LifeWatch Data Grant 2015

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

Polychaeta (Annelida)

Final Report

João Gil (aided by Geoff Read)

1. Data grant background

The World Polychaeta Database (WPolyDb) is a well established global species database within WoRMS, quickly approaching the 24,000 taxa records. As in the case of WoRMS database itself and its ancillary interfaces, the WPolyDb is being used by an increasing number of users as a reliable taxonomic tool. These users are not only relying on the accuracy of the databases, but are also demanding further categories of data related with the taxa, besides their updated taxonomy. Regardless of this demand, many of the taxonomic records in WPolyDb still include little more than the current taxonomic status of the taxon.

The present application sought a LifeWatch Data Grant to support editing priority data in WoRMS for the Polychaeta (Annelida). Thus, the main purpose of the project was to improve taxonomic data quality and to increase the diversity of available information, by providing data such as original descriptions with links to bibliographic resources, type localities with geolocation, type material with links to repository collections, environment (*e.g.* habitat and depth range), distribution, or etymology for the revised taxa. Missing authorships and new names were also to be added, and duplicates deleted.

The described actions were to be performed on taxa belonging to the Polychaeta families Acrocirridae, Alciopidae, Amphinomidae, Capitellidae, Fauveliopsidae, Lopadorrhynchidae, Nereididae, Opheliidae, Paraonide, Phyllodocidae, Pilargidae, Poebiidae, Pontodoridae, Scalibregmatidae, Spionidae, Syllidae, Terebellidae, and Trichobranchidae.

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

- Fulfil beyond the 'highly desirable information' level (basionym; original publication reference with link to the BHL or other bibliographic resources; type information such as type locality and its geolocation, museum where type material is deposited, with collection number; environment, including depth range and habitat; etymology), for the following families of Polychaeta (Annelida): Acrocirridae, Alciopidae, Amphinomidae, Capitellidae, Fauveliopsidae, Lopadorrhynchidae, Nereididae, Opheliidae, Paraonide, Phyllodocidae, Pilargidae, Poebiidae, Pontodoridae, Scalibregmatidae, Spionidae, Syllidae, Terebellidae, and Trichobranchidae, covering an estimated number of a minimum of 1250 taxa, among species and genera, or 250 working hours.

- Taxonomic update of the families covered by the application, with addition of missing taxa and authorships, and deletion of duplicates.

- Addition of ca. 100 new or missing names.

3. Results of the project:

The data grant project was developed between December 1st, 2015 and May 31st, 2016, with the following results:

About **3000** taxa (genera, species and subspecies) were examined to detect the necessity of performing urgent tasks, especially by adding authorships and publication dates, or removing duplicates. After this first step, the actions performed were as follows (statistics obtained from WPolyDb database and through the kindness of the Data Management Team):

Taxa:

A total of **498** basionyms were revised, of which:

- **88** taxa added to the database
- **364** taxa changed
- **458** taxa checked

Notes:

- **2536** notes created
- **773** notes updated

Sources:

- **15** sources created
- **291** sources updated

Specimens:

- **1243** specimens added

Some basionyms had been recently edited by other taxonomic editors, and during the present grant tasks they were revised only to add some particular detail or to link them to further information (*e.g.*: notes, specimens, literature).

4. (Brief) description of the work/methodology

The first actions performed under the scope of the present data grant were to search for missing authorships and publication dates of records already present in the database, to introduce newly published or other missing taxa, and to update the most recent taxonomic changes, especially new combinations and synonymies.

The search for missing name records was based on regular surveys using Web of Science™ and Google Scholar, to find recent bibliographic references describing new taxa, as well as by using personal lists of species made on the frame of anterior works, and also by using Olga Hartman's 1959 and 1965 publication "Catalogue of the Polychaetous Annelids of the World", as a reference work to trace old and sometimes obscure synonymised taxa.

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 updated, together with all the combinations derived from it, including misspellings.

In what concerns the bibliographic references linked to the revised basionyms, whenever necessary the bibliographic sources were linked to online databases, and as stated above, giving preference to free online services, such as Biodiversity Heritage Library or journals with online free access to their archives, whenever the work was available in more than one online resource.

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

The main problem found during the execution of the data grant was the fact that some actions to be performed could be very time consuming, delaying the rhythm of the works. Actions which could be particularly time consuming included the geolocation of type localities, the establishment of the etymology of the taxa, or retrieving relevant information from bibliographical resources in languages not familiar to the editor.

The geolocation of type localities, when a precise or even an approximate location was not given by the author, was based mainly on the use of gazetteers or tools such as Marine Regions or Google Earth. This geolocation was always based on the information provided with the original description, such as toponyms, accompanying illustrations, maps or photographs, or associated with the type material deposited in Natural History collections, if existing (online digitized labels, catalogues of type material, taxonomic revisions). Some ancient, outdated or misspelled toponyms, raised many problems in order to find the present-day equivalent toponyms and their locations. The use of ancient maps or travelling reports resulted to be particularly useful to solve some of these cases. Whenever a type locality geolocation was estimated with base on gazetteers, this was always referred in the WPolyDb.

The etymology of the taxa, whenever not clearly stated by the author, was established by checking the description of the species to determine relevant features on which the epithet could be based, with the additional use of Greek and Latin dictionaries, databases of ancient mythological figures, anatomical dictionaries, lists of prefixes and suffixes, or gazetteers, among other resources.

The difficulties with bibliographical references in languages not familiar to the editor were overcome through the use of online translators. The OCR option, present in many of the digital references, was particularly useful, not only for the use of the translators, but also to retrieve information from big publications.

Finally, the whole process of editing was more time consuming than the initially expected, and soon became evident that the initial estimate number of 1250 basionyms to be edited was too optimistic, in relation to the scheduled 250 hours of work to be dedicated to the edition of the covered families.

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

To further improve the database I think it would be useful to plot the type locality of the taxa in a distinctive way from the rest of the geographical information presented on the distribution maps, easing the location of such important information on those maps.

Besides, in order to homogenize the way the information is presented in WoRMS database, not only inside the different contexts but also through the whole database, I think it could be interesting to have a "style book" associated to the Manual with the instructions to the authors. This "style book" would include, at least for each context, the preferred way to present the information for each of the fields (for instance, the preferred way to write the type locality or to represent coordinates).

I am personally convinced that WoRMS is evolving in the right direction, by providing to the users an increasing number of options to retrieve the necessary and desired information, by presenting a growing volume of updated taxonomic information, including data associated to the covered taxa such as distribution or ecology, and by basing all the provided information on the published bibliography and on the studied biological material, through links to online bibliographical databases (many of which free, such as the Biodiversity Heritage Library), and to the existing biological collections deposited in worldwide Natural History institutions. All this is enabling WoRMS to become one of the most comprehensive taxonomic tools that ever existed since the beginning of the modern taxonomy.

Finally, I would like to thank the WoRMS' Data Management Team (VLIZ), for all their help and prompt replies during the present grant, and also the Lead Editor of Polychaeta, Dr. Geoff Read, who aided the funded editor during this data grant project, and very especially for his continuous support.

Disclaimer: Any inaccuracy, error or mistake that might exist in the present Final Report is the sole responsibility of the funded editor, João Gil. The same applies to the works performed during the present data grant.