Internship Activity Report: Contribution to finalizing the World Register of Marine Species (WoRMS)

Improving the knowledge of what lives in the world's oceans and seas

Lambert Elke
Master Oceans & Lakes
8-29 July 2019

INTRODUCING VLIZ (FLANDERS MARINE INSTITUTE)

In the period of 8 to the 29th of July 2019, I did my internship at the Flanders Marine Institute ('Vlaams Instituut voor de Zee' or short VLIZ). During this period, I was supervised by Thomas Lanssens (thomas.lanssens@vliz.be) and Leen Vandepitte (leen.vandepitte@vliz.be).

VLIZ can be defined as an autonomous and neutral non-profit organisation that is funded every year by the Flemish Government and the province of West Flanders. It serves as the main supporting institution for marine and coastal related research in Flanders. [1] The main focus areas are the ocean, seas, coast and associated tidal systems. VLIZ coordinates Flemish marine research, acts as an international contact point, supports international marine education and helps building a strong image for Flemish marine research and aim to make it more visible. [1] With their work they aspire to reach the scientific community, educational institutions, the general public and, in the light of blue economy, even the industry. One of their main goals includes a more innovative and multidisciplinary approach to marine sciences. [2]

Though a very young organisation (founded in 1999), VLIZ has a very rich backstory and has shown quick development and growth.

In 1843, Ostend housed the very first marine research station in the world. It was installed by Pierre-Joseph Van Beneden, a Zoology professor at the Catholic University of Leuven. Though it was a small and private research centre, word got around and its reputation got a lot of eminent scientists to come to Ostend and conduct marine related experiments. Thus, this first station became an important meeting place for scientists to discuss marine inspired topics inside and outside of Belgium for 30 years (**Figure 1 & 2**). [3] [4]
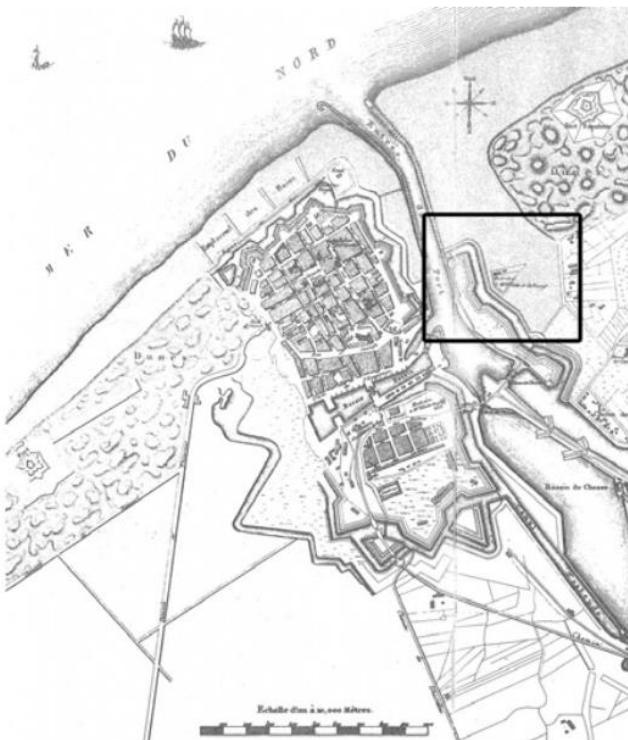

Figure 1: *A map showing the location of the first marine station in the world in Ostend, located on the east bank of the port at a distance of less than 200m from the present site of VLIZ. [4]*

Figure 2: *image of the first marine station in the world: the "Dune laboratory". [4]*

The marine station ceased to exist after 30 years and left Belgium without a marine station. This during the 1970's, when marine sciences were booming all over Europe. As the Belgian government was averse to invest in this growing field of expertise, others tried to establish a new station in order to keep up with the rest of Europe. Both the son of Pierre-Joseph, Edouard Van Beneden, and a former student of Pierre-Joseph, Gustave Gilson, attempted to establish their own research centres, but both were unsuccessful. Gilson finally convinced the Belgian Government of the need for a high-end marine research facility in 1927. This led to the birth of the organisation 'Zeewetenschappelijk Instituut (ZWI) / Institut d'Etudes Maritimes' in Ostend with Gilson as its general director. Gilson then finally managed to show its importance as ten years later, the ZWI was recognised by Royal Decree to be of great public importance. [3] The ZWI focussed on scientific research and statistical checks of the Belgian sea fisheries for over 30 years. After some financial problems and failed attempts to find the necessary funding, the ZWI was closed permanently in 1967.

Figure 3 (Above): Picture of the old IZWO offices where VLIZ operated in the first few years. Source: VLIZ

Figure 4 (Below): Current location of the VLIZ offices in the renovated warehouses of the fish mine in Ostend. Source: VLIZ

In 1970, the void was filled with the establishment of the Institute for Marine Scientific Research (IZWO) founded by the governor of West Flanders Pierre van Outryve d'Ydewalle and with the necessary financial help of the Francqui Foundation and the National Fund for Scientific Research. [5] It started out small, with 22 founding members and 13 directors. IZWO's vision then was quite like VLIZ's mission today: to coordinate and promote Belgian marine science and to boost the knowledge of the public. It became the start of an interinstitutional and multidisciplinary approach to study the Belgian seas and coasts. An important milestone in this view was the organisation of the '10th European Symposium on Marine Biology' where over 400 scientists from 40 countries came together in Ostend, giving IZWO international recognition. [5]

After 25 years, requests from inside the institution came asking for an organisation that could function as a fully-fledged international partner and is able to couple science, policy makers and the private sector. This wish was granted and led to the birth of VLIZ in 1999 with Dr. Jan Mees as General Director. [5] From this point on, VLIZ has only known growth and development

in employees but also in housing. They started out with a team of 10 permanent and two temporary employees in the old offices of the IZWO (**Figure 3**) to over 90 employees (*Source: Scheldeman Jessy*) today and a move to the renovated warehouses of the Ostend fish market, their present location (**Figure 4**). [6] Though, from 2022 onwards a completely new building will be housing VLIZ together with its (inter)national project partners and ILVO under the name InnovOcean Campus. Thus, becoming a landmark for marine sciences in Ostend.

As a growing organisation, their role in different aspects also grew. In order to be a supportive marine institution promoting marine and coastal sciences everywhere, they concluded a lot of cooperation agreements with (inter)national marine organisations, universities and research groups and are an active constituent in different national and international projects and networks. One of the major steps in this aspect was the creation of the InnovOcean site at the present warehouses of the fish mine in Ostend. Next to the VLIZ offices, these warehouses now also accommodate some outstanding international organisations and institutes, in order to promote multilateral cooperation between these partners and to support maritime research. Currently the site includes the IOC/UNESCO Project Office for IODE, the secretariat for the European Marine Board (EMB) and the European Marine Observation and Data Network (EMODnet) and regional offices of the Province of West Flanders. [7]

INTERNSHIP

Framework internship I : the LifeWatch project

During my internship at VLIZ, I performed a synchronisation between the WoRMS and AlgaeBase databases as a Belgian contribution to the LifeWatch project.

The LifeWatch project is a part of the European Strategy Forum on Research Infrastructure (ESFRI) and contributes to the European infrastructure for biodiversity data and observatories. This infrastructure will consist of scattered data of several databases and monitoring sites, a system of terrestrial, marine and freshwater biodiversity observatories, web services and modelling packages.   In total, it is a 'virtual laboratory'. The development of this integrated system already started in 2012. The main activities in this project is integrating existing systems, upgrading them and establishing new systems where needed. The aim is to facilitate the generation, the acquisition, the handling and the analysis of biodiversity data all over Europe. [8]

The Flanders Marine Institute (VLIZ) and the Research Institute for Nature and Forest (INBO) together coordinate the Flemish contribution to the LifeWatch project.

Because of its internationally recognized expertise in constructing taxonomic databases, VLIZ was chosen to set-up the 'LifeWatch Taxonomic Backbone' (TB). The aim of the TB is to facilitate the standardisation of species data and the (virtual) unification of the many dispersed biodiversity archives and data systems especially those focussing on one of these major components: taxonomy, distributional info, ecology, literature and genomic data. In the end, a data flow should be established between all the individual data systems that deal with one of these components in order to ensure that for example thematic (e.g. Fishbase) and global (e.g. WoRMS) databases continuously receive a supported flow of information and data from the regional and national databases in order to be as complete and correct as possible (**Figure 5**). [9] The Aphia platform (more information later) is a major contributor to the TB of the LifeWatch project, especially to the taxonomy component. During my internship, I explored the link between the thematic database AlgaeBase and the global database WoRMS (World Register of Marine Species). Both these databases focus on taxonomic information and contribute to the TB of the LifeWatch project in order to make it more complete.
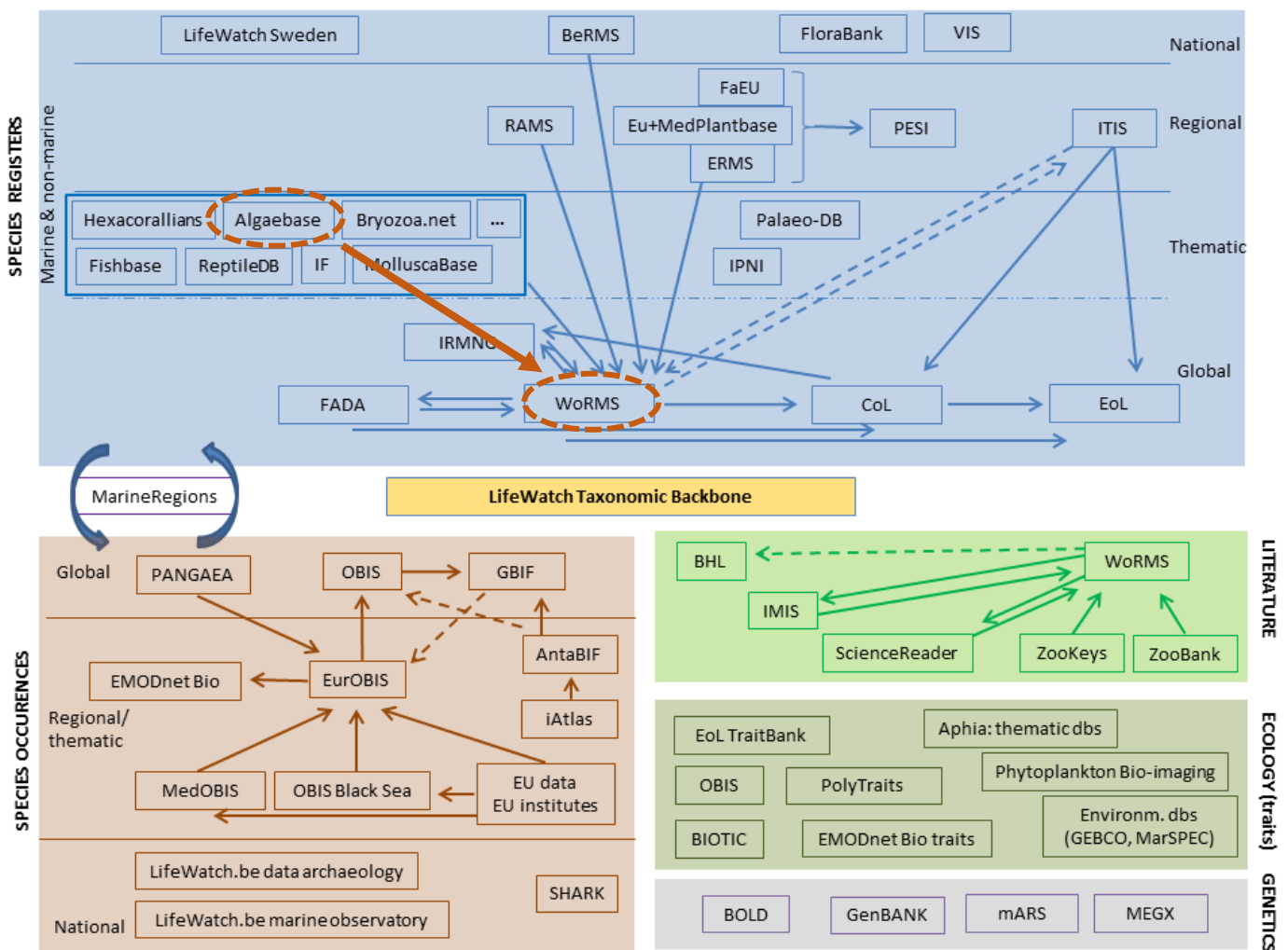

***Figure 5**: Architecture of the LifeWatch Taxonomic Backbone as constructed by VLIZ. In the future, more links will have to be explored and strengthened in order to improve the system. During my internship, I explored the link between AlgaeBase and WoRMS (as indicated in brown). [9]*

Framework internship II : The Aphia underpinned database WoRMS and the externally managed database AlgaeBase

For the synchronisation I used the Aphia database, which is a consolidated database that integrates information from several Global Species Databases (GSD), Regional Species Databases (RSD) and Thematic Species Databases (TSD). In order to be complete, Aphia also harvests information from other marine species lists managed by others (e.g. Fishbase). It provides a list of all known marine species names and even stores non-marine taxa, though it concentrates mostly on the marine ones. Next to accepted names it also stores unaccepted names (e.g. synonyms or even deleted names) and even documents the links between these different names, allowing for better quality control of the taxonomic information. Aphia also contains an editing platform that underpins databases like WoRMS and other global, thematic and regional databases. This allows us to easily access and manage the database. [10]

WoRMS

The World Register of Marine Species is one of the biggest contributors to the TB of the LifeWatch project. The aim of WoRMS is to provide an accurate and as-complete-as-possible register of all names of marine species. This includes accepted names and synonyms, with highest priority going to the valid names. Adding synonyms in its list helps users to better understand taxonomic literature which often contains a mixture of accepted and unaccepted names. [11] WoRMS combines information found in the taxonomic backbone with information extracted from externally hosted and managed (not maintained at VLIZ) species databases like Fishbase or AlgaeBase that are prepared to share their expertise and data. This with the aim to avoid duplication and use the good infrastructures that are already in place. The content on WoRMS is not checked by database managers but by taxonomic experts to ensure adequate quality control of the information.


Next to accepted names and synonyms, this database contains literature, biogeographic data and higher classification of each entry when possible. Though, this higher classification is not to be seen as precise as its aim is to aid data management rather than to suggest any opinion on species taxonomic relationships.

Because taxonomic information is added daily, sometimes even in the form of large lists by global or regional species lists, keeping WoRMS up to date is a continuous and time-consuming endeavour for the data management team! [12]Nevertheless, WoRMS is reaching the final stadium regarding its taxonomic content and then the main focal point will be adding new information and checking species traits and distributional data.

AlgaeBase

As stated before, AlgaeBase is an externally hosted and managed Global Species Database that focusses on marine, freshwater and terrestrial algae. However, information on marine algae, and more particularly seaweeds, is most complete in this database. Sea grasses are included as well, even though they are flowering plants. This purely because they used to be classified as sea weeds. Its main goal is to serve as an aid in taxonomic studies, not be an ultimate source.
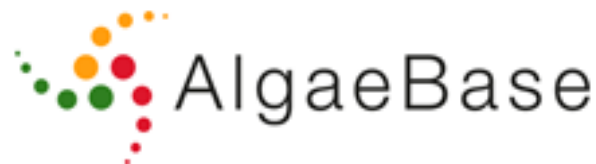
The database is linked to WoRMS through an end user agreement, which states that WoRMS can extract data from AlgaeBase and present it to its users on the online platform. So, AlgaeBase licenses its data to other websites but these may not just re-distribute its data. Responsibility for the content and the maintenance of the database remains with the original institute and not VLIZ (in this case: The National University of Ireland, Galway). [13]

Internship activity: The synchronisation between WoRMS and AlgaeBase

During my internship, I added new taxa from AlgaeBase in WoRMS and did a quality control of the taxonomic information on algae already present in WoRMS. For this task I used an interface called 'AlgaeBase Harvester'. This platform was especially designed to extract taxonomic information from the AlgaeBase database and sync it with data available in WoRMS. For this you must pick a 'harvesting period' in the module. Then, the interface will show you, one by one, all the new entries and those that were modified during this period in AlgaeBase.

I used 'cached lists', these contain entries that actually need manual intervention. This reduced the waiting time as every entry must be dealt with and none needs to be skipped (**Figure 6**).
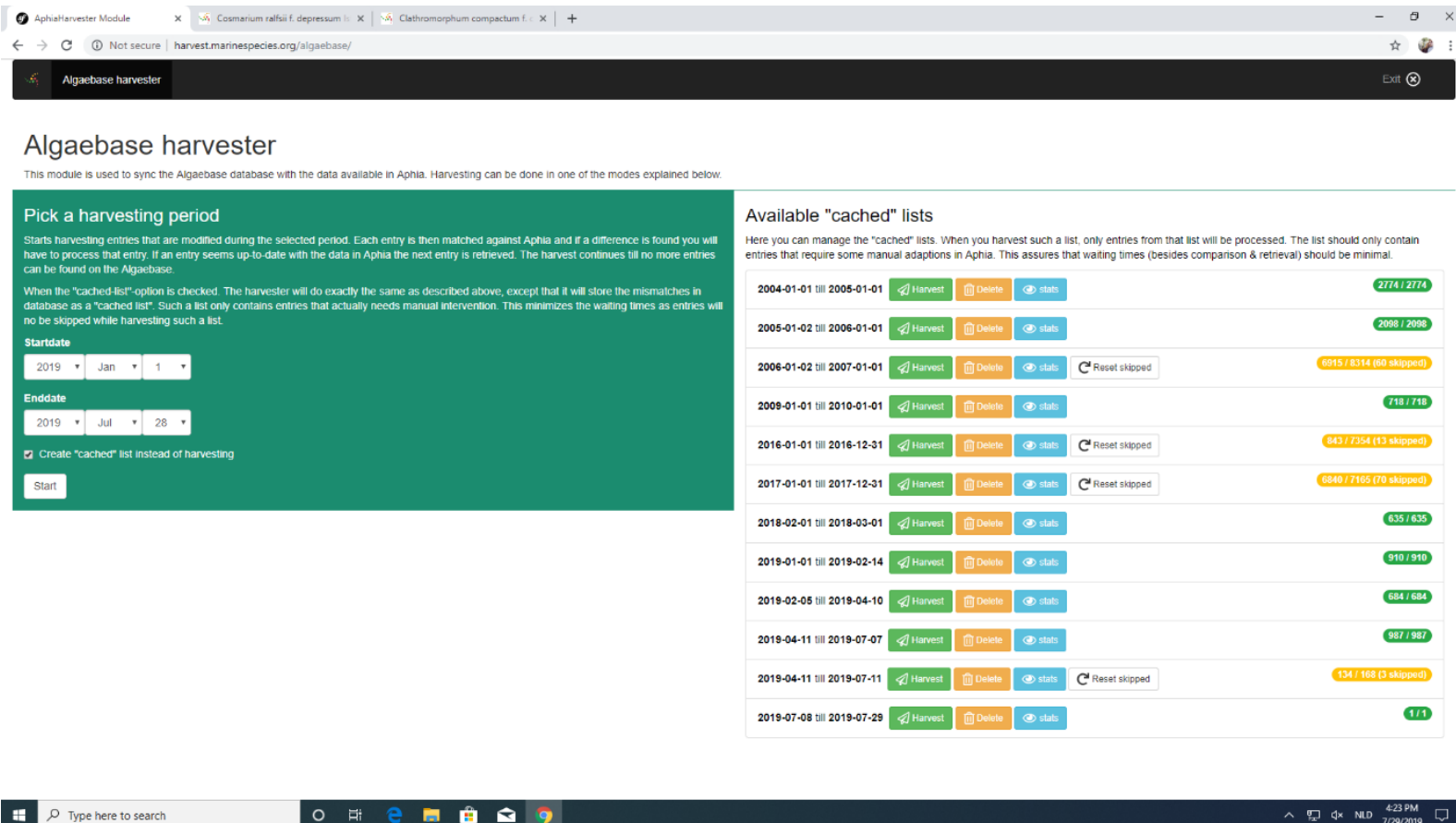

Figure 6: *Picture of the harvester interface used to add taxonomic information from AlgaeBase in WoRMS and to manage modifications made in AlgaeBase. On the left you pick the harvesting period. Once that is done, cached lists will appear on the right.*

The most common task I had to perform was keeping the information in Aphia up to date with the modifications done in AlgaeBase. In this case, the interface showed me the taxon in AlgaeBase with its information on the right (Remote taxon) and the taxon with its information in Aphia on the left. In this comparison page two colors can be apparent in case of mismatch: yellow and red. These represent two levels of discrepancy between the AB and Aphia taxon, with yellow representing a small difference that does not always need to be accounted for and red representing a large difference for which the information in Aphia must be adjusted to that of AlgaeBase. Yellow can also mean that Aphia contains more information than AlgaeBase. In this case we keep this extra information. An example of the comparison page and the process of accounting for the differences can be found in **Figure 7, 8 and 9.** Once all the differences between the 2 taxa are resolved, I moved on to the next entry.

Figure 7: *The comparison page on the AlgaeBase harvester. In this example 4 major differences are indicated by the red colour. In this case the information of the Aphia taxon needs to be matched to that of the Remote taxon (here from AlgaeBase but this interface can be used to harvest information of other externally hosted databases as well). We change the information of the Aphia taxon in red to that of the remote taxon. Here, we must adjust the author of the taxon, the year in which it was published, the taxonomic status of the name and the accepted name. Changes can be done in the editor below the comparison page (see next figure). The yellow here indicates that Aphia has more information concerning this taxon, thus we keep this information the way it is.*

Figure 8: *The editing platform below the comparison page. Here we change the information of the Aphia taxon to that delivered by the Remote taxon until it matches (see next figure).*

Aphia taxon [Refresh ↻]		Remote taxon	
SpecificEpithet	fragilis	☐ SpecificEpithet	fragilis
InfraSpecificEpithet	microptila	☐ InfraSpecificEpithet	microptila
ScientificNameAuthorship	F.M.Bailey	☐ ScientificNameAuthorship	F.M.Bailey
NamePublishedInYear	1892	☐ NamePublishedInYear	1892
ScientificName	Chara fragilis f. microptila F.M.Bailey, 1892	☐ ScientificName	Chara fragilis f. microptila F.M.Bailey
Genus	Chara	☐ Genus	Chara
TaxonRank	Forma	☐ TaxonRank	forma
Kingdom	Plantae	☐ Kingdom	Plantae
TaxonomicStatus	uncertain	☐ TaxonomicStatus	uncertain taxonomic status subjected to verification
HigherClassification	Chara fragilis A.N.Desvaux, 1810 (Species)	☐ HigherClassification	Chara Linnaeus, 1753 (genus)
	Chara Linnaeus, 1753 (Genus)		Chareae (tribe)
	Chareae (Tribe)		Characeae (family)
	Characeae S.F. gray, 1821 (Family)		Charales (order)
	Charales Dumortier, 1829 (Order)		Charophyceae (class)
	Charophyceae Rabenhorst, 1863 (Class)		Charophyta (phylum)
	Charophyta Cavalier-Smith (Phylum (Division))		Streptophyta (infrakingdom)
	Streptophyta John, Willamson & Gulry, 2011 (Infrakingdom)		Viridiplantae (subkingdom)
	Viridiplantae Cavalier-Smith, 1981 (Subkingdom)		Plantae (kingdom)
	Plantae Haeckel, 1866 (Kingdom)		Eukaryota (empire)
	Biota (Superdomain)		
isMarine	0	☐ isMarine	
isBrackish	0	☐ isBrackish	
isFreshwater	1	☐ isFreshwater	1
isTerrestrial		☐ isTerrestrial	
isExtinct		☐ isExtinct	
Visibility	Visible	☐ Visibility	Unknown
dcterms:URI	Chara fragilis f. microptila F.M.Bailey, 1892 [view]	☐ dcterms:URI	Chara fragilis f. microptila F.M.Bailey [view]
acceptedNameUsage		☐ acceptedNameUsage	
orignalNameUsage	Chara fragilis f. microptila F.M.Bailey, 1892 [view]	☐ orignalNameUsage	Chara fragilis f. microptila F.M.Bailey [view]

Figure 9: *The outcome of the update. The large differences are accounted for except for those in the higher classification. Though, these were ignored through the process. The yellow only constitutes small differences ('uncertain taxonomic status subjected to verification' and 'uncertain' is essentially the same) and extra information in Aphia about the environment. This entry can now be 'marked as done'.*

Other times, I had to add taxa that were new to literature. These were often genera or species and sometimes even families. Though adding new names was easy, every taxon had to be checked to whether it was already present in Aphia but under a slightly different name. This needs to be checked in order to avoid duplication. The last synchronisation between WoRMS and AlgaeBase happened in March 2015. Any taxon described after this date is a legit new record and can be added in Aphia without being worried of duplicates.

Lastly, I came across a lot of 'difficult' entries where more checking and work was necessary. Sometimes taxa had to be deleted from the Aphia database. Other times there were duplicates in AlgaeBase and the authorities of AlgaeBase had to be informed.

SKILLS LEARNED

During my internship, I learned most about WoRMS/Aphia: how it's constructed, how to edit a taxon on the database, how easy mistakes are made and how difficult it can be to correct for these mistakes. I also learned how to work with AlgaeBase and how to harvest data from this platform to WoRMS. Especially during this harvesting, I noticed how difficult it is to just rely on the database's information to be correct. I had to constantly be aware of mistakes in WoRMS and even in AlgaeBase (even though we considered their information to be correct). Taxonomy is a field that is under constant construction, which I noticed during my internship. I learned that adding information in a database is much more than just copy-pasting and requires constant attention. I learned

to be critical for the work I was doing, but also for the information present in both databases. I processed a few thousand entries during my 3 weeks and learned how to be fast in doing them, yet with eye for mistakes. Finally, I learned to appreciate taxonomy and its value even more.

During this short period, I met some very nice people, from interns to students to employees, that made me feel very welcome. I want to thank Thomas and Leen for supervising me and granting me the opportunity to do my internship at VLIZ!

TIMETABLE

	MONDAY 8/7	TUESDAY 9/07	WEDNESDAY 10/07	THURSDAY 11/07	FRIDAY 12/07
AM	Introduction to Aphia and the AlgaeBase Harvester	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Day of the Flemish Community (*)	Harvesting AlgaeBase data and syncing with WoRMS
PM	First time processing entries using a manual (made by Thomas)	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Day of the Flemish Community (*)	Harvesting AlgaeBase data and syncing with WoRMS
	MONDAY 15/07	TUESDAY 16/07	WEDNESDAY 17/07	THURSDAY 18/07	FRIDAY 19/07
AM	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS
PM	Feedback on first week and processing 'difficult' entries with supervisor (in order to be able to handle them alone)	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS
	MONDAY 22/07	TUESDAY 23/07	WEDNESDAY 24/07	THURSDAY 25/07	FRIDAY 26/07
AM	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS
PM	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS	Harvesting AlgaeBase data and syncing with WoRMS
	MONDAY 29/07				
AM	Harvesting AlgaeBase data and syncing with WoRMS				
PM	Harvesting AlgaeBase data and syncing with WoRMS				

(*) To account for the loss of this day, I started a half hour earlier and ended a half our later each day to complete 120 hours of internship.

REFERENCES

[1] VLIZ. (2019). *Organisation*. [Online]. Available: http://www.vliz.be/en/organisation. [Accessed 5 august 2019].

[2] VLIZ. (2019) *Strategic objectives*. [Online]. Available: http://www.vliz.be/en/strategic-objectives. [Accessed 5 august 2019].

[3] VLIZ. (2019). *History: The first marine institute by the sea (1843-1967)*. [Online]. Available: http://www.vliz.be/en/first-marine-institute-sea-1843-1967. [Accessed 5 august 2019].

[4] M. Breyne, J. Seys, A.-C. Lescrauwaet, H. Debergh, J. Haspeslagh, H. Lust en J. Mees. (2010). *The world's very first marine research station in Ostend (Belgium)*. *EARTHZINE*.

[5] VLIZ. (2019). *History: The Institute for Marine Scientific Research (1970-1999*. [Online]. Available: http://www.vliz.be/en/institute-marine-scientific-research-1970-1999. [Accessed 5 august 2019].

[6] VLIZ. (2019). *History: The Flanders Marine Institute (1999 – present)*. [Online]. Available: http://www.vliz.be/en/flanders-marine-institute-1999-present. [Accessed 5 august 2019].

[7] VLIZ. (2019). *InnovOcean site partners*. [Online]. Available: http://www.vliz.be/en/innovocean-site-partners. [Accessed 10 august 2019].

[8] LifeWatch. (2019). *The Belgian LifeWatch Infrastructure*. [Online]. Available: http://www.lifewatch.be/en/project-belgian-lifewatch-infrastructure. [Accessed 10 august 2019].

[9] LifeWatch. (2019). *Taxonomic Backbone: Architecture*. [Online]. Available: http://www.lifewatch.be/en/about-taxonomic-backbone. [Accessed 10 august 2019].

[10] L. Vandepitte, B. Vanhoorne, W. Decock, S. Dekeyzer, A.-T. Verbeeck, L. Bovit, F. Hernandez en J. Mees. (2015).*How Aphia—The Platform behind Several Online and Taxonomically Oriented Databases—Can Serve Both the Taxonomic Community and the Field of Biodiversity Informatics. Journal of Marine Science and Engineering,* pp. 1448-1473.

[11] LifeWatch. (2019). *World Register of Marine Species (WoRMS)*. [Online]. Available: http://www.lifewatch.be/en/worms. [Accessed 14 august 2019].

[12] WoRMS. (2019). *What is WoRMS?*. [Online]. Available: http://www.marinespecies.org/about.php. [Accessed 14 august 2019].

[13] LifeWatch. (2019). *AlgaeBase.* [Online]. Available: http://www.lifewatch.be/en/algaebase. [Accessed 20 august 2019].