LifeWatch Data Grant 2014

Filling the gaps in the World Register of Marine species (WoRMS)

Syllidae (Polychaeta)

Final Report

João Gil

1. Data grant background

The World Polychaeta Database (WPolyDb) is an established global species database within WoRMS, with over 23,000 taxa records. Syllidae in WPolyDb had, at the beginning of this grant, 2070 entries and 169 genera. Consequently it is the family of Polychaeta with the biggest number of described species, and also a family with a very complex taxonomy, including subfamilies, several subgenera and numerous alternate representations. These Syllidae remained largely unchecked for status and errors since importation in 2007, although a big number of errors were to be expected. Traditionally the taxonomic publishing activity in the group has been high, with significantly the most prolific author of Polychaeta taxa overall since 2000 being a Syllidae expert (Aphia stats). Besides the new taxa, the numerous recent publications on the group have introduced many taxonomic changes which were not reflected in WPolyDb. Being insufficiently edited, the group had fallen behind with the taxa addition and revision standard of the rest of the Polychaeta.

2. Agreed deliverables (as specified in the Data Grant contract)

- Taxonomic update of the whole family Syllidae (Annelida, Polychaeta), covering an estimated number of about 1000 valid taxa. The total number of taxa to be treated is estimated at 1000 valid species and genera, with 200 new or missing name records to be newly added, and 2000 name records to be revised.

- Fulfill beyond the 'highly desirable information' level (basionym; original publication reference with link to the BHL or other bibliographic resources; holotype information such as type locality and its geolocation, museum where type material is deposited, collection number), for the family Syllidae, and make corrections and remove duplicates, covering an estimated number of about 1000 valid taxa, among species and genera. Additional synonyms, further references, descriptions, feeding types, and web links ('Optional information' level) will also be included as available.

3. Results of the project:

Under the scope of the present LifeWatch Data Grant 2014 a large number of actions were performed according to the specified in the Data Grant contract,

including the creation, edition and checking of records, the addition of notes, specimens, and distributions, and the linking of all references cited in the updating process to online bibliographic resources, whenever available.

Not all the different types of actions were quantified due to the complexity of such a detailed record. The statistics provided by WoRMS show that a total of 7292 items, and 1089 taxa records were edited by the funded editor (fide Stats page). However these numbers include also actions performed previously to the period funded by the LifeWatch Data Grant 2014, from April 1st 2014 to December 15th 2014, during which the edition work covered exclusively the Syllidae. The following further details were obtained from personal registers made during the grant, from the Aphia weekly digest reports, and the WPolyDb 'Aphia' search engine for actions performed by the funded editor.

619 entries were treated, with 210 being created, 395 changed, and 30 checked without further actions (the total of these actions being higher than 619, as some entries were changed more than once).

609 of the handled entries correspond to basionyms: 58 genera, and 551 species and subspecies (of which 46 genera and 486 species are valid, being the rest synonymised taxa). These entries include 4 genera and 58 species and subspecies newly added to the database, together with further 114 new combinations, performing a total of 176 new names added.

The status of each of the 609 basionyms was revised and checked, together with their synonymies and parent taxa, original publication and other taxonomically relevant sources (with reference to pages and figures), and etymology. Besides, type specimens, whenever existing, were added for each basionym of the ranks species and subspecies (with information on museum of deposition, collection number, type locality, and geolocation), as well as notes (mainly depth range, distribution, habitat, and type locality, besides the already mentioned etymology). Distributions were also edited or added, especially whenever no information on type locality or distribution of the taxon had been previously included.

According to these actions:

581 specimens, mainly types, were added, with supporting information such as museum where they are deposited, collection number, type locality and its geolocation (being this one estimated using gazetteers, whenever not provided by authors).

3375 notes were added to the records, providing basic information about the taxa, such as depth range, distribution, etymology, habitat, and type locality, but also diagnosis of taxa, editor's comments, or type material.

228 distributions were edited or added.

3 duplicated taxa were detected and deleted from the database.

Other actions taken but not quantified included the correction and addition of synonymies, the correction of taxa status, taxa spelling, type taxa, parent taxa, or of unaccepted reason of taxa.

Considering the above numbers, it is possible to state that it was possible to fulfill beyond the 'highly desirable information' 532 valid taxa, about half the number initially estimated to be covered. Moreover, the 176 new and missing name records newly added to the database approach the estimated 200 new names to be added. It is also worth to highlight the great number of bibliographic resources which were linked, corrected and added to the literature database associated with WoRMS, as a consequence of the editing process.

4. (Brief) description of the work/methodology

Due to the urgent need to update the family Syllidae, the first action taken was the search for new and missing name records in the database, as well as the updating of the most recent taxonomic changes, like new combinations and synonymies. This work was supported by the use of Web of Science™ and Google Scholar for the search of recent bibliographic references and syllid taxa that could be missing in the WoRMS and WPolyDb databases. Besides the publications including new taxa, taxonomic revisions on the group have shown to be particularly useful.

After this first action, the taxonomic update was focused on the edition of taxa, including for each one the 'highly desirable information', namely:

a) basionym, including its current status together with synonymies and combinations, if existing;

b) the original publication reference with link to the BHL or other bibliographic resources available online (with the inclusion of abstract, DOI and other data whenever available);

c) other bibliographic references considered to be relevant for the knowledge of the taxon, with link to the BHL or other bibliographic resources available online (with the inclusion of abstract, DOI and other data whenever available);

d) type material information, such as museum where type material is deposited and its condition, collection number, type locality and its geolocation, being this one estimated whenever not provided by authors;

e) notes on the taxon, including, at least:

i) depth range

ii) distribution

iii) etymology

iv) habitat

v) type locality

The edition of taxa was always performed starting from the basionym and the original reference. After the edition of the basionym, the status of the taxon and succeeding synonymies and combinations were checked, corrected, and added, if necessary. This way it was possible to have the status of each edited taxon edited, together with all the combinations derived from it, including misspellings. Besides, in many cases much of what is known about the taxa concerning their depth range, habitat or distribution, is still what was provided with their original description, no matter the posterior taxonomic changes that they might have suffered.

5. Problems encountered and how it was solved (or expected solutions).

The main problem encountered while performing the tasks associated with the grant was the time that some actions required to be completed. These time consuming actions were particularly evident in the cases of the geolocation of type localities, the establishment of the etymology of the taxa names when not provided by the authors, or the search of particular information in the publications, like habitat, especially in those written in languages not familiar to the editor.

Many taxa were (and some still are) described without providing exact or even approximate geocoordinates of their type localities. In these cases it was necessary to geolocate them with the use of gazetteers or tools such as Marine Regions or Google Earth. This geolocation was always based on the geographic information provided with the original descriptions, such as toponyms or accompanying illustrations, such as maps or photographs, or associated with type material deposited in Natural History collections, whenever such information was available (online digitized labels, catalogues of type material, taxonomic revisions). Other problem was the use of ancient, outdated or foreign toponyms, which raised some problems in order to find the present-day toponyms and their locations. Whenever a type locality geolocation was estimated with base on gazetteers, this was always referred in the WPolyDb.

A second point was the difficulty in finding information for the required editing tasks in some of the bibliographic references. Some online translators were particularly useful to translate texts from languages with which the editor was not familiar, while the OCR option helped to retrieve information from pdfs of publications.

Finally, the whole process of editing was also more time consuming than the initially expected, reason why the initial purposes of the grant were not fully achieved by the end of the more than 250 hours of work dedicated to the edition of the Syllidae.

The editing tasks will continue for an estimated period of about 8 months, in order to reach the full taxonomic update of the whole family Syllidae. These works will follow the method of starting the edition by the basionyms, and working posteriorly the succeeding combinations and synonymies derived from them, while including the 'highly desirable information' retrieved from the bibliographic sources and linking the references to the BHL or other bibliographic resources. The duplicates that might remain in the database will be easily detected and deleted in this phase.

At the end of the process the notes containing the 'highly desirable information' of the valid taxa will be updated with base on the data contained in the 'synonymised names' of each taxon, so this information can be easily and directly retrieved from the current valid taxon name. As the feeding types in most of the cases will have to be inferred from other similar taxa, this information will be also added at the end of the updating process of the whole family.

6. Other: remarks, suggestions, other information, bibliography, …

In my opinion the WoRMS database is achieving its aim to provide the most authoritative and global list of marine taxa ever published. Moreover, it has been shown to be a very useful every day tool for the worldwide scientific community, and probably more important than that, for worldwide non scientific citizens. This is well supported by the increasing number of hits that the database receives yearly (WoRMS Users).

The evolution of the site itself shows beyond any doubt that WoRMS is developing in the good direction, integrating more and more information, and being also more user friendly.

In order to improve the database, I think it would be useful to picture or represent the type locality in a different way from the rest of the records on the distribution maps. Type locality is a very important geographic information

concerning the taxa, so I believe it should be easily and quickly identifiable when observing a distribution map of a specific taxon.

Finally, I would like to add that I hope that WoRMS and WPolyDb can be used as an authoritative and reliable tool, but without being used in a dogmatic way. First, it is a database in permanent construction, some errors, omissions and out-of-date information are just unavoidable. Secondly, WoRMS relies on Taxonomy and Systematics, humans activities with all their charge of human subjectivity. WoRMS already reflects part of this subjectivity in its database, with its 'subjective synonymies' and 'alternate representations'. It is good that users keep in mind that there will be always a certain degree of subjectivity linked to the classifications on which WoRMS relies.