2019

Conservation Internship

REPORT

KAREL DOCKX

Conservation Internship Report

VLIZ – World Register of Marine Species

Introduction

During April and May, I did an internship in Oostende, at the Flanders Marine Institute (Vlaams Instituut van de Zee: VLIZ). Within VLIZ, there are several divisions, the datacenter, my work environment, being one of these. The main work of the data center is the management of the World Register of Marine Species (WoRMS). This introduction explains what WoRMS is, based information mentioned on the website (http://www.marinespecies.org/). Afterwards, there will be a more detailed description of my tasks and the activities.

The World Register of Marine species, in the rest of this report called 'WoRMS', is a large worldwide database that provides a list of known marine species and tries to be as complete as possible. This database contains names, authorities, synonymy and so on. The primary goal is to serve as a guide to interpret different kinds of taxonomic literature. (WoRMS Editorial Board)

The content of this database is not controlled by the datacenter, where I worked. The responsible persons for the quality of the information are different taxonomic and thematic experts. Every taxonomic group within the system has its own expert. This person can get support from several specialists, but it is his/hers responsibility that the given information in the database is correct. The database managers process this information, but always need the confirmation of the expert to do adjustments in the actual database. In some cases, data is collected in Excel files, then controlled by the expert and eventually entered in the database.

WoRMS is a continually growing database. There are constantly new entries from species that are newly discovered or were still missing in the database. Sometimes there are also changes in taxonomy that have to be adjusted. As shown in figure 1, the database is continuously growing over the years. Moreover, not only the database grows, but also its popularity and number of visitors is constantly increasing. Because more and more people use the available information, it is obvious that the quality of the database is crucially important. (WoRMS Editorial Board)

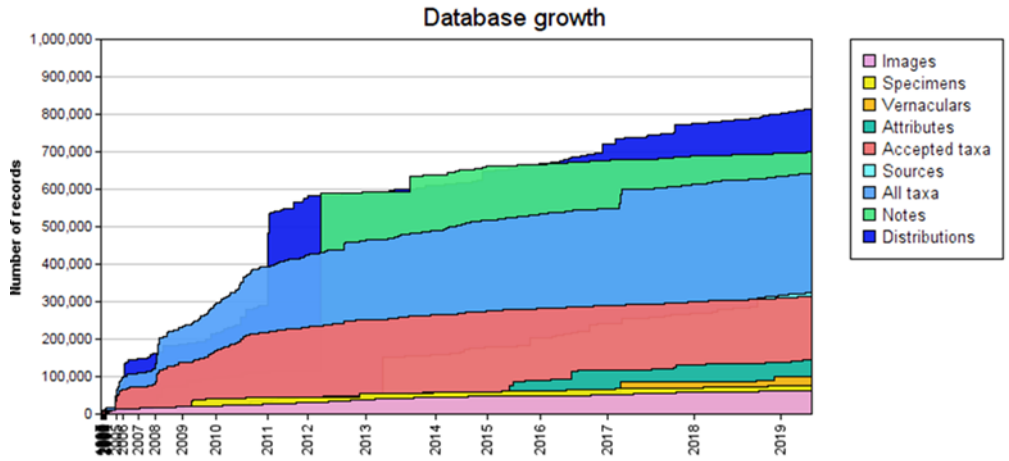

Figure 1: database growth of WoRMS trough the recent years

The problems that occur when species names are used incorrectly or inconsistently are not mainly academical, but also have conservation consequences (Mark J. Costello et al., 2013). The IUCN Red List is a well-known source of information about the conservation status of species, but the species must be properly identified to be included. Due to incorrect identification of pests and pathogens, there were wasted control measures in the past. Species are often wrongly named because of lack of taxonomical information. This causes problems in many different ways. An example of this are the problems in the tracking of fish populations. (Bartolus, 2008)

These problems can be resolved by using a generally accepted database like WoRMS. Figure 2 shows that more and more organisations indeed use the WoRMS database, which is an evolution that we can only be happy about.

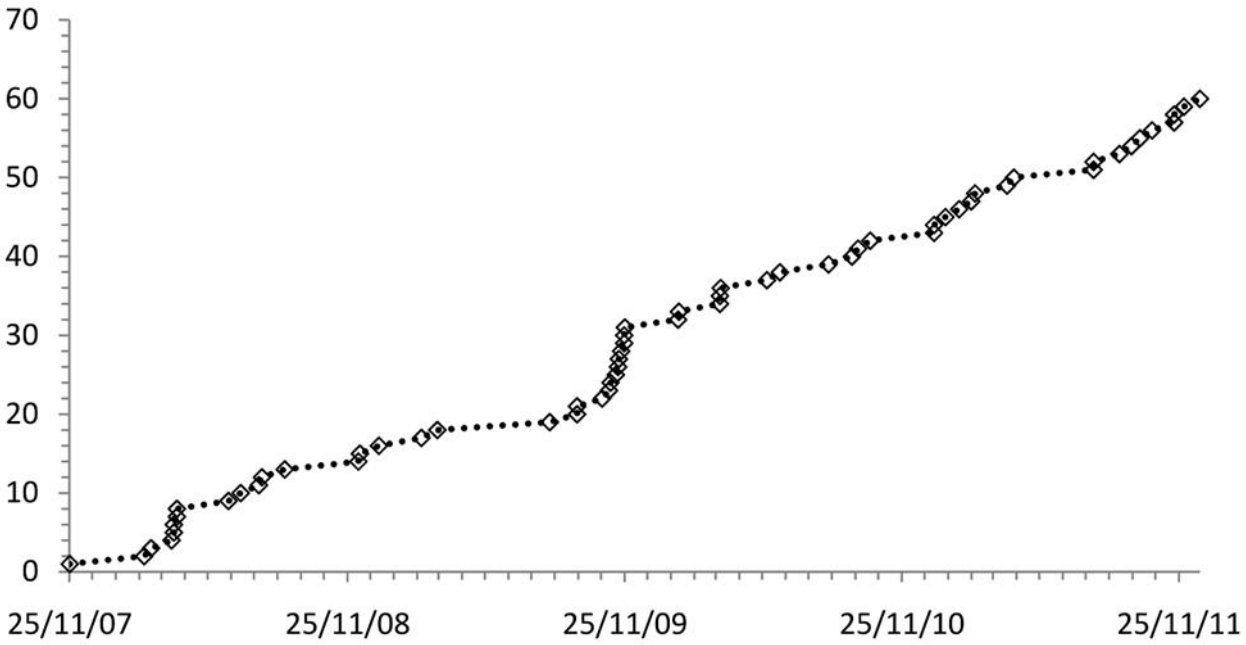

Figure 2: The number of organisations using WoRMS for their research and/or data management

Summary of activities

Calendar

Mo 22/04	Tu 23/04	We 24/04	Th 25/04	Fr 26/04	Sa 27/04	Su 28/04
	9u – 17u	8u40 – 16u45	8u40 – 16u45	8u40 – 16u45		
Mo 29/04	Tu 30/04	We 1/05	Th 2/05	Fr 3/05	Sa 4/05	Su 5/05
8u40 – 16u45	8u40 – 16u45		8u40 – 16u45	8u40 – 16u45		
Mo 6/05	Tu 7/05	We 8/05	Th 9/05	Fr 10/05	Sa 11/05	Su 12/05
8u40 – 16u45	8u40 – 16u45	8u40 – 16u45	8u40 – 16u45	8u40 – 16u45		
Mo 13/05	Tu 14/05	We 15/05	Th 16/05	Fr 17/05	Sa 18/05	Su 19/05
8u40 – 16u45	8u40 – 16u45	8u40 – 16u45	8u40 – 16u45	7u45 – 18u30 (*)		

(*) The 'working' hours were longer because this was the day of the team building

Week one and two

Both week one and two were a little bit shortened by public holidays, as shown in the calendar (fig.1). During this period, the assignment was to find the authorities of two lists of species. One list consisted of accepted species names, which are the valid names of the species. The second list consisted of the unaccepted species, these species names are not valid. This means that they are either synonyms of accepted species or that they are for another reason not commonly used. An example of such a reason is that the species is not completely characterized yet, and thus impossible to properly identify. (WoRMS Editorial Board) For this task I had to do a literature study to find the first describer of the different species and the year of discovery. For this task, there were different databases available, but the major source still was Google Scholar. Other databases that I used were Fishbase (Froese R. & Pauly D., 2019), Eshmeyer's catalogue of fishes (Fricke et al., 2019) and Algaebase (Guiry, M.D. & Guiry, G.M., 2019)

This was an important and rather urgent task because almost 6000 organisms in the database were lacking an authority. This is not only a problem because of the incomplete taxonomic name, but also because of the different statistic analyses that are made using the database. If you take into account that a large amount of species is not present in these analyses and some graphical representations, then it is easy to see that these inquiries could be rather skewed. As an illustration for this, I took two graphs of the WoRMS site which illustrate the problem.

In figure 3, the amount of species described by different authors in the last decade are illustrated. The top describer, Rolan, apparently described about 988 species present in this database. Besides him, there are also a lot of authors which described fewer species. Figure 4 is an illustration of the discovery rate, which is at it absolute top about 4000 species in one year. But, for most years, it are less than one thousand species. These are only two examples, and there are a lot more graphs and analyses, but it is clear that a list of 6000 organisms can have a major effect on how these graphs look. These are just the basic statistics from the

WoRMS site itself. After a quick search, there are more applications to discover. Costello & Wilson (2011) used the trends in discoveries, illustrated in figure 4, to make a prediction on how many remain to be discovered. This if of course based on the year of discovery, and not the author of the species.

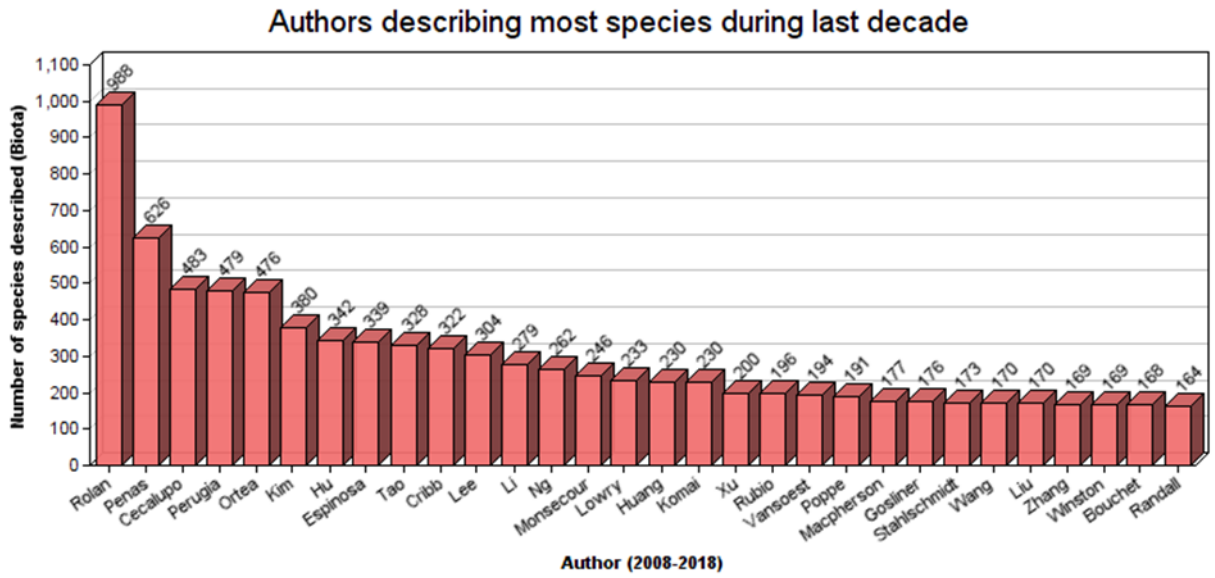

Figure 3: Authors describing most species during the last decade

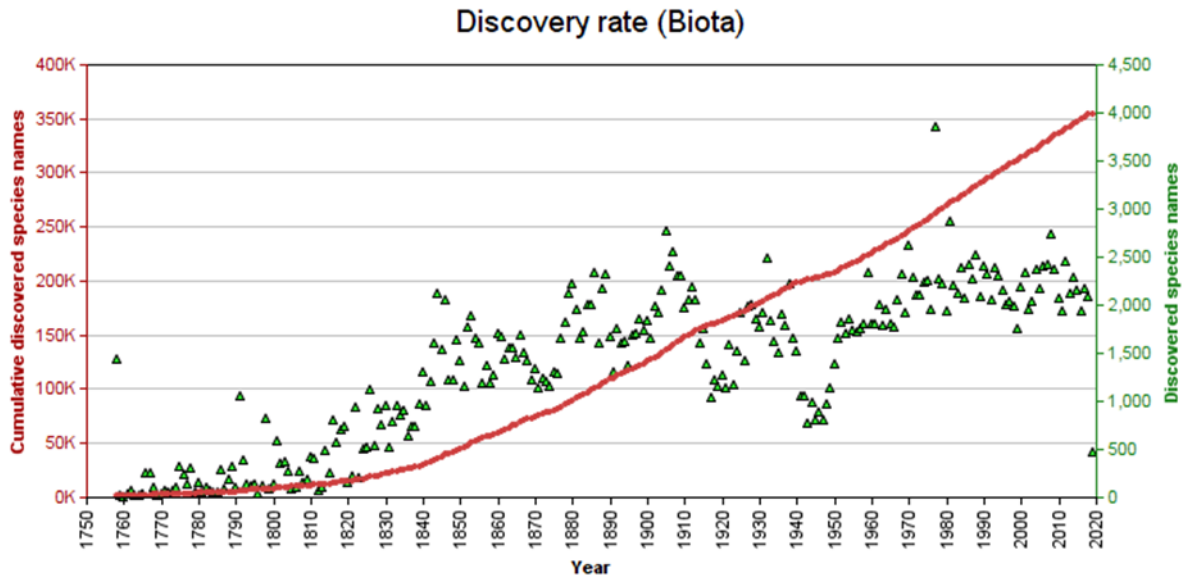

Figure 4: Discovery rate of marine species

A part of this task is given in annex 1, the document was very big and it didn't seem relevant to include the complete document.

Week three

Week 3 started with the same task as the previous weeks. Especially the list of unaccepted species was very long and still took some time to complete.

The next task was to format a list of literature references. The items of these references (Author, year, title, bibliographic reference…) were provided in one continuous single line and needed to be formatted separately in order to give users of the database the possibility to search on the different parts or parameters of the reference, which was apparently not possible or difficult in the way it was now.

Week four

In the last week of the internship I was assigned yet another task. A long list of holotypes needed to be matched with the database. Holotypes are the first discovered specimens of a certain species which are thus used for the description of the species in general.

The database has a build-in function to match this list to species in the database. First a list with holotypes in excel format is uploaded into the tool. The tool then compares the species names of this list, with extra information about the holotype, with species names present in the database. After this process the return is yet again an excel file with information about the succeeding of this process. The only problem with this automated function was that small errors in spelling complicate the process and prevent this matching from happening. So in the 'matched list' I could check if there was a 'match', which was good and no further action was necessary. A second possibility was that there was a little deviation between the name in the file and a possible match in the database. My job was to check if this was actually a match, and if this was indeed the same species. Some species had bigger deviations, which needed a bigger search. The last group of species appeared to be absent in the database.

A part of this task is given in annex 2, the document was very big and it didn't seem relevant to include the complete document. On the right side of the document, some of the codes can be spotted. 'Exact' means an exact match in the database, which is good. Codes like 'near_2' or 'ambiguous' mean that there are little deviations. An empty cell means a no match, this was often due to a divergence in authority.

The very last day of my internship, the team building was scheduled. This was a guided trip to Nausicaa in Boulogne sur Mer, the biggest aquarium of Europe. Although it is not ideal that a team building takes place on the very last day of my internship, it was, in my opinion, a worthwhile experience. Team building is part of the job in a company and it is actually very positive that this was included in the internship. The trip by itself was educational, and a kind of an illustration of the work I did in the preceding weeks. Also, because the main goal of this internship was to gain insights in the way of working within an organization, I think it was a valuable addition to the internship.

Conclusion

There was a very nice, calm and friendly work environment, questions were answered almost immediately, help was provided when needed and the description of the different tasks was clear. from my side, I think I communicated well about problems and progress. I tried to show some independence by solving the different tasks in my own structured system and by thinking about problems before asking for help.

Different trainees were placed together which was also a good thing for the atmosphere. By the start of my internship it was a little bit difficult on the social level, because it was a totally new environment for me with all new people. But in my opinion, this was already a lot better at the end of the internship. Especially for this topic, the teambuilding was a perfect addition.

Sometimes during breaks, short presentations were given about the work and projects other people were working on. This gave an even broader look on the different tasks of VLIZ, and was a nice form of scientific 'entertainment' during the breaks.

About my own work, I think I can be quite satisfied about what I realized. I completed a huge task, which was apparently important and urgent. Furthermore, I had multiple contributions to other tasks. It was not always easy to keep the focus, because the different tasks were a little monotonous. But the fact that you are doing valuable work for a prestigious project gives the much needed motivation. I tried to describe as good as possible the applications and opportunities that a database like this can give to the world of science, because it clearly is an important tool as both a source of information and to make statistical analyses and even predictions for the future. It can and will be a valuable tool in the conservation of marine species.

It was maybe not the most difficult internship, neither the most challenging. But I did learn how to efficiently work with different databases. Not only the WoRMS database, but also Fishbase, Eschmeyer's, Algaebase and others were crucial tools. The major part of the internship was to do qualitative and efficient research, which is indeed an important skill. I gained some insights and learned new things about taxonomy, and had a broader look on the life inside the oceans. On top of that, I worked for an organization with international prestige, and got an elaborate insight in the way things work in such an institute.

References

1. Bartolus A (2008), Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. Ambio 37: 114–118.
2. Costello MJ, Wilson SP (2011), Predicting the number of known and unknown species in European seas using rates of description. Global Ecology and Biogeography 20: 319–330.
3. Fricke, R., Eschmeyer, W. N. & R. van der Laan (2019). ESCHMEYER'S CATALOG OF FISHES: GENERA, SPECIES, REFERENCES. [http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp]. Electronic version accessed dd mmm 2019.
4. Froese R. & Pauly D. Editors. (2019). FishBase. World Wide Web electronic publication. [www.fishbase.org] version 02/2019
5. Guiry, M.D. & Guiry, G.M. (2019). AlgaeBase. World-wide electronic publication, National University of Ireland, Galway. [http://www.algaebase.org]
6. Mark J. Costello , Philippe Bouchet, Geoff Boxshall, Kristian Fauchald, Dennis Gordon, Bert W. Hoeksema, Gary C. B. Poore, Rob W. M. van Soest, Sabine Stöhr, T. Chad Walter, Bart Vanhoorne, Wim Decock, Ward Appeltans (2013), Global Coordination and Standardisation in Marine Biodiversity through the World Register of Marine Species (WoRMS) and Related Databases. Plos.One [https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0051629]
7. WoRMS Editorial Board (2019). World Register of Marine Species. Available from http://www.marinespecies.org at VLIZ. Accessed 2019-05-13. doi:10.14284/170

Annex 1

	A	B	I	J	
1	Color codes legend: see Sheet2 + comments				
2	ScientificName	AphiaID	Authority	Source	Link
3	Lumbricus communis	1041445	Hoffmeister, 1845	Google Scholar	https:
4	Allolobophora danieli	1041432			
5	Dendrobaena ganglbaueri	1041438	Rosa, 1894	Google Scholar	https:
6	Helodrilus ganglbaueri	1041442			
7	Allolobophora putris	1041435			
8	Dendrobaena schelkovnikovi	1041439			
9	Eisenia tetraedra	1041441	Savigny, 1826	Google Scholar	https:
10	Lumbricus tetraedrus	1041446	Savigny	Google Scholar	http:/
11	Allolobophora tetraedrus	1041436			
12	Saenuris diversisetosa	1041454			
13	Euphrosine balani	209691			
14	Ninoe americana	182812			
15	Arabella debilis	413990			
16	Scoletoma magnidentata	152260	Winsnes, 1981	Google Scholar	http:/
17	Paradiopatra piccola	736651	Paxton & Budaeva, 2013	Google Scholar	https:
18	Tomopteris (Johnstonella) aloysii	335150			
19	Sphaerosyllis arenaceus	334821			
20	Polynoe canadensis	254756	(McIntosh, 1874)	Google Scholar	https:
21	Typosyllis ehlersoides	335156			
22	Exogone furcigera	183159			
23	Polynoe minuta	335320	Potts, 1910	Google Scholar	https:
24	Sphaerodoropsis multipapillata	335138	(Hartmann-Schröder, 1974)	Google Scholar	https:
25	Sthenolepis spinosa	335145			
26	Kettnerites (Kettnerites) abraham	335094			
27	Delosites raridentatus	335080	Kozur, 1967	Google Scholar	https:
28	Kettnerites (Aeolus) sisyphi	335093			
29	Rhodine bitorquata	254758	Moore, 1923	Google Scholar	https:
30	Dasybranchus capitata	209889			
31	Euclymene gracilis	254705	(Sars, 1861)	Google Scholar	LIST,
32	Notoproctus minor	181483			
33	Spirorbis (Spirillum) lucidus	335142			
34	Colobranchus laevicornis	335078			
35	Proclea emmi	254757	Annenkova, 1937	Google Scholar	http:/
36	Cirratulus grandis	157297	Verrill	Google Scholar	https:

Annex 1a: short illustration of the first task; Authorities of given species needed to be researched, the used source had to be reported, with a link (link is outside the range of the illustration). In Annex 1b, the used color code is explained. Outside the range of this illustration was another column for remarks.

Color Code Legend		No records found in given databases or google scolar
		Authority completed
		No date found
		Species not found, but a species with very similar name found... Spelling mistake? See remarks
		Not sure: different authorities found
REMARKS	1	When the code is orange, a possible correction with authority can be found at the remarks column
		The given link and source is the page where the possible correction can be found
	2	When the name is currently regarded as a synonim of another name, this can be found in the remarks column
	3	For virusses, it was very unclear and difficult to find. So I'm not sure about these authorities

Annex 1b: Legend for the color codes + remarks

Annex 2

	A	B	C	D	AN	AO
1	ScientificName	SynTO::Author	Comment	Number	AphialD	Match type
2	Staurocephalus erucaeformis	Malmgren, 1865		251	154942	exact
3	Eteone lentigera	Malmgren, 1887		252	338701	exact
4	Petta pusilla	Malmgren, 1866		253	130597	exact
5	Axiothea catenata	Malmgren, 1865		254	155368	exact
6	Praxilla praetermissa	Malmgren, 1865		255	155125	exact
7	Eone nordmanni	Malmgren, 1865		256	155415	exact
8	Ceratocephale loveni	Malmgren, 1867		257	130367	exact
9	Amphitrite palmata	Malmgren, 1866		258	338651	exact
10	Amphitrite affinis	(Malmgren, 1866)	OK	259	131472	
11	Sigalion edwardsi	Kinberg, 1856		260	338371	exact
12	Sthenelais articulata	Kinberg, 1858		261	332162	exact
13	Sthenelais helenae	Kinberg, 1858		262	155081	exact
14	Sthenelais articulata	Kinberg, 1858		263	332162	exact
15	Sthenelais laevis	Kinberg, 1858		264	338395	exact
16	Hermadion longicirratus	Kinberg, 1858		265	328322	exact
17	Sthenelais blanchardi	Kinberg, 1858		266	332164	exact
18	Lepidonotus wahlbergi	Kinberg, 1858	Unclear... 2 possibilities	267		ambiguous
19	Aphrogenia alba	Kinberg, 1858		268	129843	exact
20	Iphione ovata	Kinberg, 1858		269	333675	exact
21	Lepidonotus jacksoni	Kinberg, 1858		270	209664	exact
22	Lepidonotus margaritaceus	Kinberg, 1858		271	328802	exact
23	Lepidonotus johnstoni	Kinberg 1855		272	338930	exact
24	Iphione spinosa	Kinberg, 1858		273	338268	exact
25	Aphrodita alta	Kinberg, 1858		274	129841	exact
26	Lepidonotus indicus	Kinberg, 1858		275	339699	exact
27	Lepidonotus pomareae	Kinberg, 1858		276	337676	exact
28	Chaetacanthus pomerae	(Kinberg, 1858)	*Chaetacanthus pomareae* (Kinberg, 1856)	277	333253	near_2
29	Lepidonotus caeruleus	Kinberg, 1858		278	328785	exact
30	Lepidonotus socialis	Kinberg, 1858		279	338939	exact
31	Antinoe pulchella	Kinberg, 1858	*Antinoe pulchellus* Kinberg, 1856	280	339646	near_2
32	Halosydna brevisetosa	Kinberg, 1856		281	333534	exact
33	Hermadion magalhaensi	(Kinberg, 1858)	OK	282	155298	
34	Halosydna parva	Kinber, 1856	OK	283	333542	
35	Halosydna australis	Kinberg, 1858		284	337631	exact
36	Halosydna elegans	Kinberg, 1858		285	337633	exact
37	Antinoe waahli	Kinberg, 1858		286	339648	exact
38	Halosydna brasiliensis	Kinberg, 1858		287	335341	exact
39	Antinoe microps	Kinberg, 1858		288	155216	exact
40	Halosydna patagonica	Kinberg, 1858		289	155283	exact
41	Harmothoë spinosa	Kinberg, 1858		290	155282	exact
42	Halosydna virgini	Kinberg, 1856		291	337642	exact

Annex 2: short illustration of the third task. This is the output given by the system.

Annex 3: Description of internship before start

Contribution to finalizing the World Register of Marine Species (WoRMS)

From the 22nd of April until the 17th of May, I will work at the Flanders Marine Institute. Here I will work at a worldwide database of marine species. This database contains contributions from volunteers worldwide, but the content is controlled in Oostende. The goal is to provide a list with valid names of marine species worldwide. Which can be used to interpret taxonomic literature. Every taxonomic group is controlled by a taxonomic expert, who is responsible for the correct content of the database. This expert works with a group of people for the data management. I will be part of the data management group.

The Flanders Marine Institute works with a 38-hours work week. This comes down on an average working day of 7 hours and 36 minutes (plus a mandatory 30-minute break). These hours are flexible, though presence between 10h and 16h30 is mandatory.

My job will be to control the taxonomy of the species in the system. Each species has a file, but some of these files are incomplete. Mainly the author names and the years of discovery are missing. These are important data for different statistics, which are incomplete without the missing information. With the information available of each species, I must do a literature study to find the missing information. For most of the species, this will be a rather simple task, but for others it will be challenging. There are multiple, more specific, databases available at the institute. And with the help of these databases, it should be possible to complete the missing information. At the moment, about 7000 species files are incomplete. And about 1000 of them are files which are part of my job.

So, the biggest part of the job will be to search for the missing links in the files. But there may come some other tasks. At the centre they said that it is difficult to define the work that will be available at the time of my internship. But ultimately, they all serve the same purpose, namely to create a species database which is as complete and correct as possible.