I L'agent maritime et le consignataire de navire. *L'Antenne - Les transports et la logistique au quotidien* [en ligne]. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://www.lantenne.com/transport-fret-maritime-agent-et-consignataire/>

EUROSTAT, 2020. Maritime ports freight and passenger statistics - Statistics Explained. *ec.europa.eu* [en ligne]. [Consulté le 5 juillet 2020]. Disponible à l'adresse : https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Maritime_ports_freight_and_passenger_statistics#Rotterdam.2C_Antwerpen_and_Hamburg_stayed_top_ports

Grand port maritime. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 11 janvier 2020 à 17:48. [Consulté le 10 juillet 2020]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Grand_port_maritime

Grand port maritime du Havre. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 8 juillet 2020 à 23:23. [Consulté le 10 juillet 2020]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Grand_port_maritime_du_Havre

HARATI-MOKHTARI, Abbas, WALL, Alan, BROOKS, Philip et WANG, Jin, 2007. Automatic Identification System (AIS): Data Reliability and Human Error Implications. *The Journal of Navigation*. Septembre 2007. Vol. 60, n° 3, p. 373-389. DOI [10.1017/S0373463307004298](https://doi.org/10.1017/S0373463307004298)

HAROPA, 2019. *Gestion des escales S-WiNG* [en ligne]. Novembre 2019. [Consulté le 24 juillet 2020]. Disponible à l'adresse : <https://www.haropaports.com/sites/default/files/media/downloads/BROCHURE%20SWING%202019-FR-BD.pdf>

INTERNATIONAL CHAMBER OF SHIPPING, 2020. ICS | Shipping Facts. *Ice-shippin.org* [en ligne]. 2020. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://www.ics-shipping.org/shipping-facts/shipping-facts>

KALYVAS, Christos, KOKKOS, Athanasios et TZOURAMANIS, Theodoros, 2017. A survey of official online sources of high-quality free-of-charge geospatial data for maritime geographic information systems applications. *Information Systems*. 1 avril 2017. Vol. 65, p. 36-51. DOI [10.1016/j.is.2016.11.002](https://doi.org/10.1016/j.is.2016.11.002)

KAZEMI, Samira, ABGHARI, Shahrooz, LAVESSON, Niklas, JOHNSON, Henric et RYMAN, Peter, 2013. Open Data for Anomaly Detection in Maritime Surveillance. *Expert Systems with Applications*. 2013. Vol. 40, n° 14, p. 5719-5729. DOI [10.1016/j.eswa.2013.04.029](https://doi.org/10.1016/j.eswa.2013.04.029)

KITE, 2020. Python Web Scraping - Should I use Selenium, BeautifulSoup or Scrapy? [2020] [enregistrement vidéo]. *YouTube* [en ligne]. 29 février 2020. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://www.youtube.com/watch?v=zucvHSQsKHA>

KOSHY, Jacob, 2020. Best Programming Languages for Web Scraping | Promptcloud. *Promptcloud.com* [en ligne]. 26 juin 2020. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://www.promptcloud.com/blog/best-programming-language-for-web-scraping/>

LANDERS, Richard N., BRUSSO, Robert C., CAVANAUGH, Katelyn J. et COLLMUS, Andrew B., 2016. A primer on theory-driven web scraping: Automatic extraction of big data from the Internet for use in psychological research. *Psychological Methods*. 2016. Vol. 21, n° 4, p. 475-492. DOI [10.1037/met0000081](https://doi.org/10.1037/met0000081)

LEI, Po-Ruey, 2020. Mining maritime traffic conflict trajectories from a massive AIS data. *Knowledge and Information Systems*. 1 janvier 2020. Vol. 62, n° 1, p. 259-285. DOI [10.1007/s10115-019-01355-0](https://doi.org/10.1007/s10115-019-01355-0). [Consulté le 5 août 2020]. Disponible à l'adresse : <https://link.springer.com/article/10.1007/s10115-019-01355-0>

LIOUDIS, Nick K., 2020. Commodities Trading: An Overview. *Investopedia.com* [en ligne]. 8 mai 2020. [Consulté le 4 août 2020]. Disponible à l'adresse : <https://www.investopedia.com/investing/commodities-trading-overview/>

LORANGER, Hoa, 2014. Infinite Scrolling is Not for Every Website. *Nielsen Norman Group* [en ligne]. 2 février 2014. [Consulté le 23 juillet 2020]. Disponible à l'adresse : <https://www.nngroup.com/articles/infinite-scrolling/>

LY, Hélène et MARTIN, Jean-Philippe, 2018. L'activité des ports maritimes français repart en 2017. *Commissariat général au développement durable* [en ligne]. décembre 2018. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://www.statistiques.developpement-durable.gouv.fr/sites/default/files/2018-12/datalab-47-activite-des-ports-maritimes-francais-repart-en-2017-decembre2018.pdf>

M2R MARITIME, 2018. Les grands armateurs mondiaux. *M2R* [en ligne]. 16 février 2018. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://m2rmaritime.com/7997-2/>

MALIK, Sanjay Kumar et RIZVI, Sam, 2011. Information Extraction Using Web Usage Mining, Web Scraping and Semantic Annotation. In : *2011 International Conference on Computational Intelligence and Communication Networks*. Gwalior : s.n. octobre 2011, pp. 465-469.

MANAADIAR, Hariesh, 2019. Difference between a shipbroker and ship charterer. *Shipping and Freight Resource* [en ligne]. 26 février 2019. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://www.shippingandfreightresource.com/shipbroker-and-ship-charterer>

MANSELL, John N. K., 2009. Flag State Responsibility: Historical Development and Contemporary Issues. S.I. : Springer Science & Business Media. ISBN 978-3-540-92933-8.

MAŠOVIĆ, Matti, 2019. Port Call Efficiency Optimization, Using Data Analysis, Process Mining and Discrete Event Simulation [en ligne]. Delft : Delft University of Technology. Master of Science. [Consulté le 25 juin 2020]. Disponible à l'adresse : <https://pdfs.semanticscholar.org/7b17/c8e11ea1ccf06581249844ecef2a6c19de2d.pdf>

MAZZARELLA, Fabio, ALESSANDRINI, Alfredo, GREIDANUS, Harm, ALVAREZ, Marlene, ARGENTIERI, Pietro, NAPPO, Domenico et ZIEMBA, Lukasz, 2013. Data Fusion for Wide-Area Maritime Surveillance. *COST MOVE Workshop on Moving Objects at Sea* [en ligne]. Brest : s.n. 27 juin 2013. [Consulté le 5 août 2020]. Disponible à l'adresse : https://www.researchgate.net/publication/259772622_Data_Fusion_for_Wide-Area_Maritime_Surveillance

MEHTA, Kunal, et al., 2020. A Comparative Study of Various Approaches to Adaptive Web Scraping. In : KUMAR, Amit, PAPRZYCKI, Marcin et GUNJAN, Vinit Kumar (éd.), *ICDSMLA 2019*. Singapore : Springer. 2020. pp. 1245-1256.

MYERS, Daniel et MCGUFFEE, James, 2015. Choosing Scrapy. *Journal of Computing Sciences in Colleges* [en ligne]. 1 octobre 2015. Vol. 31, p. 83-89. [Consulté le 3 août 2020]. Disponible à l'adresse : https://www.researchgate.net/publication/314179276_Choosing_Scrapy

Navire de charge. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 19 mars 2020 à 12:02. [Consulté le 14 juillet 2020]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Navire_de_charge

Niedersachsen Ports. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 8 mai 2020 à 01:14. [Consulté le 22 juillet 2020]. Disponible à l'adresse : https://de.wikipedia.org/wiki/Niedersachsen_Ports

ORGANISATION MARITIME INTERNATIONALE, 2020a. Automatic Identification Systems (AIS). *imo.org* [en ligne]. [Consulté le 3 juillet 2020]. Disponible à l'adresse : <http://www.imo.org/fr/OurWork/Safety/Navigation/Pages/AIS.aspx>

ORGANISATION MARITIME INTERNATIONALE, 2020b. Systèmes de numéros OMI d'identification des navires. *imo.org* [en ligne]. [Consulté le 25 juin 2020]. Disponible à l'adresse : <http://www.imo.org/fr/OurWork/MSAS/Pages/IMO-identification-number-scheme.aspx>

PARSE HUB, 2020. Web Scraping VS Data Mining : What's the Difference? [enregistrement vidéo]. *YouTube* [en ligne]. 16 mars 2020. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://www.youtube.com/watch?v=ssUeAe30UTI>

PARSERS, 2020. US court fully legalized website scraping and technically prohibited it. *Parsers.me* [en ligne]. 28 janvier 2020. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://parsers.me/us-court-fully-legalized-website-scraping-and-technically-prohibited-it/>.

PATEL, Hiren, 2020. Web Scraping vs Web Crawling: What's the Difference? *DZone.com* [en ligne]. 16 mars 2020. [Consulté le 3 août 2020]. Disponible à l'adresse : <https://dzone.com/articles/web-scraping-vs-web-crawling-whats-the-difference>

Port of Aarhus. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 20 mai 2020 à 09:23. [Consulté le 10 juillet 2020]. Disponible à l'adresse : https://en.wikipedia.org/wiki/Port_of_Aarhus

PORT OF ANTWERP, 2019. Maritime services. *Portofantwerp.com* [en ligne]. 2019. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://www.portofantwerp.com/en/maritime-services>

Port of Copenhagen. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 24 avril 2020 à 12:44. [Consulté le 10 juillet 2020]. Disponible à l'adresse : https://en.wikipedia.org/wiki/Port_of_Copenhagen

PORT OF HAMBURG, 2020. Calls at the Port of Hamburg. In : The official website of the Port of Hamburg. [en ligne]. 2020. [Consulté le 29 juillet 2020]. Disponible à l'adresse : <https://www.hafen-hamburg.de/en/calls-at-the-port-of-hamburg>

PORT OF KLAIPEDA, 2020a. Port statistics | portofklaipeda.lt. *portofklaipeda.lt* [en ligne]. 2020. [Consulté le 10 juillet 2020]. Disponible à l'adresse : <https://www.portofklaipeda.lt/port-statistics>

PORT OF KLAIPEDA, 2020b. Klaipeda Seaport is approaching the 50 mln tons of cargo handling throughput. *portofklaipeda.lt* [en ligne]. [Consulté le 22 juillet 2020]. Disponible à l'adresse : <https://www.portofklaipeda.lt/news/14897/569/>

Porte-conteneurs. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 12 juillet 2020 à 22:03. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://fr.wikipedia.org/wiki/Porte-conteneurs>

SABY, Mathieu, 2018. Nettoyer et préparer des données avec OpenRefine : atelier pour les journées du consortium MASA - 14/11/2018 [en ligne]. [Consulté le 16 juillet 2020]. Disponible à l'adresse : <https://msaby.gitlab.io/atelier-openrefine-MASA/>

SCRAPINGHUB, 2017. Scraping JavaScript pages with Scrapy and Splash [enregistrement vidéo]. *YouTube* [en ligne]. 30 mai 2017. [Consulté le 3 août 2020]. Disponible à l'adresse : <https://www.youtube.com/watch?v=VvFC93vAB7U>

SCRAPY, 2020a. Scrapy | A Fast and Powerful Scraping and Web Crawling Framework. *scrapy.org* [en ligne]. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://scrapy.org/>

SCRAPY, 2020b. Frequently Asked Questions — Scrapy 2.2.0 documentation. *scrapy.org* [en ligne]. [Consulté le 4 juillet 2020]. Disponible à l'adresse : <https://docs.scrapy.org/en/latest/faq.html>

SERRY, Arnaud et LÉVÊQUE, Laurent, 2015. Le système d'identification automatique (AIS). *Netcom* [En ligne], 2015. 29 ½, pp 177-202. [Consulté le 03 juillet 2020]. Disponible à l'adresse : <http://journals.openedition.org/netcom/1943>

SERVICE DES DONNÉES ET ÉTUDES STATISTIQUES, 2019. Le transport maritime de marchandises au 3ème trimestre 2018. *Ministère de la transition écologique - données et études statistiques* [en ligne]. Janvier 2019. [Consulté le 4 août 2020]. Disponible à l'adresse : <https://www.statistiques.developpement-durable.gouv.fr/publicationweb/158>

SHIPTRACKS, 2019. MMSI, IMO and Call Sign - What is the difference? *ShipTracks - Live AIS Tracking* [en ligne]. 1 juillet 2019. [Consulté le 25 juin 2020]. Disponible à l'adresse : <https://shiptracks.com/updates/2019/07/01/mmsi-imo-and-call-sign-what-is-the-difference/>

STATISTICS DENMARK, 2020. SKIB72: Throughput of goods in major Danish seaports by seaport, direction and type of goods. *statbank.dk* [en ligne]. 11 juin 2020. [Consulté le 22 juillet]. Disponible à l'adresse : www.statbank.dk/SKIB72

SWISS TRADING AND SHIPPING ASSOCIATION, 2020. STSA :: Shipping: the backbone of international trade. *stsa.swiss* [en ligne]. 2020. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://stsa.swiss/knowledge/shipping>

TECH WITH TIM, 2020. Python Selenium Tutorial #1 - Web Scraping, Bots & Testing [enregistrement vidéo]. *YouTube* [en ligne]. 26 avril 2020. [Consulté le 3 août 2020]. Disponible à l'adresse : <https://www.youtube.com/watch?v=Xjv1sY630Uc>

TRAFIGURA, 2020. Que sont les matières premières physiques. *Commoditiesdemystified.info* [en ligne]. 2020. [Consulté le 4 août 2020]. Disponible à l'adresse : <https://www.commoditiesdemystified.info/en/>

Transport maritime. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 14 juillet 2020 à 13:41. [Consulté le 14 juillet 2020]. Disponible à l'adresse : https://fr.wikipedia.org/wiki/Transport_maritime

TUPPER, C. Eric, 2013. Ship Types. In : Introduction to Naval Architecture (Fifth Edition). Butterworth-Heinemann, 2013, pp. 379–412. ISBN 9780080982373

U.S. COAST GUARD NAVIGATION CENTER, 2020. Maritime Mobile Service Identity. *navcen.uscg.gov* [en ligne]. [Consulté le 25 juin 2020]. Disponible à l'adresse : <https://www.navcen.uscg.gov/index.php?pageName=mtMmsi>

VARGIU, Eloisa et URRU, Mirko, 2012. Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artificial Intelligence Research* [en ligne]. 5 décembre 2012. Vol. 2, n° 1, p. 44. [Consulté le 3 août 2020]. Disponible à l'adresse : <http://www.sciedu.ca/journal/index.php/air/article/view/1390/1115>

VELDHUIS, H D, 2015. *Developing an automated solution for ETA definition concerning long distance shipping* [en ligne]. University of Twente. [Consulté le 4 août 2020]. Disponible à l'adresse : <http://pdfs.semanticscholar.org/c1de/b22d49d2eb18f2f49be45eb44b9a22a7af81.pdf>

WATSON, Richard, HOLM, Henrik et LIND, Mikael, 2015. Green Steaming: A Methodology for Estimating Carbon Emissions Avoided. *International Conference on Information Systems* [en ligne]. 13 décembre 2015. [Consulté le 4 août 2020]. Disponible à l'adresse : https://www.researchgate.net/publication/326040796_Green_Steaming_A_Methodology_for_Estimating_Carbon_Emissions_Avoided

Wilhelmshavener Häfen. *Wikipédia : l'encyclopédie libre* [en ligne]. Dernière modification de la page le 19 février 2020 à 15:15. [Consulté le 10 juillet 2020]. Disponible à l'adresse : https://de.wikipedia.org/w/index.php?title=Wilhelmshavener_H%C3%A4fen&oldid=196959262.

X-BYTE ENTERPRISE CRAWLING, 2019. Top 5 programming Languages for Web Scraping. *Medium.com* [en ligne]. 7 août 2019. [Consulté le 9 juillet 2020]. Disponible à l'adresse : <https://medium.com/@xbytecrawling/top-5-programming-languages-for-web-scraping-e592ce2192e4>

YANG, Dong, WU, Lingxiao, WANG, Shuaian, JIA, Haiying et LI, Kevin X., 2019. How big data enriches maritime research – a critical review of Automatic Identification System (AIS) data applications. *Transport Reviews*. 2 novembre 2019. Vol. 39, n° 6, p. 755-773. DOI [10.1080/01441647.2019.1649315](https://doi.org/10.1080/01441647.2019.1649315)

ZHU, Feixiang, 2011. Mining ship spatial trajectory patterns from AIS database for maritime surveillance. *2011 2nd IEEE International Conference on Emergency Management and Management Sciences*. S.l. : s.n. août 2011. p. 772-775. DOI [10.1109/ICEMMS.2011.6015796](https://doi.org/10.1109/ICEMMS.2011.6015796)

Annexe 1 : Évolution du trafic maritime mondial



Tableau 1.3 Évolution du trafic maritime international, diverses années
(En millions de tonnes chargées)

Année	Navires-citernes ^a	Principaux vracs ^b	Autres marchandises solides ^c	Total (toutes catégories)
1970	1 440	448	717	2 605
1980	1 871	608	1 225	3 704
1990	1 755	988	1 265	4 008
2000	2 163	1 186	2 635	5 984
2005	2 422	1 579	3 108	7 109
2006	2 698	1 676	3 328	7 702
2007	2 747	1 811	3 478	8 036
2008	2 742	1 911	3 578	8 231
2009	2 641	1 998	3 218	7 857
2010	2 752	2 232	3 423	8 408
2011	2 785	2 364	3 626	8 775
2012	2 840	2 564	3 791	9 195
2013	2 828	2 734	3 951	9 513
2014	2 825	2 964	4 054	9 842
2015	2 932	2 930	4 161	10 023
2016	3 058	3 009	4 228	10 295
2017	3 146	3 151	4 419	10 716
2018	3 194	3 210	4 601	11 005

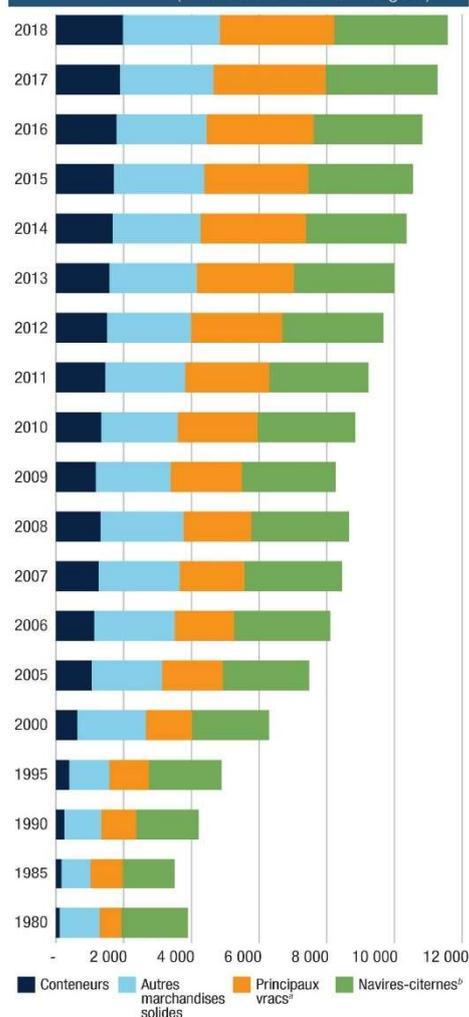
Sources : Tableau établi par le secrétariat de la CNUCED, d'après les données des pays déclarants, telles que publiées sur les sites Web des gouvernements et de l'industrie portuaire, et des sources spécialisées. Les données sur les vracs secs pour 2006 et les années suivantes ont été révisées et actualisées pour tenir compte de l'amélioration des rapports, avec des chiffres plus récents et de meilleurs renseignements concernant la ventilation par type de marchandises. Depuis 2006, la ventilation des vracs secs en principaux vracs et autres marchandises solides se base sur divers numéros de la publication *Shipping Review and Outlook* de Clarksons Research. Les totaux estimés concernant les chiffres du trafic maritime pour 2018 sont basés sur des données préliminaires ou sur l'année la plus récente pour laquelle des données étaient disponibles.

^a Pétrole brut, produits pétroliers raffinés, gaz et produits chimiques.

^b Minerai de fer, céréales, charbon, bauxite/alumine et phosphate. À partir de 2006, les principaux vracs incluent des données sur le minerai de fer, les céréales et le charbon uniquement. Les données concernant la bauxite/l'alumine et le phosphate se trouvent dans la catégorie « autres marchandises solides ».

^c Vrac de moindre importance, marchandises conteneurisées et marchandises diverses.

Graphique 1.1 Trafic maritime international, par type de marchandises, diverses années
(En millions de tonnes chargées)



Sources : CNUCED, *Étude sur les transports maritimes, diverses années*. Pour 2006-2018, la ventilation par type de marchandise est basée sur les données provenant de Clarksons Research, 2019a, *Shipping Review and Outlook*, printemps.

Note : Les chiffres des principaux vracs pour 1980-2005 intègrent le minerai de fer, les céréales, le charbon, la bauxite/l'alumine et le phosphate. À partir de 2006, les principaux vracs sont uniquement le minerai de fer, les céréales et le charbon. Les données concernant la bauxite/l'alumine et le phosphate se trouvent dans la catégorie « autres marchandises solides ».

^a Minerai de fer, céréales, charbon, bauxite/alumine et phosphate. À partir de 2006, les principaux vracs sont uniquement le minerai de fer, les céréales et le charbon. Les données concernant la bauxite/l'alumine et le phosphate se trouvent dans la catégorie « autres marchandises solides ».

^b Pétrole brut, produits pétroliers raffinés, gaz et produits chimiques.

(Conférence des Nations Unies sur le commerce et le développement, 2020)

Annexe 2 : Catégorisation des navires de commerce

A Categorisation of ships by type - Cargo carrying ships (STATCODE 5) - For Statistical Purposes

Level 5	Level 4	Level 3	Level 2	Level 1		
LNG Tanker	LNG Tanker	Liquefied Gas	Tankers	Cargo Carrying		
LPG Tanker	LPG Tanker					
LPG/Chemical Tanker	LPG Tanker					
CO2 Tanker	CO2 Tanker					
Molten Sulphur Tanker	Chemical Tanker	Chemical				
Chemical Tanker						
Chemical/Products Tanker					Chemical/Oil Products Tanker	
Wine Tanker					Wine Tanker	
Vegetable Oil Tanker					Vegetable Oil Tanker	
Edible Oil Tanker					Edible Oil Tanker	
Beer Tanker					Beer Tanker	
Latex Tanker					Latex Tanker	
Fruit Juice Tanker					Fruit Juice Tanker	
Shuttle Tanker					Crude Oil Tanker	Oil
Crude Oil Tanker						
Crude/Oil Products Tanker						
Products Tanker	Oil Products Tanker					
Tanker (unspecified)						
Asphalt/Bitumen Tanker	Bitumen Tanker					
Coal/Oil Mixture Tanker	Coal/Oil Mixture Tanker					
Water Tanker	Water Tanker	Other Liquids				
Molasses Tanker						
Glue Tanker	Glue Tanker					
Alcohol Tanker	Alcohol Tanker					
Caprolactam Tanker	Caprolactam Tanker					
Bulk Carrier	Bulk Carrier	Bulk Dry	Bulk Carriers			
Bulk Carrier, Laker Only						
Bulk Carrier (with Vehicle Decks)						
Ore Carrier	Ore Carrier	Bulk Dry/Oil				
Bulk/Oil Carrier (OBO)	Bulk/Oil Carrier					
Ore/Oil Carrier	Ore/Oil Carrier	Self Discharging Bulk Dry				
Bulk Cargo Carrier, self discharging	Self Discharging Bulk Carrier					
Bulk Cargo Carrier, self discharging, Laker		Other Bulk Dry				
Cement Carrier	Cement Carrier					
Wood Chips Carrier, self unloading	Wood Chips Carrier					
Urea Carrier	Urea Carrier					
Aggregates Carrier	Aggregates Carrier					
Limestone Carrier	Limestone Carrier					
Refined Sugar Carrier	Refined Sugar Carrier					
Powder Carrier	Powder Carrier					
				Ship Structures		

A Categorisation of ships by type - Cargo carrying ships (STATCODE 5) - For Statistical Purposes

Level 5	Level 4	Level 3	Level 2	Level 1		
General Cargo Ship (with Ro-Ro facility)	General Cargo Ship	General Cargo	Dry/Cargo/Passenger	Cargo Carrying	Ship Structures	
General Cargo, Self-discharging						
Open Hatch cargo Ship						
General Cargo/Tanker (Container/oil/bulk - COB ship)						
General Cargo/Tanker						
General Cargo Ship						
Palletised Cargo Ship	Palletised Cargo Ship					
Deck Cargo Ship	Deck Cargo Ship					
General Cargo/ Passenger Ship	Passenger/ General Cargo Ship	Passenger/General Cargo Ship				
Container Ship (Fully Cellular)	Container Ship	Container				
Container Ship (Fully Cellular with Ro-Ro Facility)						
Passenger/ Container Ship						Passenger/ Container Ship
Refrigerated Cargo Ship	Refrigerated Cargo Ship	Refrigerated Cargo Ship				
Ro-Ro Cargo Ship	Ro-Ro Cargo Ship	Ro-Ro Cargo				
Rail Vehicles Carrier						
Vehicles Carrier						Vehicles Carrier
Container/ Ro-Ro Cargo Ship						Container/ Ro-Ro Cargo Ship
Landing Craft						Landing Craft
Passenger/Ro-Ro Ship (Vehicles)	Passenger/ Ro-Ro Cargo Ship	Passenger/ Ro-Ro Cargo				
Passenger/Ro-Ro Ship (Vehicles/Rail)						
Passenger/Landing Craft	Passenger/Landing Craft					
Passenger/Cruise	Passenger (Cruise) Ship	Passenger				
Passenger Ship						
Livestock Carrier	Livestock Carrier	Other Dry Cargo				
Barge Carrier						
Heavy Load Carrier						
Heavy Load Carrier, semi submersible						
Yacht Carrier, semi submersible						
Nuclear Fuel Carrier	Nuclear Fuel Carrier					
Nuclear Fuel Carrier (with Ro-Ro facility)						

IHS MARKIT, 2019. *IHS Maritime & Trade, World Fleet Statistics 2019* [en ligne]. IHS Markit. [Consulté le 14 juillet 2020]. Disponible à l'adresse : <https://cdn.ihs.com/www/prot/pdf/0719/WorldFleetStatistics2018Report-LoRes.pdf>

Annexe 3 : Autorités portuaires nationales identifiées

Pays	Entité	Lien
Allemagne	Federal Association of German Seaport Operators	https://www.zds-seehaefen.de/
Belgique	Non identifiée	
Bulgarie	Bulgarian Ports Infrastructure Company	http://www.bgports.bg/en/
Croatie	Croatian Port Authority / Ministry of the sea, transport and infrastructure	http://mppi.hr/default.aspx?id=666
Chypre	Cyprus Ports Authority	https://www.cpa.gov.cy/CPA/page.php?pageID=1&langID=13
Danemark	Danish Ports	https://www.danskehavne.dk/
Espagne	Puertos del Estado	http://www.puertos.es/en-us
Estonie	Non identifiée	
Finlande	Finnish Port Association	http://www.finnishports.fi/eng/
France	Union des ports de France	http://www.port.fr/
Grèce	Hellenic Ports Association	https://www.elime.gr/
Irlande	Irish Ports Association	?
Italie	Associazione Porti Italiani - Assoporti	https://www.assoporti.it/en/home/
Lettonie	Non identifiée	
Lituanie	Non identifiée	
Malte	Authority for Transport Malta	https://www.transport.gov.mt/maritime-40
Norvège	Norwegian Ports Association	https://www.samfunnsbedriftene.no/bransjer/norske-havner/
Pays-Bas	Non identifiée	
Pologne	Non identifiée	-
Portugal	Association Ports of Portugal (APP)	http://www.portosdeportugal.pt/indexG.php
Roumanie	Non identifiée	
Royaume-Uni	British Ports Association	https://www.britishports.org.uk/
Slovenie	Non identifiée	
Suède	Ports of Sweden	?
Turquie	Port Operators Association of Turkey	http://www.turklim.org/en/
Europe	European Sea Ports Organisation	https://www.espo.be/

Annexe 4 : Sources web, sites ports commerciaux au 24.04.2020

Sur 190 sites web répertoriés, on en dénombre 103 qui proposent des données exploitables (~54%). Par données exploitables on entend des données qu'il est possible de collecter via des outils de *web scraping*. Ces données sont explicitées dans la colonne « Typologie des données » du tableau concernant le détail des sources web.

Cette liste de sites web a été réalisée selon la méthodologie décrite au chapitre 3.1.1. :

- Recherche d'information chez les autorités portuaires nationales des pays européens.
- Recherche d'information dans des sites web spécialisés.
- Échantillonnage des ports commerciaux et création d'une liste en prenant compte en premier lieu de l'importance du trafic maritime puis de la disponibilité de données exploitables. Les principaux ports européens sont ainsi mentionnés même s'ils ne proposent pas de données à collecter.

Index récapitulatif

Index sources web ports Europe

N°	Pays	Port	Données exploitables
1	Allemagne	Bremenports	Oui
2	Allemagne	Hambourg	Oui
	Allemagne	Kiel	Non
3	Allemagne	Lübeck	Oui
	Allemagne	Mukran	Non
4	Allemagne	Nierdersachsen	Oui
5	Allemagne	Rostock	Oui
	Allemagne	Stralsund	Non
	Allemagne	Wismar	Non
	Belgique	Anvers	Non
6	Belgique	Bruges - Zeebruges	Oui
7	Belgique	Ostende	Oui
	Bulgarie	Balchik	Non
	Bulgarie	Burgas	Non
	Bulgarie	Varna	Non
	Croatie	Ploce	Non
8	Croatie	Rijeka	Oui
	Croatie	Zadar	Non
	Chypre	Limassol	Non
	Chypre	Larnaca	Non
	Chypre	Vassiliko	Non
9	Danemark	Aabenraa	Oui
10	Danemark	Aalborg	Oui
11	Danemark	Aarhus	Oui
12	Danemark	Copenhague	Oui
13	Danemark	Esbjerg	Oui
14	Danemark	Fredericia + Nyborg	Oui

15	Danemark	Frederikshavn	Oui
	Danemark	Grenaa	Non
	Danemark	Koge	Non
	Danemark	Odense	Non
	Espagne	Algeciras	Non
	Espagne	Alicante	Non
	Espagne	Almeria	Non
16	Espagne	Avilès	Oui
	Espagne	Barcelona	Non
17	Espagne	Bilbao	Oui
18	Espagne	Cadiz Bay	Oui
19	Espagne	Cartagena	Oui
	Espagne	Castello	Non
20	Espagne	Gijón	Oui
	Espagne	Huelva	Non
21	Espagne	Las Palmas (+ Puerto del Rosario, Arrecife)	Oui
22	Espagne	Malaga	Oui
23	Espagne	Marin	Oui
	Espagne	Melilla	Non
	Espagne	Motril	Non
24	Espagne	Pasaia	Oui
25	Espagne	Santander	Oui
26	Espagne	Seville	Oui
27	Espagne	Tarragona	Oui
28	Espagne	Tenerife	Oui
29	Espagne	Vigo	Oui
30	Espagne	Valencia	Oui
31	Espagne	Vilagarcía	Oui
	Estonie	Kunda	Non
32	Estonie	Tallinn	Oui
33	Finlande	Hanko	Oui
34	Finlande	HaminaKotka	Oui
35	Finlande	Helsinki	Oui
36	Finlande	Inkoo	Oui
37	Finlande	Jakobstad & Pietarsaari	Oui
	Finlande	Joensuu	Non
38	Finlande	Kemi	Oui
39	Finlande	Kokkola	Oui
40	Finlande	Naantali	Oui
41	Finlande	Oulu	Oui
42	Finlande	Rauma	Oui
	Finlande	Sköldvik	Non
43	Finlande	Turku	Oui
	France	Ajaccio	Non
	France	Bastia	Non
	France	Bayonne	Non
44	France	Bordeaux	Oui

	France	Boulogne	Non
	France	Brest	Non
	France	Caen	Non
	France	Calais	Non
	France	Cherbourg	Non
45	France	Dunkerque	Oui
	France	La Réunion	Non
46	France	La Rochelle	Oui
47	France	Le Havre	Oui
48	France	Lorient	Oui
	France	Marseille	Non
49	France	Nantes-Saint-Nazaire	Oui
50	France	Nouméa (Nouvelle-Calédonie)	Oui
	France	Pointe-à-Pitre (Guadeloupe)	Non
	France	Port-la-Nouvelle	Non
51	France	Rouen	Oui
	France	Saint-Malo	Non
	France	Sète	Non
	France	Toulon	Non
	Grèce	Heraklion	Non
	Grèce	Igoumenitsa	Non
	Grèce	Patras	Non
	Grèce	Piraeus	Non
52	Grèce	Thessaloniki	Oui
	Grèce	Volos	Non
53	Irlande	Cork	Oui
	Irlande	Drogheda	Non
54	Irlande	Dublin	Oui
55	Irlande	Galway	Oui
	Irlande	New Ross	Non
56	Irlande	Waterford	Oui
	Italie	Ancona	Non
	Italie	Augusta (+ Catania)	Non
57	Italie	Bari	Oui
	Italie	Civitavecchia (+ Fiumicino, Gaeta)	Non
	Italie	Gênes (+ Vado, Savona, Pra')	Non
	Italie	Cagliari	Non
	Italie	Gioia Tauro	Non
	Italie	Livourne (+ Capraia Isola, Piombino, Portoferraio, Rio Marina, Cavo)	Non
	Italie	Messina (+ Milazzo)	Non
	Italie	Napoli (+ Salerno, Castellammare di Stabia)	Non
	Italie	Palermo	Non
	Italie	Ravenna	Non
58	Italie	Trieste	Oui
	Italie	Tarente	Non
	Italie	Venise	Non

	Lettonie	Liepaja	Non
	Lettonie	Riga	Non
	Lettonie	Ventspils	Non
59	Lituanie	Klaipeda	Oui
	Malte	Valletta	Non
	Malte	Malta Freeport	Non
60	Norvege	Bergen	Oui
61	Norvege	Drammen	Oui
62	Norvege	Karmsund	Oui
63	Norvege	Narvik	Oui
64	Norvege	Oslo	Oui
65	Norvege	Stavanger (+Risavika, Mekjarvik, Cruise)	Oui
66	Norvege	Tromsø	Oui
67	Pays-Bas	Amsterdam	Oui
68	Pays-Bas	Aruba	Oui
	Pays-Bas	Groningen	Non
	Pays-Bas	Moerdijk	Non
69	Pays-Bas	North Sea Ports	Oui
70	Pays-Bas	Rotterdam	Oui
71	Pologne	Gdansk	Oui
72	Pologne	Gdynia	Oui
73	Pologne	Świnoujście	Oui
74	Pologne	Szczecin	Oui
75	Portugal	Açores	Oui
76	Portugal	Aveiro	Oui
	Portugal	Faro	Non
77	Portugal	Figueira de Foz	Oui
78	Portugal	Leixoes	Oui
79	Portugal	Lisbonne	Oui
80	Portugal	Madère	Oui
81	Portugal	Setubal	Oui
82	Portugal	Sines	Oui
83	Portugal	Viana do Castelo	Oui
	Roumanie	Constanța	Non
84	Royaume-Uni	Aberdeen	Oui
85	Royaume-Uni	Belfast	Oui
86	Royaume-Uni	Blyth	Oui
87	Royaume-Uni	Bristol	Oui
	Royaume-Uni	Douvres	Non
88	Royaume-Uni	Felixtowe	Oui
	Royaume-Uni	Forth	Non
	Royaume-Uni	Glasgow	Non
89	Royaume-Uni	Grimsby / Immingham / Hull / Goole	Oui
	Royaume-Uni	Liverpool	Non
90	Royaume-Uni	Londres	Oui
91	Royaume-Uni	Milford Haven	Oui
92	Royaume-Uni	Southampton	Oui

93	Royaume-Uni	Southwales Ports (Barry, Cardiff, Newport, Port Talbot, Swansea)	Oui
94	Royaume-Uni	Tees / Hartlepool	Oui
95	Royaume-Uni	Tyne	Oui
96	Slovenie	Koper	Oui
97	Suède	Gavle	Oui
98	Suède	Goteborg	Oui
	Suède	Falkenbergs	Non
99	Suède	Helsingborg	Oui
	Suède	Kalmar	Non
	Suède	Karlskrona	Non
100	Suède	Malmö	Oui
101	Suède	Norrköping	Oui
	Suède	Oxelösunds	Non
102	Suède	Stockholm (+ Nynäshamn, Kapellskär, Cruise calls)	Oui
103	Suède	Wallhamn	Oui
	Turquie	Atakas	Non
	Turquie	Bandirma	Non
	Turquie	Belde	Non
	Turquie	Canakkale	Non
	Turquie	Gemlik	Non
	Turquie	Istanbul	Non
	Turquie	Kumport	Non
	Turquie	Limar	Non

Détails

N°	Pays	Nom du port	Données disponibles	Typologie des données	Lien
1	Allemagne	Bremenports (Bremen, Bremerhaven)	Ships in Port	Name, Port of Origin, Berth	https://bremenports.de/en/ships-in-port/ships-tracker/
2	Allemagne	Hambourg	Expected Vessels	Shipname, IMO, Shipstyp, TEU, Length, ETA, Berth + DETAILS (Technical data : IMO, Vesseltype, Built, GRT, DWT, Container, Length, Breadth, Draught)	https://www.hafen-hamburg.de/en/vessels/eta
	Allemagne	Kiel	Aucune données exploitables		https://www.portofkiel.com/home-en.html
3	Allemagne	Lübeck (Lübecker Hafengesellschaft)	Current arrivals	Name, Company, From, Time of Arrival, Terminal, Berth, Status	https://www.lhg.com/index.php?id=192&L=1
			Current departure	Name, Company, To, Time of Sailing, Terminal, Berth, Status	
	Allemagne	Mukran	Aucune données exploitables		https://www.mukran-port.de/home.html
4	Allemagne	Niedersachsen Ports	Ship arrivals and Departures	Status, Schiffsart, Ankunft/Abfahrt, Rufzeichen/Schiffsname, Flagge, BRZ Abfahrtshafen, Zielhafen, Kai	https://www.nports.de/en/ship-arrivals-and-departures/
5	Allemagne	Rostock	Ships in the Port	BT, Ship, Arrival, Import / Export, From / To + DETAILS (redirection sur fleetmon.com)	https://www.rostock-port.de/en/shipping/ships-in-the-port.html
			Ships Arrivals	BT, Ship, Arrival, Import / Export, From / To + DETAILS (redirection sur fleetmon.com)	https://www.rostock-port.de/en/shipping/ship-arrivals.html
	Allemagne	Stralsund	Aucune données exploitables		https://www.seehafen-stralsund.de/en/
	Allemagne	Wismar	Aucune données exploitables		https://www.hafen-wismar.de/de
	Belgique	Anvers	Aucune données exploitables		https://www.portofantwerp.com/en/maritime-services
6	Belgique	Bruges - Zeebruges	Ship arrivals	Arrival date/time, Departure date/time, Ship name, Location, Lloyds	https://portofzeebrugge.be/en/arrivals-timetable
		Gand (North Sea Port Gent)	Voir port n°69		
7	Belgique	Ostende	Expected arrivals	Ship, Ship no, LOA, Dr., GT, Flag, Agent, Berth, Time, Port	https://ensor.be/Site/VerwachteOpvaarten.aspx?lang=en
			Expected departures	Ship, Ship no, LOA, Dr., GT, Flag, Agent, Berth, Time, Destination	https://ensor.be/Site/RecenteAfvaarten.aspx?lang=en
	Bulgarie	Balchik	Aucune données exploitables		http://www.port-balchik.com/bg/

	Bulgarie	Burgas	Aucune données exploitables		https://port-burgas.bg/
	Bulgarie	Varna	Aucune données exploitables		https://www.port-varna.bg/
	Croatie	Ploce	Aucune données exploitables		https://www.luka-ploce.hr/?lang=en
8	Croatie	Rijeka	Arrivals	Vessel, Arrival Time, Berth, Agent, Status + DETAILS (Length, Flag, Vessel type, IMO-number, Signal)	https://www.portauthority.hr/en/arrivals-and-departures/
			Departures	Vessel, Departure Time, Berth, Agent, Status + DETAILS (Length, Flag, Vessel type, IMO-number, Signal)	
	Croatie	Zadar	Aucune données exploitables		https://www.luka-zadar.hr/en
	Chypre	Limassol	Aucune données exploitables		https://www.cpa.gov.cy/CPA/page.php?pageID=52&mpath=/12/23/24
	Chypre	Larnaca	Aucune données exploitables		https://www.cpa.gov.cy/CPA/page.php?pageID=53&mpath=/12/23/26
	Chypre	Vassiliko	Aucune données exploitables		https://www.cpa.gov.cy/CPA/page.php?pageID=55&mpath=/12/23/80
9	Danemark	Aabenraa	List of ships	Quay, Facility, Shipname, IMO Number, Cargo, Recipient, ETA, Status, ETD, BT, LOA, DW	https://www.aabenraahavn.dk/en/maritime-information/list-of-ships/
10	Danemark	Aalborg	Ships in port	Vessel, Arrival, Departure, Dock, Agent, Status + DETAILS (Vessel length, Vessel width, GRT, Flag code, IMO, Vessel Type)	https://portofaalborg.com/maritime-information/ships-in-port/
				Arrival, Departure, Dock, StatusKey, StatusClass, VesselName, VesselLength, VesselWidth, VesselFlagCode, VesselImo, VesselWeight, VesselType, VesselAgent, SortDate	https://portofaalborg.com/umbraco/api/portinfo/getcalls?cultureCode=en-US
11	Danemark	Aarhus	Vessels in port	Pier, Shipsname, Shipstype, Exp. arrival, Status, Exp. departure	https://www.aarushavn.dk/en/ship_calls/vessels_in_port/
12	Danemark	Copenhagen	Port Calls	Arrival, ETA, Ship, From, Quay, Agency, Status + DETAILS (Arrival from, Incl. port, Cargo, Quantity, Loading recipient, Departure to, Loading port, Cargo, Quantity, Loading recipient, Call sign, IMO, Nationality, BT, Vessel type, Ship length, Passenger capacity, Captain, Call no., Main purpose, Quay name, Dangerous goods	http://www.cmport.com/ships-in-port/calls/copenhagen
13	Danemark	Esbjerg	Ships in port	Name/Call Sign, Date, Time, Berth, GT, DWT, BR, DR ARR., LOA	http://port-of-esbjerg.dk/en/ships/port
			Scheduled	Name/Call Sign, Date, ETA, Homeport, Arrives from, Berth, GT, BR, LOA	http://port-of-esbjerg.dk/en/ships/inbound
			Departed	Name/Call Sign, Date, Time, Homeport, Next port, GT, BR, DR ARR., DR DEP., LOA	http://port-of-esbjerg.dk/en/ships-port/departed
14	Danemark	Fredericia + Nyborg	Calls	Name, Quay, Status, ETA, Time, ETD, Time, Broker	https://www.adp-as.dk/en/maritime-services/calls/

15	Danemark	Frederikshavn	Arrivals and departures overview	Status, IMO, Nom, Date d'arrivée, Heure d'arrivée, But, Quai déchargement, Quai chargement, Quai autre, De, Date de départ prévue	https://complet.pof.dk:20001/webparts/anlob.aspx
	Danemark	Grenaa	Aucune données exploitables		https://www.port-of-grenaa.com/
	Danemark	Koge	Aucune données exploitables		https://koegehavn.dk/
	Danemark	Odense	Aucune données exploitables		https://www.lpo.dk/
	Espagne	Algeciras	Aucune données exploitables		https://www.apba.es/fr/
	Espagne	Alicante	Aucune données exploitables		https://www.puertoalicante.com/en/
	Espagne	Almeria	Aucune données exploitables		https://apalmeria.com/
16	Espagne	Avilès	Vessels in Port	State, Scale / Berth, Arrival / Departure, Vessel, Origin / Destination, Draft / LOA, Agent / Stevedore, Quay, Oper, Tonnage / Cargo	https://www.puertoaviles.es/en/servicioscomerciales/buquesenelpuerto.asp
	Espagne	Barcelona	Aucune données exploitables		http://www.portdebarcelona.cat/en/web/port-dels-negocis/contenedores
17	Espagne	Bilbao	Vessels Arrivals	Flag, Vessel, Home Port, Arrival	https://www.bilbaoport.eus/en/vessel-activity/vessel-arrivals/
			Vessels Departures	Flag, Vessel, Destination, Arrival	https://www.bilbaoport.eus/en/vessel-activity/vessel-departures/
			Vessels Operating	Flag, Vessel, GT100, Home Port, Arrival, Destination, Departure	https://www.bilbaoport.eus/en/vessel-activity/vessels-operating/
			Scheduled Calls	Flag, Vessel, Length, GT, Expected Arrival, Home Port, Destination	https://www.bilbaoport.eus/en/vessel-activity/scheduled-calls/
			Inactive Stay	Flag, Vessel, Length, GT 100, Situation	https://www.bilbaoport.eus/en/vessel-activity/inactive-stay/
18	Espagne	Cadiz Bay	Vessel arrivals	No information available in this section	https://www.puertocadiz.com/en/communication/vessel-movements-daily-bulletin/
			Vessels operating	Scale n°, Name, Flag, GT, Length, Draft, Origin, Destination, Dock, Bollard, Stevedore	
			Vessels departures		
			Inactive vessels		
			Inside operating vessels		
			Vessels for scraping	No information available in this section	
			Vessels in legal custody	Scale n°, Name, Flag, GT, Length, Draft, Origin, Destination, Dock, Bollard, Stevedore	
			Pilotage		
Tugs					
19	Espagne	Cartagena	En opération, Arrivée, Départ	Escale, Navire, Quai, Arrivée, Départ, Consignataire, Longueur, Type, Marchandise, Tn, Manutention	http://atraques.apc.es/

	Espagne	Castello	Aucune données exploitables		https://www.portcastello.com/en/
20	Espagne	Gijón	En opération, Arrivée, Départ	Escale, Navire, Quai, Arrivée, Départ, Consignataire, Longueur, Type, Marchandise, Tn, Manutention	http://gijon.posidoniaport.com/
	Espagne	Huelva	Aucune données exploitables		https://www.puertohuelva.com/es/
21	Espagne	Las Palmas (+ Puerto del Rosario, Arrecife)	En opération, Arrivée, Départ	Escale, Navire, Quai, Arrivée, Départ, Consignataire, Longueur, Type, Marchandise, Tn, Manutention	http://laspalmas.posidoniaport.com/
22	Espagne	Malaga	Berths Information	Ship / Flag, GT / Length, Consignee, Entry, Dock, Origin / Destination	https://www.puertomalaga.com/en/berths-information/
23	Espagne	Marin	Arrived Vessels	Vessel, Number, Ship agent, Quay, Bollard, Operation, ETD	http://www.apmarin.com/en_buques_puerto.php
			Expected Vessels	Vessel, Number, Ship agent, Quay, Bollard, Operation, ETA	http://www.apmarin.com/en_previ_buques.php
	Espagne	Melilla	Aucune données exploitables		https://www.puertodemelilla.es/
	Espagne	Motril	Aucune données exploitables		http://www.apmotril.com/
24	Espagne	Pasaia	Ships in port	Reference, Arrival, IMO, Ship, G.T., Length, Draught, Consignee, Stevedore, Loading/Unloading, Arrival Cargo, Arrival Tm., Departure Cargo, Departure Tm., Dock	https://sede.pasaiaport.gob.es/elPuertoHoy/BuquesEnPuerto.do?lang=en
			Arrivals	Reference, Arrival, IMO, Ship, G.T., Length, Draught, Cargo, Loading/Unloading, Tm., Consignee, Dock, Previous Port	https://sede.pasaiaport.gob.es/elPuertoHoy/Entradas.do?lang=en
			Departures	Reference, Arrival, Departure, IMO, Draught, Consignee, Next Port, Loading/Unloading, Cargo, Tm., Dock	https://sede.pasaiaport.gob.es/elPuertoHoy/Salidas.do?lang=en
			Announced Vessels	Reference, Arrival, IMO, Length, G.T., Consignee, Stevedore, Dock, Loading/Unloading, Arrival Cargo, Arrival Tm., Departure Cargo, Departure Tm.	https://sede.pasaiaport.gob.es/elPuertoHoy/BuquesAnunciados.do?lang=en
25	Espagne	Santander	Buques en Puerto, Entradas, Salidas	Registro, Buque y procedencia, Muelle, Flag, Inicio, Final, Consignatario	http://www.puertosantander.es/cas/buques_puerto.aspx
26	Espagne	Seville	Ships at port	Date, Vessel, IMO, Flag, Mooring, Consignee	http://portal.apsevilla.com/wps/portal/puerto_en/buques_en?WCM_GLOBAL_CONTEXT=/EN/puertosevilla/buquesEN
27	Espagne	Tarragona	En opération, Arrivée, Départ	Navire, Quai, Arrivée, Départ, Consignataire, Longueur, Type, Marchandise, Tn, Manutention	http://tarragona.posidoniaport.com/
28	Espagne	Tenerife	En opération, Arrivée, Départ	Navire, Quai, Arrivée, Départ, Consignataire, Longueur, Type, Marchandise, Tn, Manutention	http://gisweb.puertosdetenerife.org/
29	Espagne	Vigo	Diverses	Diverses	https://www.apvigo.es/elpuertohoy/
30	Espagne	Valencia	Port calls	Vessel, Port call, Status, Arrival, Departure, Port call port, UN/LOCODE, Container terminal, Vessel agent, Regular line	https://www.valenciaportpcs.net/portcalls/Search?lang=en
31	Espagne	Vilagarcía	Arrivals	Ship name, Flag, From/To, Arrival date, Departure date, Dock, Operation, Type of good, Ship agent	http://www.portovilagarcia.es/en/traffics/arrivals.html

			Departures	Ship name, Flag, From/To, Arrival date, Departure date, Dock, Operation, Type of good, Ship agent	http://www.portovilagarca.es/en/traffics/departures.html
			Ship in port	Ship name, Flag, From/To, Arrival date, Departure date, Dock, Operation, Type of good, Ship agent	http://www.portovilagarca.es/en/traffics/ships-in-port.html
	Estonie	Kunda	Aucune données exploitables		https://www.kundasadam.ee/welcome.html
32	Estonie	Tallinn	List of ships arriving to cargo harbours	Port, ETA, Nom Type, GT, L, B, Kai	https://veeb.ts.ee/cargo_ships_arrivals.php
33	Finlande	Hanko	Arrivée	Vessel, Nat., From/To, ETA, Berth, ETS, Arrival Agent & Departure Agent, Updated	https://portofhanko.fi/en/laivalistat/
			Au port	Vessel, Nat., From/To, Arrived, Berth, ETS, Arrival Agent & Departure Agent, Updated	
			Départ	Vessel, Nat., From/To, Arrived, Berth, Sailed, Arrival Agent & Departure Agent, Updated	
34	Finlande	HaminaKotka	Arrivals	ETA, Vessel, Nat., From / To, Berth, ETS, Agent of Arrival / Agent of Departure, Updated	https://www.haminakotka.com/transport/arrivals-and-departures
			At Berth	Arrived, Vessel, Nat., From / To, Berth, ETS, Terminal, Agent of Arrival / Agent of Departure, Updated	https://www.haminakotka.com/transport/arrivals-and-departures
			Departed	ATS, Vessel, Nat., From / To, Arrived, Berth, Agent of Arrival / Agent of Departure, Updated	https://www.haminakotka.com/transport/arrivals-and-departures
35	Finlande	Helsinki	Outbound Vessels and Vessels in Port	Vessel, Representative, Arrival time, Berth, Terminal, Harbour, Departure time, Destination	https://www.portofhelsinki.fi/en/cargo-traffic-and-ships/arrivals-and-departures-cargo
			Arriving Vessels	Vessel, Representative, Arrival time, From, Berth, Terminal, Harbour	https://www.portofhelsinki.fi/en/cargo-traffic-and-ships/arrivals-and-departures-cargo
			Traffic of five previous days	Ship, Shipping company, Arrived, From, Harbour, Berth, Departured, Destination	https://www.portofhelsinki.fi/en/cargo-traffic-and-ships/traffic-five-previous-days
36	Finlande	Inkoo	Traffic	Vessel, Nationality, Arrival, From, Departure, To, Import, Export, Cargo, Amount	https://xml.inkooshipping.com/
37	Finlande	Jakobstad & Pietarsaari	Arrivals	ETA, Vessel, Flag, Nt, Dwt, LOA, From, To, Quay / Note	https://portofpietarsaari.fi/en/alusliikenne/
			Vessels in port and departures	ETD, Vessel, Flag, Nt, Dwt, LOA, From, To, Quay / Note	
			Vessels sailed	ETD, Vessel, Flag, Nt, Dwt, LOA, From, To, Quay / Note	
	Finlande	Joensuu	Aucune données exploitables		https://www.saimaansatamat.fi/
38	Finlande	Kemi	Arrivals	Updated, Vessel, Nat., From/To, ETA, Berth, ETS, Agent of Arrival / Agent of Departure, Discharge, Load	http://laivalista.keminsatama.fi/wpds.asp?0?1?10
			At Berth	Updated, Vessel, Nat., From/To, ATA, Berth, ETS, Agent of Arrival / Agent of Departure, Discharge, Load	http://laivalista.keminsatama.fi/wpds.asp?0?8?1

			Departed	Updated, Vessel, Nat., From/To, ATA, Berth, ATS, Agent of Arrival / Agent of Departure, Discharge, Load	http://laivalista.keminsatama.fi/wpds.asp?0?9?8
39	Finlande	Kokkola	Arrivals	Vessel, DWT, Nat., ETA, Agent, From / To, Berth	https://portofkokkola.fi/en/arrivals-and-departures/
			At Berth	Vessel, DWT, Nat., Arrival, Agent, From / To, Berth	
			Departed	Vessel, DWT, Nat., Departure, Agent, From / To, Berth	
40	Finlande	Naantali	Arriving Vessels	Vessel, Agent, Arr., Dep., From, To, Berth	http://laivalista.portofnaantali.fi/liikenne/Saapuvat.aspx
			Vessels in Harbour	Vessel, Agent, Arr., Dep., From, To, Berth	http://laivalista.portofnaantali.fi/liikenne/Satamassa.aspx
			Departed Vessels	Vessel, Agent, Arr., Dep., From, To, Berth	http://laivalista.portofnaantali.fi/liikenne/Lahteneet.aspx
41	Finlande	Oulu	Arrivals	Updated, Vessel, Nat., From/To, ETA, Berth, ETS, Agent of Arrival / Agent of Departure	https://www.oukapalvelut.fi/pds/default.aspx?0?1?9
			At Berth	Updated, Vessel, Nat., From/To, ATA, Berth, ETS, Agent of Arrival / Agent of Departure	https://www.oukapalvelut.fi/pds/default.aspx?0?8?1
			Departed	Updated, Vessel, Nat., From/To, ATA, Berth, ATS, Agent of Arrival / Agent of Departure	https://www.oukapalvelut.fi/pds/default.aspx?0?9?8
42	Finlande	Rauma	Arrivals	Vessel, Nat., From / To, ETA, Berth, ETS, Agent of Arr. / Dep., Vessel owner + DETAILS OF VESSEL (Name, Type, Radio call sign, LR Number, Ship Owner, Nationality, Length, Width, max Draught, Dead Wt, Gross, Net, Ice Classification, Double Bottom, Inert, Segregated Ballast, Gas Free Cargo Tanks, Engine, max Speed, Year of Build)	https://extra.portofrauma.com/laivalista/default.aspx?cmd=1&arg=0&prev=8
			At Berth	Vessel, Nat., From / To, ATA, Berth, ETS, Agent of Arr. / Dep., Vessel owner + DETAILS OF VESSEL (Name, Type, Radio call sign, LR Number, Ship Owner, Nationality, Length, Width, max Draught, Dead Wt, Gross, Net, Ice Classification, Double Bottom, Inert, Segregated Ballast, Gas Free Cargo Tanks, Engine, max Speed, Year of Build)	https://extra.portofrauma.com/laivalista/default.aspx?cmd=8&arg=0&prev=1
			Departed	Vessel, Nat., From / To, ATA, Berth, ATS, Agent of Arr. / Dep., Vessel owner + DETAILS OF VESSEL (Name, Type, Radio call sign, LR Number, Ship Owner, Nationality, Length, Width, max Draught, Dead Wt, Gross, Net, Ice Classification, Double Bottom, Inert, Segregated Ballast, Gas Free Cargo Tanks, Engine, max Speed, Year of Build)	https://extra.portofrauma.com/laivalista/default.aspx?cmd=9&arg=0&prev=8
	Finlande	Sköldvik	Aucune données exploitables		https://www.neste.com/en
43		Turku	Arrivals, At Berth, Departed	Arrives, Vessel / Broker, Berth, From / To, Date of Departure + DETAILS (Contact details, Vessel Information : Vessel type, Length, Width, Draught, DWT, GR, NT)	https://www.portofurku.fi/en/cargo-traffic/vessels/arrivals-and-departures/
	France	Ajaccio	Aucune données exploitables		http://www.2a.cci.fr/Port-d-ajaccio.html
	France	Bastia	Aucune données exploitables		http://www.bastia.port.fr/

	France	Bayonne	Aucune données exploitables		https://www.bayonne.port.fr/
44	France	Bordeaux	Navires à quais	Nom, Poste, Accostage, Sortie prévue, Agent montée, Agent sortie, n° d'escale, Provenance, Destination, Longueur, Déchargement, Chargement, Port en lourd, Armateur, Pavillon, Longueur, Jauge nette, Jauge brute, Largeur, Type, Port d'attache, Port en lourd	https://www.bordeaux-port.fr/fr/navires-quais-attendus
			Navires attendus	Nom, Poste, Date prévue arrivée, Provenance, Agent, Longueur, Déchargement, Chargement, Port en lourd, Armateur, Pavillon, Tirant d'eau AR montée, Longueur, Jauge nette, Jauge brute, Largeur, Type, Port d'attache, Port en lourd	
	France	Boulogne	Aucune données exploitables		https://www.portboulognecalais.fr/fr/
	France	Brest	Aucune données exploitables		http://www.brest.port.fr/fr/activites-commerciales/entree-sortie-navire
	France	Caen	Aucune données exploitables		https://www.caen.port.fr/
	France	Calais	Aucune données exploitables		https://www.portboulognecalais.fr/fr/
	France	Cherbourg	Aucune données exploitables		https://www.port-cherbourg.fr/
45	France	Dunkerque	Navires présents	Numéro poste à quai, Poste à quai, Shipflow-paq_id, Nom du navire, Ligne régulière, Agent consignataire, Date d'entrée jetée, Provenance, Destination, Prévision de départ	http://www.dunkerque-port.fr/fr/activites-commerciales/navires-presents-port-dunkerque.html?dunkerqueParam%5Borderby%5D=previsiondepart&dunkerqueParam%5Borderdir%5D=4
			Navires attendus en entrée/transit	Prévision d'arrivée, Nom du navire, Ligne régulière, Agent consignataire, Provenance, Poste à quai prévu, Destination	http://www.dunkerque-port.fr/fr/activites-commerciales/navires-entree-transit.html
			Navires attendus en sortie	Prévision de départ, Nom du navire, Ligne régulière, Agent consignataire, Destination	http://www.dunkerque-port.fr/fr/activites-commerciales/navires-sortie.html
	France	La Réunion	Aucune données exploitables		https://reunion.port.fr/
46	France	La Rochelle	Diverses	Diverses	https://larochelle.vigiesip.eu/vigiesip-larochelle/listeDemandes.jsf
47	France	Le Havre	Arrivées	N°AP, n°VOS, Bateau, Type, Consignataire, Arrivée, Infrastructure, Provenance, Destination, Longueur, Largeur, TE av, TE ar, Pavillon	https://www.havre-port.com/fr/liste/128/arrivees-au-grand-port-maritime-du-havre
			Départs	N°Escale, n°AP, n°VOS, Bateau, Type, Consignataire, Départ, Infrastructure, Provenance, Destination, Longueur, Largeur, TE av, TE ar, Pavillon	https://www.havre-port.com/fr/liste/129/departs-au-grand-port-maritime-du-havre
			Navires à quai	N°Escale, n°VOS, Navire, Type, Consignataire, Accostage, Appareillage, Infrastructure, Provenance, Destination, Longueur, Largeur, TE av, TE ar, Pavillon	https://www.havre-port.com/fr/liste/2000/navires-a-quai
48	France	Lorient	Mouvement des navires (À Quai, Attendus, Mouvements prévus)	Date d'arrivée, Navire, Produit, Tonnage, Agent du navire, Nombre de remorqueurs, Provenance, Destination, Localisation	http://lorient.port.fr/port-de-commerce/mouvements-des-navires

	France	Marseille	Aucune données exploitables		https://www.marseille-port.fr/fr/Accueil/
49	France	Nantes-Saint-Nazaire	Navires à quai	Accostage, Navire, Site, Chargement / Déchargement, Marchandise, Tonnage	http://www.nantes.port.fr/le-port-en-direct/mouvements-des-navires-en-temps-reel/navires-a-quai/
			Prévisions de mouvements	ETA, ETD, Nom du navire, Longueur du navire, E/D/S, Poste, Site, Agent	http://www.nantes.port.fr/le-port-en-direct/previsions-de-mouvements/
50		Nouméa (Nouvelle-Calédonie)	Le mouvement des navires	Étape, Nom du navire, Type de navire, Code port de N-C, Code d'amarrage, Date d'arrivée prévue, heure d'arrivée prévue, Port de provenance, Date d'arrivée, heure d'arrivée, Code port de destination, Date de départ, Heure de départ, Agence	https://noumeaport.nc/la-capitainerie/mouvement-des-navires/
	France	Pointe-à-Pitre (Guadeloupe)	Aucune données exploitables		https://guadeloupe-portcaraibes.com/sites-portuaires/pointe-a-pitre/
	France	Port-la-Nouvelle	Aucune données exploitables		https://www.port-la-nouvelle.com/
51	France	Rouen	Navires attendus	Date d'entrée, Navire, Pavillon, Agent, Provenance, Destination, Observation	https://www.capitainerie-rouen.com/fr/navires-attendus
			Navires présents	Arrivée sur rade, Navire, Pavillon, Agent, Provenance, Destination, Observation	https://www.capitainerie-rouen.com/fr/navires-presents
			Navires sortis	Date de sortie, Navire, Pavillon, Agent, Provenance, Destination, Observation	https://www.capitainerie-rouen.com/fr/navires-sortis
	France	Saint-Malo	Aucune données exploitables		https://saintmalo.port.fr/
	France	Sète	Aucune données exploitables		http://www.sete.port.fr/fr
	France	Toulon	Aucune données exploitables		https://www.portsradetoulon.com/
	Grèce	Heraklion	Aucune données exploitables		https://portheraklion.gr/index.php/en/organization/marine-traffic-en
	Grèce	Igoumenitsa	Aucune données exploitables		https://olig.gr/en/
	Grèce	Patras	Aucune données exploitables		http://www.patrasport.gr/cms/?page_id=1192&lang=en
	Grèce	Piraeus	Aucune données exploitables		http://www.olp.gr/en/arrivals-and-departures
52	Grèce	Thessaloniki	Ship Arrivals	Vessel Name, Storage ID, Arrival Date, Departure Date, Total Hours on Dock	https://portal.thpa.gr/fnet5/shiparrivals/?lang=en
	Grèce	Volos	Aucune données exploitables		https://www.port-volos.gr/cgi-bin/pages/index.pl?type=index&arlang=Greek
53	Irlande	Cork	Vessels Expected	Status, Vessel, ETA, CFM, OAL, DFT, Berth	https://www.portofcork.ie/index.cfm/page/shippingtimes
	Irlande	Drogheda	Aucune données exploitables		http://www.droghedaport.ie/

54	Irlande	Dublin	Ships in Port	Vessel Name, Vessel Type, Arrival Date, Departure Time, Berth	https://www.dublinport.ie/wp-ison/ships/v2/all
			Arrivals	Date, Time, Vessel, Vessel Type, To Location	https://www.dublinport.ie/wp-ison/arrivals/v2/all
			Departures	Date, Time, Vessel, Vessel Type, From Location	https://www.dublinport.ie/wp-ison/departures/v2/all
55	Irlande	Galway	Arrivals / Departures	Date, Time, Vessel, Vessel Type	https://theportofgalway.ie/shippingschedule/
	Irlande	New Ross	Aucune données exploitables		http://newrossport.ie/
56	Irlande	Waterford	Arrivals	Date, Time, Vessel, Vessel Type, To	http://www.portofwaterford.com/arrivals
			Departures	Date, Time, Vessel, Vessel Type, From	http://www.portofwaterford.com/departures
	Italie	Ancona	Aucune données exploitables		https://porto.ancona.it/
	Italie	Augusta (+ Catania)	Aucune données exploitables (via MarineTraffic.com)		https://www.adspmaresiciliaorientale.it/
57	Italie	Bari	Navi in Porto	Nave, IMO, Flag, Tipo, Posizione, Data Arrivo	https://www.adspmam.it/ship-info/gaia-ship-info-bari/
	Italie	Civitavecchia (+ Fiumicino, Gaeta)	Aucunes données exploitables		https://www.portidiroma.it/
	Italie	Gênes (+ Vado, Savona, Pra')	Aucune données exploitables		https://www.portsofgenoa.com/fr/ http://www.pcs-eport.it/eportHomePage/ http://pcs.porto.sv.it/
	Italie	Cagliari	Aucune données exploitables		http://www.adspmaredisardegna.it/
	Italie	Gioia Tauro	Aucune données exploitables		http://www.portodigioiatauro.it/?lang=en
	Italie	Livourne (+ Capraia Isola, Piombino, Portoferraio, Rio Marina, Cavo)	Aucune données exploitables		https://www.portialtotirreno.it/en/ http://www.shipinfo.it/web/guest/home
	Italie	Messina (+ Milazzo)	Aucune données exploitables		http://www.porto.messina.it/
	Italie	Napoli (+ Salerno, Castellammare di Stabia)	Aucune données exploitables		https://adsptirrenocentrale.it/
	Italie	Palermo	Site inaccessible		http://www.portpalermo.it/
	Italie	Ravenna	Aucune données exploitables		http://www.port.ravenna.it/
58	Italie	Trieste	Navi in Porto	Nave, Tipo nave, Operatore, Terminal, Da, Per	http://sinfomar.porto.trieste.it/index.jsp

	Italie	Tarente	Aucune données exploitables		https://www.port.taranto.it/index.php/en/ship-movement
	Italie	Venise	Aucune données exploitables		https://www.port.venice.it/
	Lettonie	Liepaja	Aucune données exploitables		https://liepaja-sez.lv/en
	Lettonie	Riga	Aucune données exploitables		https://rop.lv/en/
	Lettonie	Ventspils	Aucune données exploitables		http://www.portofventspils.lv/en/
59	Lituanie	Klaipeda	List of vessel within the port	Ship, IMO No., Flag, Date time, Agent, Quay No., Port Company	https://www.portofklaipeda.lt/list-of-vessels-within-the-port
			Planned calls	Ship, IMO No., Flag, Date time, Agent, Quay No.	https://www.portofklaipeda.lt/planned-calls
			Last calls	Ship, IMO No., Flag, Date time, Agent, Quay No.	https://www.portofklaipeda.lt/last-calls
			Last departures	Ship, IMO No., Flag, Date time, Agent, Quay No.	https://www.portofklaipeda.lt/last-departures
	Malte	Valletta	Aucune données exploitables		https://www.transport.gov.mt/maritime/local-waters/ports-in-malta/port-of-valletta-118
	Malte	Malta Freeport	Aucune données exploitables		http://www.maltafreeport.com.mt/home.aspx
60	Norvège	Bergen	Call Schedules	Name, Time of arrival, Time of departure, Berth, Flag state, Gross tonnage, From / to	https://bergenhavn.no/wp-ison/portcalls/v1/calls?type=cruise&duration=year
61	Norvège	Drammen	Boats	Nom du navire, Nationalité, Arrivée prévue, Quai, De, À, Status + DETAILS	http://drammenhavn.no/boats/
62	Norvège	Karmsund	Expected port calls	Vessel, Mooring place, Arrival, Departure, From harbor, To harbor	https://karmsundhavn.no/en/maritime/traffic-arrival/
			Vessels in port	Vessel, Mooring place, Arrival	
63	Norvège	Narvik	Latest arrivals	Vessel, Area, Entry, Stop	https://www.narvikhavn.no/Get_Hbm_html.aspx?u=a
			Latest departures	Vessel, Area, Start, Exit	https://www.narvikhavn.no/Get_Hbm_html.aspx?u=a
			Expected vessels	Vessel, ETA, Calc ETA, Destination	https://www.narvikhavn.no/Get_Hbm_html.aspx?u=e
64	Norvège	Oslo	Port calls	Arrivée, Bateau, Quai / Terminal, Date de départ, Type de navire + DETAILS (Tonnage, Drapeau, De, À)	https://oslohavn.no/en/port-calls/
65	Norvège	Stavanger (+Risavika, Mekjarvik, Cruise)	Expected, arrived, departed	Name, Arrival/Departure, Mooring location, Quay code, Vessel Type, Status + DETAILS (Arrived from, departing to, departure, mooring location, agent, length, vessel width, flag, deadweight, imo, signal, vessel type)	https://www.stavangerhavn.no/en/maritim/vessel-traffic/

66	Norvège	Tromsø	Expected vessels	Navire / Quai, Arrivée, Départ, De, À	https://www.tromso.havn.no/en/traffic/traffic-overview/
			Vessels in port	Navire, Amarrage, Départ	
67	Pays-Bas	Amsterdam	Arrival, Departure, In port, Shifting	Ship, Nationality, Time (ETA, ATA, ETD, ATD), Harbour, Berth + DETAIL (Agent, Length, Largeur, Draft, Deadweight)	https://www.portofamsterdam.com/fr/scheepvaart/aankomst-en-vertrek
68	(Pays-Bas)	(Aruba)	Arrivals	Arrival Date, Vessel Name, Agent, ETA, Arrival Dock, Vessel Type	https://www.arubaports.com/tracker/
			Departures	Departure Date, Vessel Name, Agent, ETD, Departure Dock, Vessel Type	
	Pays-Bas	Groningen seaports	Aucune données exploitables (erreur 404)		https://www.groningen-seaports.com/en/shipping/current-port-information/
	Pays-Bas	Moerdijk	Aucune données exploitables		https://www.portofmoerdijk.nl/en/
69	Pays-Bas	North Sea Port (Vlissingen & Terneuzen + Gent)	Expected arrivals	Ship, Ship no, LOA, Dr., Flag, Location, Berth, Time, Port, Agent	https://en.northseaport.com/enigma/detail/19/400
			Ships in port	Ship, Ship no, LOA, Dr., Flag, Location, Berth, Agent	https://en.northseaport.com/enigma/detail/20/ships-in-port
			Expected sailings	Ship, Ship no, LOA, Dr., Flag, Location, Berth, Time, Agent	https://en.northseaport.com/enigma/detail/21/expected-sailings
			Sailing ships	Ship, Ship no, LOA, Dr., Flag, Location, Berth, I/O, Agent	https://en.northseaport.com/enigma/detail/22/sailing-ships
			Recent arrivals	Ship, Ship no, LOA, Dr., Flag, Location, Berth, Time, Origine, Agent	https://en.northseaport.com/enigma/detail/23/recent-arrivals
			Recent sailings	Ship, Ship no, LOA, Dr., Flag, Location, Berth, Time, Agent	https://en.northseaport.com/enigma/detail/24/recent-sailings
70	Pays-Bas	Rotterdam	Expected ships	IMO, UCRN, Berth, datetime (ETA), name, MMSI, call sign, length, length_registered, breadth, breadth_extreme, draught, ship_category_code, power_kw_max, gross_tonnage, nett_tonnage, deadweight, flag_code, port_of_registry, year_of_build, status_date, operator, style_image.	https://www.portofrotterdam.com/ship_visits/all
			Present ships	Idem, sauf datetime (ETD)	
			Departed ships	Idem, sauf datetime (ETD)	
71	Pologne	Gdansk	Ships in port	Ship's name, Flag, Additionnal information	http://www.portgdansk.pl/shipping/ships-in-port
72	Pologne	Gdynia	Ships in port	Name, Flag, Quay	https://www.port.gdynia.pl/files/wydarzenia/statki/ships_en.php
73	Pologne	Świnoujście	Vessel Traffic in Swinoujscie	Name, Flag, Agent, Group of cargo, Quay, Arrival, Departure	http://www.port.home.pl/ruch/Nruch_SWa.htm
74	Pologne	Szczecin	Vessel Traffic in Szczecin	Name, Flag, Agent, Group of cargo, Quay, Arrival, Departure	http://www.port.home.pl/ruch/Nruch_SZa.htm
75	Portugal	Açores (plusieurs ports)	Mouvements portuaires	Date, Opération, Navire, IMO, Type de navire, Pax (?), Tripulantes (?), Origine, Destination, Escale, Contremarque	https://portosdosacores.pt/movimento-portuario/

76	Portugal	Aveiro	Entrées	ETA (Date, Horaires), N°Imo, Navire, Drapeau, LOA, GT, Port (Origine, Destination), Terminal, Poste attribué, Agent, Marcardoria (?)	http://movimentoportuario.portodeaveiro.pt/apa_portal/do?com=DS;9095988476;9001000;+RCNT(1000);+RCNT(1000)
			Navires au port	N°Processus, Entrée (Date, Horaire), ETD (Date, Horaire), N°Imo, Navire, Drapeau, LOA, GT, Port (Origine, Destination), Poste attribué, Agent, Marcardoria (?)	http://movimentoportuario.portodeaveiro.pt/apa_portal/do?com=DS;9095988476;9001010;+RCNT(1000);+RCNT(1000)
			Sorties	N°Processus, ETD (Date, Horaire), N°Imo, Navire, Drapeau, LOA, GT, Port (Origine, Destination), Poste attribué, Agent, Marcardoria (?)	http://movimentoportuario.portodeaveiro.pt/apa_portal/do?com=DS;9095988476;9001040;+RCNT(1000);+RCNT(1000)
	Portugal	Faro	Berthed Ships	No exist information	http://www.portodefaro.pt/en/ships/
			Ships at Anchor	No exist information	
			Ships in maneuvers	No exist information	
			Ships hovering	No exist information	
			Arrival Forecasts	No exist information	
			Output Forecasts	No exist information	
77	Portugal	Figueira da Foz	Arrivals	IMO, Ship, Type, Agent, Terminal, Arrival Date, Arrival Time, Departure Date, Departure Time, Port of Origin, Port of Dest, GT, DWT	http://iup.portofigueiradafoz.pt/apffiup/gespor/public/PVEntrada APFF.aspx
			Departures	IMO, Ship, Type, Agent, Terminal, Arrival Date, Arrival Time, Departure Date, Departure Time, Port of Origin, Port of Dest	http://iup.portofigueiradafoz.pt/apffiup/gespor/public/PVPartidas APFF.aspx
78	Portugal	Leixoes (Porto)	ETA - Estimated Time of Arrival	Estimated Arrival Time, Process, Custom process ID, Imo Nr, Vessel, Call Reference, Flag, Comp., GT, Berthing Place, Draught (Entrance / Departure), Port (Origin / Destination), Shipping Agent, Vessel Type	http://www.apdl.pt/site-apdl/planeamento/naviosprevchegada.jsp?lang=en
			Vessels in the approaching area	Estimated Arrival Time, Process, Imo Nr, Vessel, Call Reference, Flag, Comp., GT, Draught (Entrance, Departure), Port (Origin, Destination), Shipping Agent, Vessel Type	http://www.apdl.pt/site-apdl/planeamento/navioszonaaprox.jsp?lang=en
			Berth Vessels	Estimated Departure Time, Process, Imo Nr, Vessel, Comp., GT, Draught (Entrance / Departure), Port (Origin / Destination), Berthing Place, Shipping Agent, Vessel Type	http://www.apdl.pt/site-apdl/planeamento/naviosporto.jsp?lang=en
			Vessels - Foreseen Movements	Movements (Type, Place, Date, Hour, Tugs n°), Process, Vessel, Comp., GT, Shipping Agent, Vessel Type	http://www.apdl.pt/site-apdl/planeamento/naviosmanoprev.jsp?lang=en
			Vessels - Executed Movements on the last days	Movements (Type, Place, Date, Hour, ???), Process, Vessel, Comp., GT, Shipping Agent, Vessel Type	http://siga.apdl.pt/site-multiporto/planeamento/naviosmanoreal.jsp?lang=en&portoescala=PTLEI
			Arrival Time at 6 Miles	Arrival Time at 6 miles, Process, Imo Nr, Vessel, Call Reference, Comp., GT, Draught (Entrance / Departure), Port (Origin / Destination), Shipping Agent, Vessel Type	http://siga.apdl.pt/site-multiporto/planeamento/6m.jsp?lang=en&portoescala=PTLEI
79	Portugal	Lisbonne	All Ships in Port	Process n°, Countermark, Ship Name, IMO N°, ATA, ETD, Flag, Ship Type, Agent, Following Port, Previous Port, Reason for Stopover, Attributed Location	http://www.portodelisboa.pt/portal/page/portal/PORTAL_PORTO_LISBOA_ING/SHIPS_P ILOTAGE/IN_PORT

			Estimated Arrivals	Process n°, Countermark, Ship Name, IMO N°, ATA, ETD, Flag, Ship Type, Agent, Following Port, Previous Port, Reason for Stopover, Attributed Location	http://www.portodelisboa.pt/portal/page/portal/PORTAL_PORTO_LISBOA_ING/SHIPS_P_ILOTAGE/ARRIVALS
80	Portugal	Madère	Ship Movements	Day, Ship, Agent, Arrival, Departure, Port of Origin, Port of Destiny, Port of Call	http://www.apram.pt/site/index.php/en/ports/ship-movements
81	Portugal	Setubal	Arrivées	IMO, Ship, Type, Agent, Terminal, Arrival Time, Departure Time, Port of Origin, Port of Dest, GT, DWT, Contremarque	https://iup2.portodesetubal.pt/iup/public/ecraNaviosChegada.jspx
			Départs	IMO, Ship, Type, Agent, Terminal, Arrival Time, Departure Time, Port of Origin, Port of Dest, GT, DWT, Contremarque	https://iup2.portodesetubal.pt/iup/public/ecraNaviosSaida.ispx
82	Portugal	Sines	Berthed Ships	Call, Countermark, IMO, Ship, Ship Type, Origin / Destiny, Since	http://www.portodesines.pt/en/ships/
			Ships at Anchor	Call, Countermark, IMO, Ship, Ship Type, Origin / Destiny, Since	
			Ships in maneuvers	Call, Countermark, IMO, Ship, Ship Type, Origin / Destiny, Since	
			Ships hovering	Call, Countermark, IMO, Ship, Ship Type, Origin / Destiny, Since	
			Arrival Forecasts	Call, Countermark, IMO, Ship, Ship Type, Origin, Forecast	
			Output Forecasts	Call, Countermark, IMO, Ship, Ship Type, Destiny, Forecast	
83	Portugal	Viana do Castelo	ETA - Estimated Time of Arrival	Estimated Arrival Time, Process, Custom process ID, Imo Nr, Vessel, Call Reference, Flag, Comp., GT, Berthing Place, Draught (Entrance / Departure), Port (Origin / Destination), Shipping Agent, Vessel Type	http://viana.apdl.pt/site-apdlvc/planeamento/naviosprevchegada.jsp?lang=en
			Vessels in the approaching area	Estimated Arrival Time, Process, Imo Nr, Vessel, Call Reference, Flag, Comp., GT, Draught (Entrance, Departure), Port (Origin, Destination), Shipping Agent, Vessel Type	http://viana.apdl.pt/site-apdlvc/planeamento/navioszonaaprox.jsp?lang=en
			Berth Vessels	Estimated Departure Time, Process, Imo Nr, Vessel, Comp., GT, Draught (Entrance / Departure), Port (Origin / Destination), Berthing Place, Shipping Agent, Vessel Type	http://viana.apdl.pt/site-apdlvc/planeamento/naviosporto.jsp?lang=en
			Vessels - Foreseen Movements	Movements (Type, Place, Date, Hour, Tugs n°), Process, Vessel, Comp., GT, Shipping Agent, Vessel Type	http://viana.apdl.pt/site-apdlvc/planeamento/naviosmanoprev.jsp?lang=en
			Vessels - Executed Movements on the last days	Movements (Type, Place, Date, Hour, ???), Process, Vessel, Comp., GT, Shipping Agent, Vessel Type	http://siga.apdl.pt/site-multiporto/planeamento/naviosmanoreal.jsp?lang=en&portoescala=PTVDC
			Arrival Time at 6 Miles	Arrival Time at 6 miles, Process, Imo Nr, Vessel, Call Reference, Comp., GT, Draught (Entrance / Departure), Port (Origin / Destination), Shipping Agent, Vessel Type	http://siga.apdl.pt/site-multiporto/planeamento/naviosmanoreal.jsp?lang=en&portoescala=PTVDC
	Roumanie	Constanța	Aucune données exploitables		https://www.portofconstantza.com/pn/home
84	Royaume-Uni	Aberdeen	Vessel Arrivals	Vessel Name, Arrival Date, Time, Arrived From, Pilot, IMO No., Visit ID	http://www.aberdeen-harbour.co.uk/article/vessel-arrivals/

			Vessels in Port	Vessel Name, Arrival Date, Time, Berth Offset, Arrival Port, IMO No., Visit ID	http://www.aberdeen-harbour.co.uk/article/vessels-in-port/
			Vessel Sailings	Vessel Name, Sailing Date, Time, Sailed To, Pilot, IMO No, Visit ID	http://www.aberdeen-harbour.co.uk/article/vessel-sailings/
			Expected Arrivals	Vessel Name, Arrival Date, Time, IMO No, Visit ID	http://www.aberdeen-harbour.co.uk/article/vessel-expected-arrivals/
85	Royaume-Uni	Belfast	Traffic Expected	Date & Time (Estimated), Ship, Job Type, To/From, Cargo, Last/Next Port, Inward Agent, Outward Agent	https://booking.belfast-harbour.co.uk/
			Traffic History	Date & Time (Actual), Type, Ship, To/From, Inward Agent, Outward Agent	
			In Port	Arr Date & Time (Actual), Dep Date & Time (Estimated), Ship, Berth, Port, Inward Agent, Outward Agent	
86	Royaume-Uni	Blyth	In Port, Due	No., Name, Flag, DWT, Berth, Last Port, ETA / Arrival, Next Port, ETS / Depart	http://portofblyth.co.uk/port-live/
87	Royaume-Uni	Bristol	Forward Movements	Rotn, Vessel, Agent, ETA, L x B, D, A - PO, Dock, Berth, ETD, D - PO, Flag, From, To, Purpose, A_Tugs, D_Tugs, PL, Comments, dD, Cargo	https://www.bristolport.co.uk/shipping/forward-movements
			In Dock	Vessel, Dock, Berth, Purpose, ETD, Flag, D, From, To, PO time, PL, L x B, Roth, Berthing Comment, Comment, Cargo, Critical Vessel	https://www.bristolport.co.uk/shipping/dock
	Royaume-Uni	Douvres / Dover	Aucune données exploitables		https://www.doverport.co.uk/
88	royaume-Uni	Felixtowe	Arrivals	Arrival Date and Time, Vessel Name, Terminal, Nationality, Gross Tonnage, Overall Length, Last Port, Next Port, Ship Agent	https://www.portoffelixstowe.co.uk/sailing-schedule/shipping-information/#Arrivals
			Departures	Departure Date and Time, Vessel Name, Terminal, Nationality, Gross Tonnage, Overall Length, Last Port, Next Port, Ships Agent	https://www.portoffelixstowe.co.uk/sailing-schedule/shipping-information/#Departures
			Closing	Estimated Arrival, Vessel Name, Rotation, Ships Agent, Last Port, Commence Time, Closing Time	https://www.portoffelixstowe.co.uk/sailing-schedule/shipping-information/#ClosingTimes
	Royaume-Uni	Forth	Aucune données exploitables		https://www.forthports.co.uk/
	Royaume-Uni	Glasgow	Aucune données exploitables		http://www.peelports.co.uk/
89	Royaume-Uni	Grimsby / Immingham / Hull / Goole	Planned Voyage Summary	Ship name, Draft, LOA, DWT, On passage Date & Time, Origin, Destination	http://www.humber.com/admin/ftp/Planned%20Voyages%20Summary.pdf
			Ships at Berth Summary	Ship name, Draft, LOA, DWT, Berthed, Berthed at	http://www.humber.com/admin/ftp/Ships%20at%20Berth%20Summary.pdf
			Ships on Passage Summary	Ship name, Draft, LOA, DWT, On Passage, Origin, Destination, Last Reported At	http://www.humber.com/admin/ftp/Ships%20on%20Passage%20Summary.pdf
	Royaume-Uni	Liverpool	Aucune données exploitables		https://www.peelports.com/marine-information?port=liverpool
90	Royaume-Uni	Londres	Ships expected arrivals	Date, Time, At, Vessel Name, Agent, Nationality, From, To	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=5
			Ships arrivals list	Time, Vessel Name, Ref, Nationality, From, To	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=2
			Ships expected departures	Date, Time, At, Vessel Name, Agent, Nationality, From, To	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=6
			Ships departures list	Time, Vessel Name, Ref, Nationality, From, To	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=3

			Ships expected in-port shifts	Date, Time, At, Vessel Name, Agent, Nationality, From, To	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=7
			Ships in-port list	Location, Vessel name, Ref, Nationality, Berthed	http://www.pla.co.uk/Port-Trade/Ship-movements/Ship-movements?flag=4
91	Royaume-Uni	Millford Haven	Vessels Arriving	ETA, Ship, Ship Type, GT, From, To, Agent	https://www.mhpa.co.uk/vessels-arriving/
			Vessels at Berth	Berth name, Ship, Ship name, GT, Berthed Time, Next Movement Time	https://www.mhpa.co.uk/vessels-at-berth/
92	Royaume-Uni	Southampton	Planned Shipping Movements : Arrivals, Movements, Departures	Ship Name, Lloyds No, Type, Flag, ETA, Dir, Origin, Destination, Side To, Draft, LOA, DWT, Agent	http://www.southamptonvts.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southampton\Shipping_Movements\sot_extract.xml&xslFilePath=D:\wwwroot\CMS_Southampton\content\files\assets\sotextract.xsl&w=0
			Vessels Alongside	Ship Name, Lloyds No, Type, Flag, Actual Arrival Time, Berth, Origin, Side To, Draft, LOA, DWT, Agent	http://www.southamptonvts.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southampton\Ships_Along_Side\sotberthed.xml&xslFilePath=D:\wwwroot\CMS_Southampton\content\files\assets\sotberthed.xsl&w=59
			Vessels Underway	Ship Name, Lloyds No, Flag, On Passage Time, Origin, Destination, Draft, LOA, DWT, Agent	http://www.southamptonvts.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southampton\Ships_On_Passage\sotonpassage.xml&xslFilePath=D:\wwwroot\CMS_Southampton\content\files\assets\sotonpassage.xsl&w=10
			Vessel Sailed	Ship Name, Lloyds No, Type, Flag, Actual Time of Departure, Origin, Destination, Side to, Draft, LOA, DWT, Agent	http://www.southamptonvts.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southampton\Vessels_Sailed\sotsailed.xml&xslFilePath=D:\wwwroot\CMS_Southampton\content\files\assets\sotsailed.xsl&w=53
			Ships At Anchor	Ship Name, Anchorage, Estimated Departure, GT, LOA, Flag, Origin, Actual Arrival Time, Draft, Lloyds Number, Purpose, Type, DWT, Agent	http://www.southamptonvts.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southampton\Ships_Anchored\sotananchored.xml&xslFilePath=D:\wwwroot\CMS_Southampton\content\files\assets\sotananchored.xsl&w=17
93	Royaume-Uni	Southwales Ports (Barry, Cardiff, Newport, Port Talbot, Swansea)	Arrivals	Ship Name, Type, Flag, Estimated Time of Arrival, Dir, Origin, Destination, Side To, Draft, Agent, Lloyds No, LOA, Beam, DWT, GT, Tugs, Aids	http://www.southwalesports.co.uk/BackgroundSite/Ajax/LoadXmlFileWithTransform?xmlFilePath=D:\ftp\southwales\Shipping_Movements\sw_extract.xml&xslFilePath=D:\wwwroot\CMS_SWales\content\files\assets\sw_extract.xsl&w=34
			Movements	Ship Name, Type, Flag, Estimated Departure Time, Dir, Origin, Destination, Side To, Draft, Agent, Lloyds No, LOA, Beam, DWT, GT, Tugs, Aids, Boatmen, POB, Purpose	
			Departures	Ship Name, Type, Flag, Estimated Departure Time, Dir, Origin, Destination, Side To, Draft, Agent, Lloyds No, LOA, Beam, DWT, GT, Tugs, Aids, Boatmen, POB, Purpose	
94	Royaume-Uni	Tees / Hartlepool	Shipping Movements	HM No., Order Time, Order Type, Vessel, IMO No., Call Sign, Move Type, From, To, Agent, Operation, Tugs Required	https://www.pdports.co.uk/marine-information/shipping-reports/
			Arrivals	HM No., Vessel, IMO No., Vessel Anchor Time, Passing Breakwater, Vessel Arrival Time, From, To, Agent, Operation, Tugs Used	

			Sailings	HM No., Vessel, IMO No., Vessel Departure Time, Passing Breakwater, From, To, Agent, Tugs Used	
			Ship Location	HM No., Berth, Vessel, IMO No., Arrival Time, Agent, Estimated Departure, Operation	
95	Royaume-Uni	Tyne	Expected	Ship Name, ETA/ETD, From, To, Order + DETAILS (ARR/MOVE/SAIL, Agent, Flag, Call sign, IMO)	https://www.portoftyne.co.uk/marine-information/shipping-movements
			At Berth	Ship Name, ETD, From, Flag, IMO + DETAILS (Agent, Call Sign)	
			History	Ship Name, Start, From, To, IMO + DETAILS (ARR/MOVE/SAIL, Agent, Flag, GRT)	
96	Slovenie	Koper	Arriving Vessels	Call, Name, Date, Type + DETAILS (Call, Vessel code, Date, LOA, Draft, Cargo (t), Type of cargo, GT, Agent, Shipowner, Voyage, Last change)	https://www.luka-kp.si/eng/announcements-of-ships-pilot-planes-and-mooring
			Pilotage Plan	Call no., Vessel code, Berth, Date + DETAILS (Call no., Date, Vessel code, Vesse type, Draft, Vessel length, BT, Berth, Pilots, Agent, Action)	
			Ships Alongside	Date, Call no., Vessel code, Berthe, Operation, Type of cargo + DETAILS (Call no., Date, Vessel code, Type of cargo, Cargo weight (t), Operation, No. of working groups in 1st shift, Operation, Berthe)	
97	Suède	Gavle	Ship information	Ship, From/to, Terminal, Berth, Departure berth, Agent	https://gavlehamn.se/EN/moored-ships
			Expected vessels	Ship, From/to, Terminal, Berth, Arrival berth, Agent	
98	Suède	Goteborg	Arrivals and departures	Vessel, Arrival time / Departure time, Port name, Agent, Status + DETAILS (Length, width, Flag, IMO-number, Signal, Vessel Type) + VESSEL HISTORY	https://www.portofgothenburg.com/maritime/arrivals-and-departures/get_arrivals?query=&harbour_part_id=all&from_date=&to_date=
	Suède	Falkenbergs	Aucune données exploitables		http://www.falkenbergs-terminal.se/en/home/
99	Suède	Helsingborg	Navires dans le port	Statut, Arrivée, Départ prévu, Nom, Courtier (agent ?), Kaj, LO/LA	http://www.port.helsingborg.se/
	Suède	Kalmar	Aucune données exploitables		https://www.kalmar.se/kalmarhamn
	Suède	Karlskrona	Aucune données exploitables		https://www.karlskrona.se/naringsliv/naringslivet-i-karlskrona/karlskrona-hamn/
100	Suède	Malmö	Port Calls	Arrival, ETA, Ship, From, Quay, Agency, Status + DETAILS (Arrival from, Incl. port, Cargo, Quantity, Loading recipient, Departure to, Loading port, Cargo, Quantity, Loading recipient, Call sign, IMO, Nationality, BT, Vessel type, Ship length, Passenger capacity, Captain, Call no., Main purpose, Quay name, Dangerous goods)	http://www.cmport.com/ships-in-port/calls/malmo
101	Suède	Norrköping	Vessels in port	Vessel, Berth, Dim (L/B/DW), Arrival / Departure, Goods, Agent, Arriving from / Sailing to	https://www.norrkopingshamn.se/en/ankomstinformation/fartyg-i-hamnen

			Arriving vessels	Vessel, Berth, Dim (L/B/DW), Arrival, Goods, Agent, Arriving from / Sailing to	https://www.norrkopingshamn.se/en/ankomstinformation/fartyg-i-hamnen-anlop
	Suède	Oxelösunds	Aucune données exploitables		https://www.oxhamn.se/
102	Suède	Stockholm (+ Nynäshamn, Kapellskär, Cruise calls)	Vessel calls	Vessel, Arrival / Departure, Berth, Vessel type, Status + DETAILS (Arrives from, Departs to, Departure, Berth, Agency, Length, Width, Flag, Deadweight, IMO-number, Signal, Vessel type)	https://www.portsofstockholm.com/vessel-calls/
103	Suède	Wallhamn	Traffic	Date d'arrivée, Navire, Agent	http://www.wallhamn.com/
	Turquie	Atakas	Aucune données exploitables		http://atakasliman.com.tr/tr
	Turquie	Bandirma	Aucune données exploitables		http://www.portofbandirma.com.tr/?lang=en
	Turquie	Belde	Aucune données exploitables		http://www.beldeport.com.tr/
	Turquie	Canakkale	Aucune données exploitables		http://portofcanakkale.com/
	Turquie	Gemlik	Aucune données exploitables		http://www.gemport.com.tr/
	Turquie	Istanbul	Aucune données exploitables		https://www.tdi.gov.tr/
	Turquie	Kumport	Vessel Monitoring	Navire, arrivée, départ, ligne, charge, quai	http://online.kumport.com.tr/#!/vessels
	Turquie	Limar	Aucune données exploitables		http://www.limar.com.tr/en/home_page.html

Annexe 5 : Sources web, agents maritimes

Ports	Agents / Armateurs	Sites web	Données disponibles	Liens
Hambourg DE				
Niedersachsen DE				
Copenhague DK	BALTIC SHIPPING COMPANY	https://balticshipping.dk/	Fleet	
	SCHULTZ SHIPPING	http://schultzshipping.dk/	Fleet list	
	SDK SHIPPING	https://sdkgroup.com/	-	
	UNIFEEDER	https://www.unifeeder.com/	Operational Schedule	https://www.unifeeder.com/schedule
Aarhus DK				
Fredericia DK	BLUE WATER SHIPPING	https://www.bws.net/	-	
	FREDERICIA SHIPPING	https://www.fredericiashipping.dk/	-	
	SDK SHIPPING	https://sdkgroup.com/	-	
Bordeaux FR	BAGGERWERKEN	http://www2.deme-group.com/	-	
	HUMANN-TACONET	http://www.humann-taconet.fr/	-	
	MARMEDSA	https://www.marmedsabemarine.com/?lang=fr	-	
	NORIENT	https://www.norientpool.com/	Fleet	
	SEA INVEST	https://www.sea-invest-sa.com/	-	
	SOCATRA	https://socatra.com/	Fleet	
	TAYLOR MARITIME	https://www.taylormaritime.com/	Fleet	
	WORMS SM	https://www.worms-sm.com/	-	
Dunkerque FR	AMLF AGENA	https://www.agenatramp.fr/	-	
	CMA/CGM	http://www.cma-cgm.fr/	Horaires par port	http://www.cma-cgm.fr/ebusiness/schedules/port
	EURO DOCKS SERVICES	https://eurodocks.fr/	-	
	LEMAIRE-V-SHIPPING	http://www.lemvshipping.com/	-	
Le Havre FR	BURGER PORT AGENCIES FRANCE	-	-	
	INCHCAPE SHIPPING SERVICES	https://www.iss-shipping.com/	-	
	OCEAN NETWORK EXPRESS	https://www.one-line.com/	Fleet list	-

Amsterdam	NL	DEN HELDER SUPPORT SERVICE	https://dhss.nl/	-	
		EURO NORDIC LOGISTICS	https://www.euronordic.nl/	-	
		KUHLMAN REPKO SHIPPING	https://www.krshipping.nl/	-	
		KVSA	https://www.kvsa.nl/	Via Ship2Report (payant)	http://www.ispsamsterdam.nl/
		RHENUS LOGISTICS	https://www.rhenus.com/	-	-
		SILVER STAR AGENCIES	https://www.silver-star-agencies.com/	-	-
		TRANSITO AGENCY	http://www.transito.nl/	-	-
		TRANS-SAAR	http://www.transsaar.nl/	-	-
		VERTOM AGENCIES BV	https://www.vertom.nl/	ETA & ISPS Page	https://isps.vertom.nl/fmi/webd#ISPSVisitorsList
		VOPAK AGENCIES	https://www.vopakagencies.com/	Arrivals & Departures	https://portview.agencies.vopak.com/ords/f?p=PUB%3A1%3A0%3A%3A%3A%3A
		WAGENBORG AGENCIES	https://www.wagenborg.com/	-	-
Rotterdam	NL	CARL BUTTNER GMBH & CO	https://www.carlbuettnner.de/	Fleet	
		ESSBERGER JT GMBH	https://www.essberger.biz/	Fleet	
		HANSA TANKERS	http://www.hansa-tankers.com	Fleet	
		SEACONTRACTORS BV	https://www.seacontractors.com/	Live map	https://www.seacontractors.com/live-map.htm
		WILSON EUROCARRIERS AS	https://www.wilsonship.no/	Fleet	
Klaipeda	LT	ALL OVERSEAS	https://alloverseas.net/	-	
		ARGOTRANSA	http://argotransa.lt/	-	
		BALTIC MARINE SERVICES	https://bms-overseas.com/	-	
		BEGA	http://www.bega.lt/	-	
		CONTAINERSHIPS	http://www.containershipsgroup.com/	Sailing Schedule	http://www.containershipsgroup.com/schedules
		EUROCHEM LOGISTICS INTL	https://www.eurochem.lt/	-	
		EUROGA	http://euroga.lt/	-	
		FERTIMARA	https://fertimar.lt/en/	-	
		FRIGATE SHIPPING AGENCY	http://www.frigate.lt/	-	
		GOLFINAS SHIPPING AGENCY	http://www.golfinas.lt/	-	
		GREEN TERMINAL	https://www.green.lt/	Fleet, Vessel's arrival (vide)	https://www.green.lt/article/archive/175/
		LIMARKO GROUP	https://www.limarko.com/	Fleet	
		NAVISOLIS	https://navigatiosolis.lt/	-	
		NORDIS SHIPPING SERVICE	http://www.nordis.lt/	-	

NORSPEDA LOGISTICS	https://norspeda.it/	-	
TLC LOGISTICS	https://www.tlclogistics.it/	-	
WESTERN SHIPS AGENCY	http://www.wsa.it/en/	-	

Annexe 6 : Référentiel normalisé

Données "artificielles" rajoutées			Données d'identification				Données liées au voyage										
Date scraping	Port source	Page(s) scrapée(s)	Nom	Pavillon	N° IMO	Type	Statut	Prov.	Dest.	ETA, ATA, Date arrivée	ETD, ATD, Date départ	Terminal	Quai	March. déchargée	March. chargée	Agent	
x	Hambourg (DE)	Expected vessels	Shipname		IMO	Shiptype				ETA			Berth				
		Ship movements	Shipname		IMO	Shiptype	Direction			Time	Time						
x	Niedersachsen (DE)	Ships arrivals and departures	Schiffsname	Flag		Schiffsart	Status	Abfahrts hafen	Zielhafen	Ankunft	Abfahrt	Terminal	Kai				
x	Copenhague (DK)	Port Calls Oil	Ship	Nationality	IMO	Vessel type	Status	From2 (Arriving from)	Departing to	Arrival + ETA	Departure + time		Quay			Agency	
x	Aarhus (DK)	Vessels in port	Shipsname			Shiptype	Status			Exp. Arrival	Exp. Departure	Terminal	Pier				
x	Fredericia (DK)	Port Calls	Name				Status			ETA + Time	ETD + Time		Quay			Broker	
x	Bordeaux (FR)	Navires à quai	Nom	Pavillon		Type		Provenance	Destination	Accostage	Sortie prévue		Poste	Déchargement	Chargement	Agent montée	
		Navires attendus	Nom	Pavillon		Type		Provenance		Date prévue arrivée				Poste	Déchargement	Chargement	Agent
x	Dunkerque (FR)	Navires présents	Nom du navire					Provenance	Destination	Date d'entrée jetée	Prévision de départ		N° poste à quai + poste à quai			Agent consignataire	
		Navires attendus en entrée / transit	Nom du navire					Provenance	Destination	Prévision d'arrivée				Poste à quai prévu			Agent consignataire
		Navires attendus en sortie	Nom du navire						Destination			Prévision de départ					Agent consignataire
x	Le Havre (FR)	Arrivées	Bateau	Pavillon		Type		Provenance	Destination	Arrivée			Quai ? (infrastructure ?)			Consignataire	
		Départs	Bateau	Pavillon		Type		Provenance	Destination		Départ			Quai ? (infrastructure ?)			Consignataire

		Navires à quai	Navire	Pavillon		Type		Provenance	Destination	Accostage	Appareillage		Infrastructure ?			Consignataire
x	Amsterdam (NL)	Arrivals & Departures	Ship name	Flag			Ship status			ETA/ATA	ETD/ATD	Harbour	Berth			Agent
x	Rotterdam (NL)	Expected	Name	Flag code	IMO					ETA			Berth			Operator
		Present	Name	Flag code	IMO						ETD		Berth			Operator
		Departed	Name	Flag code	IMO						ETD		Berth			Operator
x	Klaipeda (LT)	List of vessels within the port	Ship	Flag	IMO no.					Date, time			Quay No.			Agent
		Planned calls	Ship	Flag	IMO no.					Date, time			Quay No.			Agent
		Last calls	Ship	Flag	IMO no.					Date, time			Quay No.			Agent
		Last departure	Ship	Flag	IMO no.						Date, time		Quay No.			Agent

Annexe 7 : Scripts des *spiders* de Scrapy

Aarhus / *vessels_in_port*

```
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour). Des cas ont été rencontrés pour Aarhus.
import dateparser

class VesselsInPortSpider(scrapy.Spider):
    name = 'vessels_in_port'
    allowed_domains = ['aarhushavn.dk']
    start_urls = ['https://www.aarhushavn.dk/en/ship_calls/vessels_in_port/']

    def parse(self, response):
        for ships in response.xpath('//*[@id="subcontentwrapper"]/div[3]/div'):
            nom = ships.xpath('div[2]/div[2]/span/text()').extract_first()
            if nom == None:
                continue
            yield{
                'Port source' : "Aarhus (DK)",
                'Statut' : ships.xpath('div[5]/div[2]/span/text()').extract_first().replace("\xa0", "").replace("Expected", "Attendu").replace("Arrived", "À quai"),
                'Nom' : nom.replace("\xa0", ""),
                'Pavillon' : "",
                'IMO' : "",
                'Type' : ships.xpath('div[3]/div[2]/span/text()').extract_first().replace("\xa0", ""),
                'Provenance' : "",
                'Destination' : "",
                'ETA' : dateparser.parse(ships.xpath('div[4]/div[2]/span/text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'ETD' : dateparser.parse(ships.xpath('div[6]/div[2]/span/text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'Terminal' : ships.xpath('preceding-sibling::div[@class="portarea"][position() < 2]/b/text()').extract_first(),
                'Quai' : ships.xpath('div[1]/div[2]/span/text()').extract_first().replace("\xa0", ""),
                'Marchandise chargée' : "",
                'Marchandise déchargée' : "",
                'Agent' : "",
            }
}
```

Amsterdam / arrivals_departures

```
import scrapy
from scrapy_splash import SplashRequest
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour). Des cas ont été rencontrés pour Amsterdam.
import dateparser

class ArrivalsDeparturesSpider(scrapy.Spider):
    name = 'arrivals_departures'

    allowed_domains = ['portofamsterdam.com']
    start_urls = ['https://myport.portofamsterdam.com/arrivals/']

    def start_requests(self):
        yield SplashRequest(
            url='https://myport.portofamsterdam.com/arrivals/',
            callback=self.parse,
            args={"wait": 5},
        )

    def parse(self, response):
        for ships in response.xpath('//main/div/ul/li')[1:]:
            #On ne veut pas prendre le début du texte "Current position :", on supprime donc les 16 premiers caractères
            statut = ships.xpath('div[1]/div[1]/span/span/text()').re_first('.{16}{.*}')
            #Les en-têtes des temps d'arrivée et de départ varient selon le statut du bateau
            if statut == "Arriving" or statut == "Shifting":
                ETA_time = dateparser.parse(ships.xpath('div[2]/div[2]/ul[1]/li[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M")
                ETD_time = dateparser.parse(ships.xpath('div[2]/div[2]/ul[1]/li[2]/text()').extract_first()).strftime("%Y-%m-%d %H:%M")
            elif statut == "Departing":
                ETA_time = ""
                ETD_time = dateparser.parse(ships.xpath('div[2]/div[2]/ul[1]/li[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M")
            elif statut == "In_port":
                ETA_time = dateparser.parse(ships.xpath('div[2]/div[2]/ul[1]/li[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M")
                ETD_time = ""
            yield {
                'Port source' : "Amsterdam (NL)",
                'Statut': statut.replace("In_port", "À quai").replace("Departing", "Partant").replace("Arriving", "Attendu").replace("Shifting", "En déplacement"),
```

```
'Nom': ships.xpath('div[1]/div[2]/span/text()').extract_first(),
'Pavillon': ships.xpath('normalize-space(div[2]/div[2]/ul[2]/li/span[contains(., "Flag:")] /following-sibling::node())').extract_first(),
'IMO' : "",
'Type' : "",
'Provenance' : "",
'Destination' : "",
'ETA': ETA_time,
'ETD': ETD_time,
'Terminal': ships.xpath('normalize-space(div[1]/div[4]/text()').extract_first(),
'Quai': ships.xpath('normalize-space(div[1]/div[5]/text()').extract_first(),
'Marchandise déchargée' : "",
'Marchandise chargée' : "",
'Agent': ships.xpath('normalize-space(div[2]/div[2]/ul[2]/li/span[contains(., "Agent:")] /following-sibling::node())').extract_first(),
}
```

Bordeaux / navires_attendus

```
# -*- coding: utf-8 -*-
import scrapy
from scrapy_splash import SplashRequest
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour Bordeaux.
import dateparser

class NaviresAttendusSpider(scrapy.Spider):
    name = 'navires_attendus'
    allowed_domains = ['bordeaux-port.fr']
    start_urls = ['https://www.bordeaux-port.fr/fr/navires-quais-attendus']

    def start_requests(self):
        yield SplashRequest(
            url='https://www.bordeaux-port.fr/fr/navires-quais-attendus',
            callback=self.parse,
            args={"wait": 3},
        )

    def parse(self, response):
        ships = response.xpath('//div[@class="pbx-navire-attendu"]')
        for ship in ships:
            ETD = ship.xpath('table/tbody[2]/tr[1]/td[2]/text()').extract_first()
            if ETD :
                ETD = dateparser.parse(ETD).strftime("%Y-%m-%d %H:%M")
            yield {
                'Port source' : "Bordeaux (FR)",
                'Statut' : "Attendu / A quai",
                'Nom' : ship.xpath('table//tr/td[@class="pbx-navire-name"]//text()').extract_first(),
                'Pavillon' : ship.xpath('table/tbody[2]/tr[4]/td[2]/text()').extract_first(),
                'IMO' : "",
                'Type de bateau' : ship.xpath('table/tbody[3]/tr[2]/td[2]/text()').extract_first(),
                'Provenance' : ship.xpath('table/tbody[2]/tr[2]/td[1]//text()').re_first('^.(+)( \\(\\w+\\))'),
                'Destination' : "",
                'ETA' : dateparser.parse(ship.xpath('table/tbody[2]/tr[1]/td[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
```

```
'ETD' : ETD,  
'Terminal' : "",  
'Quai' : "",  
'Marchandise déchargée' : ship.xpath('table/tbody[2]/tr[3]/td[1]/text()').extract_first(),  
'Marchandise chargée' : ship.xpath('table/tbody[2]/tr[3]/td[2]/text()').extract_first(),  
'Agent' : ship.xpath('table/tbody[2]/tr[2]/td[2]/text()').extract_first(),  
'Armateur' : ship.xpath('table/tbody[2]/tr[4]/td[1]/text()').extract_first(),  
}
```

Copenhague / port_calls_oil

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Aucun cas n'a été rencontré pour Copenhague.
import dateparser

class PortCallsOilSpider(scrapy.Spider):
    name = 'port_calls_oil'
    allowed_domains = ['cmport.com']
    start_urls = ['http://www.cmport.com/ships-in-port/oil/copenhagen']

    def parse(self, response):
        for ships in response.xpath('//tr[@class="toggleVesselDetails"]'):
            #Difficile de parser les dates car elle se trouvent dans des champs qui ne sont pas bien analysés par dateparser
            ETA_date = dateparser.parse(ships.xpath('td[1]/span//text()').extract_first()).strftime("%Y-%m-%d")
            ETA_time = ships.xpath('td[2]/span//text()').extract_first()
            ETA_time2 = ETA_time[:2] + ":" + ETA_time[2:]
            ETA = ETA_date + " " + ETA_time2

            ETD_date = dateparser.parse(ships.xpath('following-sibling::tr[3]/td/table/tr[2]/td[3]/span//text()').extract_first()).strftime("%Y-%m-%d")
            ETD_time = ships.xpath('following-sibling::tr[3]/td/table/tr[2]/td[4]/span//text()').extract_first()
            ETD_time2 = ETD_time[:2] + ":" + ETD_time[2:]
            ETD = ETD_date + " " + ETD_time2

            yield {
                'Port source' : "Copenhague (DK)",
                'Statut' : ships.xpath('td/img/@title').extract_first().replace("Expected", "Attendu").replace("Arrived", "À quai"),
                'Nom' : ships.xpath('td[3]/span//text()').extract_first(),
                'Pavillon' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[4]/td[4]/span//text()').extract_first(),
                'IMO' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[3]/td[4]/span//text()').extract_first(),
                '#Call sign' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[2]/td[4]/span//text()').extract_first(),
                'Type' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[6]/td[4]/span//text()').extract_first(),
                'Provenance' : ships.xpath('td[4]/span//text()').extract_first(),
                'Destination' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[9]/td[2]/span//text()').extract_first(),
                'ETA' : ETA,
                'ETD' : ETD,
```

```
'Terminal' : ships.xpath('following-sibling::tr/td/table[@class="vesselDetailsTable"]/tr[12]/td[4]/span//text()).extract_first(),
'Quai' : ships.xpath('td[5]/span//text()).extract_first(),
'Marchandise chargée' : "",
'Marchandise déchargée' : "",
'Agent' : "",
}
```

Dunkerque / navires_attendus

```
# -*- coding: utf-8 -*-
import scrapy
from scrapy_splash import SplashRequest
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour Dunkerque.
import dateparser

class NaviresPresentsSpider(scrapy.Spider):
    name = 'navires_attendus'

    allowed_domains = ['dunkerque-port.fr']
    start_urls = ['http://www.dunkerque-port.fr/fr/activites-commerciales/navires-entree-transit.html']

    def start_requests(self):
        yield SplashRequest(
            url='http://www.dunkerque-port.fr/fr/activites-commerciales/navires-entree-transit.html',
            callback=self.parse,
            args={"wait": 3},
        )

    def parse(self, response):
        for ships in response.xpath('//*[@class="normal removeborder"]//tbody/tr')[1:]:
            yield {
                'Port source' : "Dunkerque (FR)",
                'Statut' : "Attendu",
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : "",
                'IMO' : "",
                'Type' : "",
                'Provenance' : ships.xpath('td[5]//text()').extract_first(),
                'Destination' : ships.xpath('td[7]//text()').extract_first(),
                'ETA' : dateparser.parse(ships.xpath('td[1]//text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'ETD' : "",
                'Terminal' : "",
                'Quai' : ships.xpath('td[6]//text()').extract_first(),
            }
```

```
'Marchandise chargée' : "",  
'Marchandise déchargée' : "",  
'Agent' : ships.xpath('td[4]//text()').extract_first(),  
}
```

Dunkerque / navires_partants

```
# -*- coding: utf-8 -*-
import scrapy
from scrapy_splash import SplashRequest
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour Dunkerque.
import dateparser

class NaviresPresentsSpider(scrapy.Spider):
    name = 'navires_partants'

    allowed_domains = ['dunkerque-port.fr']
    start_urls = ['http://www.dunkerque-port.fr/fr/activites-commerciales/navires-sortie.html']

    def start_requests(self):
        yield SplashRequest(
            url='http://www.dunkerque-port.fr/fr/activites-commerciales/navires-sortie.html',
            callback=self.parse,
            args={"wait": 3},
        )

    def parse(self, response):
        for ships in response.xpath('//*[@class="normal removeborder"]//tbody/tr')[1:]:
            yield {
                'Port source' : "Dunkerque (FR)",
                'Statut' : "Partant",
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : "",
                'IMO' : "",
                'Type' : "",
                'Provenance' : "",
                'Destination' : ships.xpath('td[5]//text()').extract_first(),
                'ETA' : "",
                'ETD' : dateparser.parse(ships.xpath('td[1]//text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'Terminal' : "",
                'Quai' : "",
            }
```

```
'Marchandise chargée' : "",  
'Marchandise déchargée' : "",  
'Agent' : ships.xpath('td[4]//text()').extract_first(),  
}
```

Dunkerque / navires_presents

```
# -*- coding: utf-8 -*-
import scrapy
from scrapy_splash import SplashRequest
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interversion du mois et du jour).
#Des cas ont été rencontrés pour Dunkerque.
import dateparser

class NaviresPresentsSpider(scrapy.Spider):
    name = 'navires_presents'

    allowed_domains = ['dunkerque-port.fr']
    start_urls = ['http://www.dunkerque-port.fr/fr/activites-commerciales/navires-presents-port-dunkerque.html']

    def start_requests(self):
        yield SplashRequest(
            url='http://www.dunkerque-port.fr/fr/activites-commerciales/navires-presents-port-dunkerque.html',
            callback=self.parse,
            args={"wait": 3},
        )

    def parse(self, response):
        for ships in response.xpath('//*[@class="normal removeborder"]/tbody/tr')[1:]:
            #Si la date estimée de départ existe
            if ships.xpath('td[10]//text()'):
                ETD = dateparser.parse(ships.xpath('td[10]//text()').extract_first()).strftime("%Y-%m-%d %H:%M")
            else:
                ETD = ""
            yield {
                'Port source' : "Dunkerque (FR)",
```

```
'Statut' : "À quai",
'Nom' : ships.xpath('td[4]//text()').extract_first(),
'Pavillon' : "",
'IMO' : "",
'Type' : "",
'Provenance' : ships.xpath('td[8]//text()').extract_first(),
'Destination' : ships.xpath('td[9]//text()').extract_first(),
'ETA' : dateparser.parse(ships.xpath('td[7]//text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
'ETD' : ETD,
'Terminal' : "",
'Quai' : ships.xpath('td[2]//text()').extract_first(),
'Marchandise chargée' : "",
'Marchandise déchargée' : "",
'Agent' : ships.xpath('td[6]//text()').extract_first(),
}
```

Fredericia / port_calls

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour Fredericia.
import dateparser

class PortCallsSpider(scrapy.Spider):
    name = 'port_calls'
    allowed_domains = ['https://www.adp-as.dk/']
    start_urls = ['https://www.adp-as.dk/en/maritime-services/calls/']

    def parse(self, response):
        for ships in response.xpath('//div[2]/div/table/tbody/tr'):
            #Pour préciser que le champ agent est en fait un courtier
            if ships.xpath('td[8]//text()').extract_first():
                agent = ships.xpath('td[8]//text()').extract_first() + " (broker)"
            else:
                agent = ""
            yield {
                'Port Source' : "Fredericia (DK)",
                'Statut' : ships.xpath('td[3]//text()').extract_first().replace("Arrived", "À quai").replace("Expected", "Attendu"),
                'Nom' : ships.xpath('td[1]//text()').extract_first(),
                'Pavillon' : "",
                'IMO' : "",
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : dateparser.parse(ships.xpath('td[4]//text()').extract_first() + " " + ships.xpath('td[5]//text()').extract_first() + ":00").strftime("%Y-%m-%d %H:%M"),
                'ETD' : dateparser.parse(ships.xpath('td[6]//text()').extract_first() + " " + ships.xpath('td[7]//text()').extract_first() + ":00").strftime("%Y-%m-%d %H:%M"),
                'Terminal' : "",
                'Quai' : ships.xpath('td[2]//text()').extract_first(),
                'Marchandise chargée' : "",
                'Marchandise déchargée' : "",
                'Agent' : agent,
            }
        }
```

Hambourg / *ships_expected*

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (intersion du mois et du jour).
#Pas de cas pour Hambourg.
import dateparser
from scrapy_splash import SplashRequest

# pour permettre le infinite scrolling, on détermine le nombre scoll (num_scrolls) qu'on veut,
# en sachant qu'un scroll donne 20 résultats et qu'on peut aller jusqu'à 56406 résultats (testé le 12 juin)
# https://stackoverflow.com/questions/40325657/how-does-scrapy-splash-handle-infinite-scrolling
script = """
function main(splash)
  local num_scrolls = 20
  local scroll_delay = 1.0

  local scroll_to = splash:jsfunc("window.scrollTo")
  local get_body_height = splash:jsfunc(
    "function() {return document.body.scrollHeight;}"
  )
  assert(splash:go(splash.args.url))
  splash:wait(splash.args.wait)

  for _ = 1, num_scrolls do
    scroll_to(0, get_body_height())
    splash:wait(scroll_delay)
  end
  return splash:html()
end
"""

class ShipExpectedSpider(scrapy.Spider):
  name = 'ships_expected'
  allowed_domains = ['hafen-hamburg.de']
  start_urls = ['https://www.hafen-hamburg.de/en/vessels/eta']
```

```

def start_requests(self):
    yield SplashRequest(
        url='https://www.hafen-hamburg.de/en/vessels/eta',
        callback=self.parse,
        endpoint='execute',
        args={"wait": 3, 'lua_source': script}, #on attend 3 secondes pour permettre au JS de se charger correctement et on appelle le script pour gérer le scroll infini
    )

def parse(self, response):
    ships = response.xpath('//li[@class="col-25 tile-box-entry"']
    for ship in ships:
        yield{
            'Port source' : "Hambourg (DE)",
            'Statut' : "Attendu",
            'Nom' : ship.xpath('normalize-space(ul/li[2]/a/text)').extract_first(),
            'Pavillon' : "",
            'IMO' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[1]/text)').extract_first(),
            'Type' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[2]/text)').extract_first(),
            'Provenance' : "",
            'Destination' : "",
            'ETA' : dateparser.parse(ship.xpath('normalize-space(ul/li[3]/a/dl/dd[5]/text)').extract_first()).strftime("%Y-%m-%d %H:%M"),
            'ETD' : "",
            'Terminal' : "",
            'Quai' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[6]/text)').extract_first(),
            'Marchandise chargée' : "",
            'Marchandise déchargée' : "",
            'Agent' : "",
        }

```

Hambourg / *ships_movements*

```
# -*- coding: utf-8 -*-
import scrapy
import dateparser
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Pas de cas pour Hambourg.a
from scrapy_splash import SplashRequest

# pour permettre le infinite scrolling, on détermine le nombre scoll (num_scrolls) qu'on veut,
# en sachant qu'un scroll donne 20 résultats et qu'on peut aller jusqu'à 56406 résultats (testé le 12 juin)
# https://stackoverflow.com/questions/40325657/how-does-scrapy-splash-handle-infinite-scrolling
script = """
function main(splash)
  local num_scrolls = 20
  local scroll_delay = 1.0

  local scroll_to = splash:jsfunc("window.scrollTo")
  local get_body_height = splash:jsfunc(
    "function() {return document.body.scrollHeight;}"
  )
  assert(splash:go(splash.args.url))
  splash:wait(splash.args.wait)

  for _ = 1, num_scrolls do
    scroll_to(0, get_body_height())
    splash:wait(scroll_delay)
  end
  return splash:html()
end
"""

class ShipMovementsSpider(scrapy.Spider):
  name = 'ships_movements'
  allowed_domains = ['hafen-hamburg.de']
  start_urls = ['https://www.hafen-hamburg.de/en/vessels/ship-movements']
```

```

def start_requests(self):
    yield SplashRequest(
        url='https://www.hafen-hamburg.de/en/vessels/ship-movements',
        callback=self.parse,
        endpoint='execute',
        args={"wait": 3, 'lua_source': script}, #on attend 3 secondes pour permettre au JS de se charger correctement et on appelle le script pour gérer le scroll infini
    )

def parse(self, response):
    ships = response.xpath('//li[@class="col-25 tile-box-entry"']
    for ship in ships:
        yield{
            'Port source' : "Hambourg (DE)",
            'Statut' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[4]/text())').extract_first().replace("Arrived", "À quai").replace("Departed", "Parti"),
            'Nom' : ship.xpath('normalize-space(ul/li[2]/a/text())').extract_first(),
            'Pavillon' : "",
            'IMO' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[1]/text())').extract_first(),
            'Type' : ship.xpath('normalize-space(ul/li[3]/a/dl/dd[2]/text())').extract_first(),
            'Provenance' : "",
            'Destination' : "",
            'ETA' : dateparser.parse(ship.xpath('normalize-space(ul/li[3]/a/dl/dd[5]/text())').extract_first()).strftime("%Y-%m-%d %H:%M") if ship.xpath('normalize-space(ul/li[3]/a/dl/dd[4]/text())').extract_first() == "Arrived" else None,
            'ETD' : dateparser.parse(ship.xpath('normalize-space(ul/li[3]/a/dl/dd[5]/text())').extract_first()).strftime("%Y-%m-%d %H:%M") if ship.xpath('normalize-space(ul/li[3]/a/dl/dd[4]/text())').extract_first() == "Departed" else None,
            'Terminal' : "",
            'Quai' : "",
            'Marchandise chargée' : "",
            'Marchandise déchargée' : "",
            'Agent' : "",
        }

```

Klaipeda / last_calls

```
# -*- coding: utf-8 -*-
import scrapy

class LastCallsPySpider(scrapy.Spider):
    name = 'last_calls'
    allowed_domains = ['portofklaipeda.lt']
    start_urls = ['https://www.portofklaipeda.lt/last-calls']

    def parse(self, response):
        for ships in response.xpath('//*[@contains(concat(" ", @class, " "), concat(" ", "ships", " "))]/tbody/tr):
            yield {
                'Port source' : "Klaipeda (LT)",
                'Statut' : 'Arrivé',
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : ships.xpath('td[4]//text()').extract_first(),
                'IMO' : ships.xpath('td[3]//text()').extract_first(),
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : ships.xpath('td[5]//text()').extract_first(),
                'ETD' : "",
                'Terminal' : "",
                'Quai' : ships.xpath('td[7]//text()').extract_first(),
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships.xpath('td[6]//text()').extract_first().replace("", " ").replace(', ', ' '),
            }
```

Klaipeda / last_departures

```
# -*- coding: utf-8 -*-
import scrapy

class LastDeparturesSpider(scrapy.Spider):
    name = 'last_departures'
    allowed_domains = ['portofklaipeda.lt']
    start_urls = ['https://www.portofklaipeda.lt/last-departures']

    def parse(self, response):
        for ships in response.xpath('//*[@contains(concat(" ", @class, " "), concat(" ", "ships", " "))]/tbody/tr'):
            yield {
                'Port source' : "Klaipeda (LT)",
                'Statut' : "Parti",
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : ships.xpath('td[4]//text()').extract_first(),
                'IMO' : ships.xpath('td[3]//text()').extract_first(),
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : "",
                'ETD' : ships.xpath('td[5]//text()').extract_first(),
                'Terminal' : "",
                'Quai' : ships.xpath('td[7]//text()').extract_first(),
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships.xpath('td[6]//text()').extract_first().replace("", " ").replace(',', ' '),
            }
```

Klaipeda / *planned_calls*

```
# -*- coding: utf-8 -*-
import scrapy

class PlannedCallsSpider(scrapy.Spider):
    name = 'planned_calls'
    allowed_domains = ['portofklaipeda.lt']
    start_urls = ['https://www.portofklaipeda.lt/planned-calls']

    def parse(self, response):
        for ships in response.xpath('//*[@contains(concat(" ", @class, " "), concat(" ", "ships", " "))]/tbody/tr):
            yield {
                'Port source' : "Klaipeda (LT)",
                'Statut' : "Attendu",
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : ships.xpath('td[4]//text()').extract_first(),
                'IMO' : ships.xpath('td[3]//text()').extract_first(),
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : ships.xpath('td[5]//text()').extract_first(),
                'ETD' : "",
                'Terminal' : "",
                'Quai' : ships.xpath('td[7]//text()').extract_first(),
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships.xpath('td[6]//text()').extract_first().replace('"', '').replace(',', ' '),
            }
```

Klaipeda / *planned_calls*

```
# -*- coding: utf-8 -*-
import scrapy

class VesselsWithinPortSpider(scrapy.Spider):
    name = 'vessels_within_port'
    allowed_domains = ['portofklaipeda.lt']
    start_urls = ['http://portofklaipeda.lt/list-of-vessels-within-the-port']

    def parse(self, response):
        for ships in response.xpath('//*[@contains(concat(" ", @class, " "), concat(" ", "ships", " "))]/tbody/tr'):
            yield {
                'Port source' : "Klaipeda (LT)",
                'Statut' : "À quai",
                'Nom' : ships.xpath('td[2]//text()').extract_first(),
                'Pavillon' : ships.xpath('td[4]//text()').extract_first(),
                'IMO' : ships.xpath('td[3]//text()').extract_first(),
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : ships.xpath('td[5]//text()').extract_first(),
                'ETD' : "",
                'Terminal' : "",
                'Quai' : ships.xpath('td[7]//text()').extract_first(),
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships.xpath('td[6]//text()').extract_first().replace("", " ").replace(',', ' '),
            }
```

Le Havre / arrivées

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour le Havre.
import dateparser
```

```
class ArriveesSpider(scrapy.Spider):
    name = 'arrivees'
    allowed_domains = ['havre-port.com']
    start_urls = ['https://www.havre-port.com/fr/liste/128/arrivees-au-grand-port-maritime-du-havre']

    def parse(self, response):
        for href in response.xpath('//tr/td/a/@href'):
            url = response.urljoin(href.extract())
            yield scrapy.Request(url, callback = self.parse_dir_contents)

    def parse_dir_contents(self, response):
        for ships in response.xpath('//section[2]/div'):
            #On prend l'heure d'arrivée prévue si l'heure d'arrivée effective n'existe pas
            if response.xpath('//table[2]/tr/td[4]/text()').extract_first() :
                eta_path = 'table[2]/tr/td[4]/text()'
            else:
                eta_path = 'table[2]/tr/td[1]/text()'
            yield{
                'Port source' : "Le Havre (FR)",
                'Statut' : "Attendu",
                'Nom' : ships.xpath('table[6]/tr[1]/td[1]/div/div[2]/text()').extract_first(),
                'Pavillon' : ships.xpath('table[6]/tr[2]/td[1]/div/div[2]/text()').re_first('{6}(.*)'),
                'IMO' : ships.xpath('table[6]/tr[1]/td[2]/div/div[2]/text()').extract_first(),
                'Type' : ships.xpath('table[6]/tr[2]/td[2]/div/div[2]/text()').re_first('{6}(.*)'),
                'Provenance' : ships.xpath('table[5]/tr/td[2]/text()').re_first('{8}(.*)'),
                'Destination' : ships.xpath('table[5]/tr/td[3]/text()').re_first('{8}(.*)'),
                'ETA' : dateparser.parse(ships.xpath(eta_path).extract_first()).strftime("%Y-%m-%d %H:%M"),
                'ETD' : dateparser.parse(ships.xpath('table[3]/tr/td[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
```

```
'Terminal' : ships.xpath('table[4]/tr/td[3]/text()').extract_first().replace("-", ""),
'Quai' : ships.xpath('table[4]/tr/td[2]/text()').extract_first(),
'Marchandise déchargée' : "",
'Marchandise chargée' : "",
'Agent' : "",
}
```

Le Havre / départs

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour le Havre.
import dateparser

class DepartsSpider(scrapy.Spider):
    name = 'departs'
    allowed_domains = ['havre-port.com']
    start_urls = ['https://www.havre-port.com/fr/liste/129/departs-au-grand-port-maritime-du-havre']

    def parse(self, response):
        for href in response.xpath('//tr/td/a/@href'):
            url = response.urljoin(href.extract())
            yield scrapy.Request(url, callback = self.parse_dir_contents)

    def parse_dir_contents(self, response):
        for ships in response.xpath('//section[2]/div'):
            #On prend l'heure d'arrivée prévue si l'heure d'arrivée effective n'existe pas
            if response.xpath('//table[2]/tr/td[4]/text()').extract_first() :
                eta_path = 'table[2]/tr/td[4]/text()'
            else:
                eta_path = 'table[2]/tr/td[1]/text()'
            yield{
                'Port source' : "Le Havre (FR)",
                'Statut' : "Parti",
                'Nom' : ships.xpath('table[6]/tr[1]/td[1]/div/div[2]/text()').extract_first(),
                'Pavillon' : ships.xpath('table[6]/tr[2]/td[1]/div/div[2]/text()').re_first('{6}(.*)'),
                'IMO' : ships.xpath('table[6]/tr[1]/td[2]/div/div[2]/text()').extract_first(),
                'Type' : ships.xpath('table[6]/tr[2]/td[2]/div/div[2]/text()').re_first('{6}(.*)'),
                'Provenance' : ships.xpath('table[5]/tr/td[2]/text()').re_first('{8}(.*)'),
                'Destination' : ships.xpath('table[5]/tr/td[3]/text()').re_first('{8}(.*)'),
                'ETA' : dateparser.parse(ships.xpath(eta_path).extract_first()).strftime("%Y-%m-%d %H:%M"),
                'ETD' : dateparser.parse(ships.xpath('table[3]/tr/td[1]/text()').extract_first()).strftime("%Y-%m-%d %H:%M"),
```

```
'Terminal' : ships.xpath('table[4]/tr/td[3]/text()').extract_first().replace("-", ""),
'Quai' : ships.xpath('table[4]/tr/td[2]/text()').extract_first(),
'Marchandise déchargée' : "",
'Marchandise chargée' : "",
'Agent' : "",
}
```

Niedersachsen / arrivals_departures

```
# -*- coding: utf-8 -*-
import scrapy
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour).
#Des cas ont été rencontrés pour Niedersachsen.
import dateparser

class ArrivalsDeparturesSpider(scrapy.Spider):
    name = 'arrivals_departures'
    allowed_domains = ['nports.de']
    start_urls = ['https://www.nports.de/en/ship-arrivals-and-departures/']

    def parse(self, response):
        ships = response.xpath('//table/tr')
        for ship in ships[1:]:
            yield {
                'Port source' : "Niedersachsen (DE)",
                'Statut' : ship.xpath('@data-status').extract_first().replace("abgefahren", "Parti").replace("angekommen", "À quai").replace("avisiert", "Attendu"),
                'Nom' : ship.xpath('td[@class="rz name"]/a/span[2]//text()').re_first('^(.+)( \\(\\w+\\))'), #regex pour ne pas prendre la longueur entre ()
                'Pavillon' : ship.xpath('td[@class="flagge"]//text()').extract_first(),
                '#Call sign' : ships.xpath('td[@class="rz name"]/a/span//text()').extract_first(),
                'IMO' : "",
                'Type' : ship.xpath('td[@class="typ"]//text()').extract_first(),
                'Provenance' : ship.xpath('td[@class="sport"]//text()').extract_first(),
                'Destination' : ship.xpath('td[@class="tport"]//text()').extract_first(),
                'ETA' : dateparser.parse(ship.xpath('normalize-space(td[@class="an ab"]//text())').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'ETD' : dateparser.parse(ship.xpath('normalize-space(td[@class="an ab"]//text())[2]').extract_first()).strftime("%Y-%m-%d %H:%M"),
                'Terminal' : ship.xpath('@data-terminal').extract_first(),
                'Quai' : ships.xpath('td[@class="kai"]//text()').extract_first(),
                'Marchandise chargée' : "",
                'Marchandise déchargée' : "",
                'Agent' : "",
                '#Jauge brute' : ships.xpath('td[@class="brz"]//text()').extract_first(),
                '#Longueur' : ships.xpath('td[@class="rz name"]/a/span[2]//text()').re_first('\\d+m'), #regex pour prendre la longueur entre ()
            }
        }
```

Rotterdam / *departed_ships*

```
# -*- coding: utf-8 -*-
import scrapy
import json
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (intersion du mois et du jour). Pas de cas pour le port de Rotterdam.
import dateparser

class PresentShipsSpider(scrapy.Spider):
    name = 'departed_ships'
    allowed_domains = ['portofrotterdam.com']
    start_urls = ['https://www.portofrotterdam.com/ship_visits/all']

    def parse(self, response):
        jsonresponse = json.loads(response.text)
        for ships in jsonresponse["Departed"]:
            yield {
                'Port source' : "Rotterdam (NL)",
                #On rajoute manuellement le statut, basé sur le membre JSON Departed
                'Statut' : "Parti",
                'Nom' : ships['name'],
                'Pavillon' : ships['flag_code'],
                'IMO' : ships['imo'],
                #'UCRN' : ships['ucrn'],
                #'MMSI' : ships['mmsi'],
                #'Call sign' : ships['call_sign'],
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : "",
                'ETD' : dateparser.parse(ships['datetime']).strftime("%Y-%m-%d %H:%M"),
                'Terminal' : "",
                'Quai' : ships['berth'],
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships['operator'],
                #'Port_of_registry' : ships['port_of_registry'],
```

```
#'Gross_tonnage' : ships['gross_tonnage'],  
#'Nett_tonnage' : ships['nett_tonnage'],  
#'Deadweight' : ships['deadweight'],  
#'Length' : ships['length'],  
#'Breadth' : ships['breadth'],  
#'Draught' : ships['draught'],  
}
```

Rotterdam / *expected_ships*

```
# -*- coding: utf-8 -*-
import scrapy
import json
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interverson du mois et du jour). Pas de cas pour le port de Rotterdam.
import dateparser

class PresentShipsSpider(scrapy.Spider):
    name = 'expected_ships'
    allowed_domains = ['portofrotterdam.com']
    start_urls = ['https://www.portofrotterdam.com/ship_visits/all']

    def parse(self, response):
        jsonresponse = json.loads(response.text)
        for ships in jsonresponse["Expected"]:
            yield {
                'Port source' : "Rotterdam (NL)",
                #On rajoute manuellement le statut, basé sur le membre JSON Expected
                'Statut' : "Attendu",
                'Nom' : ships['name'],
                'Pavillon' : ships['flag_code'],
                'IMO' : ships['imo'],
                #'UCRN' : ships['ucrn'],
                #'MMSI' : ships['mmsi'],
                #'Call sign' : ships['call_sign'],
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : dateparser.parse(ships['datetime']).strftime("%Y-%m-%d %H:%M"),
                'ETD' : "",
                'Terminal' : "",
                'Quai' : ships['berth'],
                'Marchandise déchargée' : "",
                'Marchandise chargée' : "",
                'Agent' : ships['operator'],
                #'Port_of_registry' : ships['port_of_registry'],
```

```
#'Gross_tonnage' : ships['gross_tonnage'],  
#'Nett_tonnage' : ships['nett_tonnage'],  
#'Deadweight' : ships['deadweight'],  
#'Length' : ships['length'],  
#'Breadth' : ships['breadth'],  
#'Draught' : ships['draught'],  
}
```

Rotterdam / *present_ships*

```
# -*- coding: utf-8 -*-
import scrapy
import json
#Attention, l'utilisation de dateparser créé parfois des erreurs dans le champ date (interversion du mois et du jour). Pas de cas pour le port de Rotterdam.
import dateparser

class PresentShipsSpider(scrapy.Spider):
    name = 'present_ships'
    allowed_domains = ['portofrotterdam.com']
    start_urls = ['https://www.portofrotterdam.com/ship_visits/all']

    def parse(self, response):
        jsonresponse = json.loads(response.text)
        for ships in jsonresponse["Present"]:
            #Pour éviter qu'un champ sans date fasse planter le module dateparser
            if ships['datetime'] == "Unknown" :
                datetime = ships['datetime']
            else:
                datetime = dateparser.parse(ships['datetime']).strftime("%Y-%m-%d %H:%M")
            yield {
                'Port source' : "Rotterdam (NL)",
                #On rajoute manuellement le statut, basé sur le membre JSON Present
                'Statut' : "À quai",
                'Nom' : ships['name'],
                'Pavillon' : ships['flag_code'],
                'IMO' : ships['imo'],
                #'UCRN' : ships['ucrn'],
                #'MMSI' : ships['mmsi'],
                #'Call sign' : ships['call_sign'],
                'Type' : "",
                'Provenance' : "",
                'Destination' : "",
                'ETA' : "",
                'ETD' : datetime,
                'Terminal' : "",
            }
```

```
'Quai' : ships['berth'],  
'Marchandise déchargée' : "",  
'Marchandise chargée' : "",  
'Agent' : ships['operator'],  
# 'Port_of_registry' : ships['port_of_registry'],  
# 'Gross_tonnage' : ships['gross_tonnage'],  
# 'Nett_tonnage' : ships['nett_tonnage'],  
# 'Deadweight' : ships['deadweight'],  
# 'Length' : ships['length'],  
# 'Breadth' : ships['breadth'],  
# 'Draught' : ships['draught'],  
}
```