

Review

How Aphia—The Platform behind Several Online and Taxonomically Oriented Databases—Can Serve Both the Taxonomic Community and the Field of Biodiversity Informatics

Leen Vandepitte ^{1,*}, Bart Vanhoorne ¹, Wim Decock ¹, Stefanie Dekeyzer ¹, Aina Trias Verbeeck ¹, Lorenzo Bovit ¹, Francisco Hernandez ¹ and Jan Mees ^{1,2}

¹ Flanders Marine Institute/Vlaams Instituut voor de Zee (VLIZ), Wandelaarkaai 7, Oostende 8400, Belgium; E-Mails: bart.vanhoorne@vliz.be (B.V.); wim.decock@vliz.be (W.D.); stefanie.dekeyzer@vliz.be (S.D.); aina.triasverbeeck@vliz.be (A.T.V.); lorenzo.bovit@vliz.be (L.B.); francisco.hernandez@vliz.be (F.H.); jan.mees@vliz.be (J.M.)

² Ghent University (UGent), Sint-Pietersnieuwstraat 25, Ghent 9000, Belgium

* Author to whom correspondence should be addressed; E-Mail: leen.vandepitte@vliz.be; Tel.: +32-(0)59-340-155; Fax: +32-(0)59-342-131.

Academic Editor: Angelika Brandt

Received: 14 September 2015 / Accepted: 17 November 2015 / Published: 1 December 2015

Abstract: The Aphia platform is an infrastructure designed to capture taxonomic and related data and information, and includes an online editing environment. The latter allows easy access to experts so they can update the content of the database in a timely fashion. Aphia is the core platform that underpins the World Register of Marine Species (WoRMS) and its more than 80 related global, regional and thematic species databases, but it also allows the storage of non-marine data. The content of Aphia can be consulted online, either by individual users or via machine-to-machine interactions. Aphia uses unique and stable identifiers for each available name in the database through the use of Life Science Identifiers (LSIDs). The system not only allows the storage of accepted and unaccepted names, but it also documents the relationships between names. This makes it a very powerful tool for taxonomic quality control, and also allows the linking of different pieces of information through scientific names, both within the Aphia platform and in relation to externally hosted databases. Through these LSIDs, Aphia has become an important player in the field of

(marine) biodiversity informatics, allowing interactions between its own taxonomic data and e.g., biogeographic databases. Some applications in the field of biodiversity informatics encompass the coupling of species traits and taxonomy, as well as the creation of diverse, expert validated data products that can be used by policy makers, for example. Aphia also supplies (part of) its content to other data integrators and the infrastructure can be used to host orphan databases in danger of being lost.

Keywords: Aphia; taxonomy; biodiversity informatics; online database; online editing environment; data rescue; World Register of Marine Species; web services

1. Introduction

Ever since the rise of the World Wide Web, databases have been made available online to offer people easy access to data from anywhere at any time. The biodiversity research field has played a significant role in this and, after some early experimental systems, global steps were taken in the late 1990s–early 2000s to make existing taxonomic databases available online and create new ones to meet the needs of the scientific community [1]. The main goals were to offer easy access to the scattered taxonomic data and to establish collaborations amongst these different taxonomic initiatives in order to avoid duplication of effort e.g., ([2–5]).

In parallel, biodiversity informatics came to life, stressing the importance of getting biodiversity into the digital era and—more importantly—to ensure interoperability among the variety of existing biodiversity-related databases, offering scientists the opportunity to get easy access to data and allowing them to tackle overarching biodiversity-related questions [1,3,6–8]. Although emerging in the late 1990s–early 2000s, the roots of biodiversity informatics can be traced back to Australia, where Australian herbaria had been digitizing their data cooperatively since the mid-1970s. The set-up of and experiences with the Australian Environmental Resources Information Network (ERIN) paved the way for several similar initiatives worldwide [1].

Sarkar [9] depicts biodiversity informatics “as a discipline that brings together biological information from a range of contemporary and historical sources across the spectrum of life, using organisms as the linking thread”. Several other sources/definitions/people refer to the species name as the crucial element in being able to link together information from different disciplines such as e.g., literature, geography and genetics, although this might also pose problems, e.g., due to the fact that species have received different names over time and the same name has been used for different species [8,10–12].

In the early days of biodiversity informatics, the Aphia database—hosted at the Flanders Marine Institute (VLIZ, Oostende)—was still embryonic. Aphia started as a small-scale taxonomic database with an MS Access editing interface, allowing the easy documentation of taxon names by assigning a unique identifier to each name, making links between accepted and unaccepted names, and linking this information to published sources, thereby allowing traceability of the information. Now—15 years later—it can be seen as a mature, dynamic and interoperable taxonomically-oriented data platform that allows easy access and continuous online management of taxonomic information by a worldwide team of experts—the taxonomic and thematic editors of World Register of Marine Species (WoRMS) and

other databases —, supported by a small, dedicated data management team. In the meantime, Aphia has kept track of the latest developments and needs in the field of biodiversity and biodiversity informatics: through awareness of the growing importance of linking taxonomic information with literature, specimen information, ecological traits and basic species distribution data. Based on the needs of the scientific and user communities, the Aphia structure was further extended and is now also able to store e.g., information of ecological importance such as biological and ecological species traits as well as human-defined traits, e.g., in relation to environmental legislation. IT-wise, Aphia has kept pace with the growing requirement for machine-to-machine communication, by developing a multitude of web services for both its human and machine users.

Over the years, relationships were built with other global and big regional players in the field of biodiversity and biodiversity informatics, such as the Catalogue of Life (CoL) [13], the Encyclopedia of Life (EoL) [5], the Integrated Taxonomic Information System (ITIS) [14], AlgaeBase [15], FishBase [16] and the Ocean Biogeographic Information System (OBIS) [17] and has been part of several biodiversity-related projects such as the Marine Biodiversity and Ecosystem Functioning Network of Excellence (EU-FP6—MarBEF) and the European Marine Data and Observation Network (EMODnet), representing the taxonomic backbone of these projects. Recently, the Aphia platform also became part of the LifeWatch project, where it provides a major part of the LifeWatch Taxonomic Backbone. This Taxonomic Backbone (TB) facilitates the standardization of species data and the (virtual) integration of the many distributed biodiversity data repositories and operating facilities. It virtually brings together different component databases and data systems, dealing with five major components, being taxonomy, biogeography, ecology, genetics and literature [18].

Through its original goal, Aphia is mainly marine-oriented but perfectly capable of dealing with non-marine taxonomy as well, as is demonstrated by some of its Global Species Portals.

This paper gives an overview of the history and current state of the Aphia data platform and its online editing interface. In addition, its relationship with other large data systems and projects will be explained to demonstrate how it has contributed to the field of biodiversity informatics.

2. History

Soon after its establishment in 1999, the Flanders Marine Institute (VLIZ) started building the Aphia platform—named after a fish genus—to manage TISBE, a Taxonomic Information System for the Belgian Continental Shelf, within its new Marine Data Centre [19]. The main goal of TISBE was to have a comprehensive list of all records of marine species from the Belgian part of the North Sea and the adjacent coastal areas, including the Scheldt Estuary. Aphia was built to store all the species names. Links with the taxonomic literature would help to trace the history of the documented species names and distribution records. By linking accepted and synonymized taxa, TISBE would offer the possibility of better assessing marine biodiversity and monitoring actual species composition along the Belgian coast over time [20].

When developing Aphia, international database formats were kept in mind and the original database structure of Aphia was based on the structure of the Integrated Taxonomic Information System (ITIS), hosted at the National Museum of Natural History (Washington D.C.) [14,21]. The original modules of

Aphia were similar to those available in ITIS, namely taxonomy, distributions, vernaculars, sources and comments or notes.

Five years later—in 2004—the European Register of Marine Species (ERMS) [2,22] was made available through Aphia [23,24] and, in 2006, the online Aphia editing interface became functional [25]. This online editing interface speeded up the process of adding new information to and changing existing information in ERMS, thereby serving the European scientific community by both making taxonomic information readily available and ensuring that it was as up-to-date as possible. The development of Aphia in relation to ERMS was made possible through the EU-FP6 project Marine Biodiversity and Ecosystem Functioning Network of Excellence (MarBEF) that was coordinated by the late Carlo Heip.

Around the same time, another regional register became part of Aphia: “MASDEA”, the Marine Species Database for Eastern Africa. This MS Access database was originally developed in 1996, to list all marine species from Eastern Africa, focusing on their taxonomy and distribution. In 1999, MASDEA moved to the Flanders Marine Institute (VLIZ), where an online interface was built on top [26]. It only became part of Aphia in 2004, allowing its content to be extended by making use of the already developed Aphia tools and functionalities.

From 2004 onwards, more institutes and organizations started to manage their species lists in the Aphia database, making use of the online edit interface (since 2006) and the support of the Data Management Team (DMT). In 2007, the idea was first raised to create a World Register of Marine Species (WoRMS) in Aphia [27–30].

3. Infrastructure: Database, Online Editing Environment and Available User Tools

3.1. Database

The Aphia platform is an MS SQL database. The database contains over 400 fields, spread over 81 related tables. For a full overview of the structure, see [31]. Aphia has an MS Access backend which is used for administration functions such as managing the editing rights of the involved experts (see also [29]). In addition, there is an SQL management interface which is used for bulk uploads and for quality control purposes.

Content-wise, the Aphia structure can be roughly subdivided into 10 modules (Figure 1): taxonomy, distribution, traits, specimen information, vernacular names, notes, links, images, identification keys and sources. Each module consists of several tables and interlinks with other modules and management tables. The structure of the Aphia database and its attributes/traits tables is based upon key value pairs (KVP), allowing for easy future extension without having to modify existing code or data.

All information in the database—with exception of images, see 3.1.10—is licensed under CC-BY, allowing users to share and adapt the content, provided they give appropriate credit and indicate if changes were made.

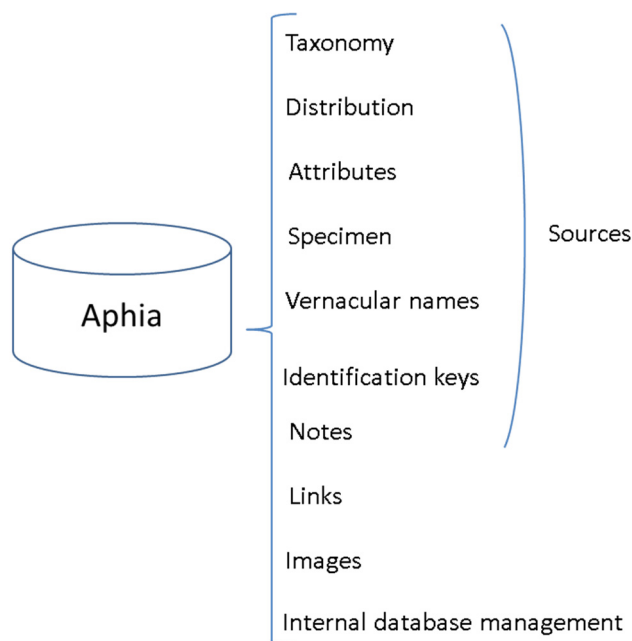


Figure 1. Conceptual representation of the modules in the Aphia database. The sources module is an overarching module: each piece of information added to Aphia—with the exception of links and images—needs to be linked to a source.

Below, we will elaborate on each module, focusing on their general content and possibilities, and noting the framework within which they were developed.

3.1.1. Taxonomy

All information added to Aphia is linked to a taxonomic name and this forms the core information of the database. All entered data needs to conform to the existing International Codes of Nomenclature [32–35]. The database allows the linking of accepted and unaccepted names, and multiple references can be added, each with a specified use (e.g., original description, source of synonymy, new combination reference...), making it possible for users to trace information. Each taxon added to Aphia receives a unique and persistent identifier, the AphiaID. This AphiaID can be expanded to a Life Science Identifier (LSID). These identifiers are a way to name and locate data on the web and ensure the enduring storage of digital objects. In a way, they can be compared to the Digital Object Identifiers (DOIs) currently used by publishers. Once assigned, the name and its corresponding ID cannot be physically deleted from the database.

Unpublished names or names of doubtful validity can be added in quarantine, meaning they are only accessible to the editors. Once their status is resolved, they can be made publicly available. The quarantine section thus serves as a kind of workbench for editors, allowing them to keep track of unpublished names.

In 2015, it became possible to document the original name combination of a species, a long-standing request from the editor community. Once fully populated, this field will facilitate making the distinction between objective synonyms (homotypic, same type) and subjective synonyms (heterotypic, different type).

3.1.2. Distribution

Distribution information is not systematically entered into the Aphia database, but the system does allow an extensive documentation of the whereabouts of a taxon. As a rule of thumb, editors are advised to document the distribution as it appears in literature, be it a local beach or a general area. Three levels of certainty or “status” for a distribution are available: valid, doubtful or inaccurate. The latter is used if an identification in a specific region is wrong, and the source mentioning this misidentification should be linked, allowing verification. It is preferable to document a distribution as inaccurate rather than removing it from the database: flagging will alert users to the problem, while removal might lead to subsequent re-entering into the database.

The distribution module of Aphia is linked to Marine Regions [36], a standardized hierarchical list of marine georeferenced place names and areas. Marine Regions also contains place names unrelated to the marine environment, e.g., all countries and the faunistic regions used by the Freshwater Animal Diversity Assessment [37], making the system also suitable for documenting non-marine place names. Through the link with Marine Regions, the Aphia system is able to generate and display distribution maps on the species web pages, based on shape files from Marine Regions. Any point location—e.g., the holotype type locality—can be plotted on top of these general shapes, providing a full overview of the documented distributions of a taxon. This visual geographical display is available to Global, Regional and Thematic Registers upon request, and is for example activated in the World Echinoidea Database [38] and illustrated on the *Psammechinus miliaris* species page of this Portal.

Since 2014, the option to document whether a species is introduced in a specific area has been added; this information is spread over three distinct fields—origin, invasiveness and occurrence—all following internationally agreed standards [39]. By storing this information in a structured way, it can easily be queried and used for statistical, analysis and management purposes.

During 2015, a new feature has been added to the distribution module, a so-called simplified, “visual distribution entry interface”. This feature allows editors to quickly select the region(s) where a species is known to occur on an interactive map and upload this to the database. Different internationally accepted standards for the regions are offered, such as sea areas defined by the International Hydrographic Organization (IHO), Exclusive Economic Zones (EEZ), Marine Regions and the Geographic Regions of the Biodiversity Information Standards (TDWG, also known as the Taxonomic Databases Working Group).

3.1.3. Attributes

After taxonomic information, there has been continuously growing demand for ecological information (e.g.[40–43]). A strong focus on traits is also apparent at the European policy level. This focus, amongst others, has been translated into the EMODnet Biology project, where the documentation of traits for European species—and specifically policy relevant species—is seen as a priority [44]. Given that the content of Aphia is also used as the taxonomic backbone for EMODnet Biology, it was logical to include attributes into Aphia, thereby establishing the link between taxonomy and attributes, but also a link between taxonomic and thematic editors. The attributes include three different types: (1) biological and ecological traits-specific characteristics of a taxon, such as body size or feeding method–,

(2) taxonomic traits—whether a taxon is part of a paraphyletic group, e.g., algae or fish—and (3) human-defined traits—whether a taxon is e.g., considered to be under threat, or if it is covered by specific legislation.

The attributes part of Aphia is designed in a generic way, allowing the capture of a diverse range of ecologically-relevant data, and is able to deal with both qualitative and quantitative data. Examples of attributes are body size, species importance to society (e.g., Habitat & Bird Directives), feeding method and host-parasite relationships. Although using key value pairs (KVP) is generic in approach, the information is stored in a structured and hierarchical way that facilitates display and the extraction of both raw data and general statistics. At the database level, attributes do not necessarily need to be assigned to a species; if an attribute is the same for a whole family, the attribute can easily be assigned to this higher level and a top-down inheritance can be applied. If exceptions exist, these can be assigned at a more finely-grained taxon level and will automatically overrule the inheritance functionality. Through this, a number of attributes can be assigned quickly to a majority of taxa and this greatly simplifies the work of the editors.

3.1.4. Specimens

The specimen module was improved and extended during 2009, to meet the needs of the Mangrove Global Species Database, known as the “Mangrove Reference Database and Herbarium” [45]. Rather than just capturing the details of the locality of the specimen, this module allows the documentation of the collection reference number, the museum where it is stored, the identifier, the collector and the preservation method. It can thus serve as a digital overview of different museum collections, allowing users to trace where the holotype of a species is kept for example, although it is not intended as a true specimen collection management tool.

3.1.5. Vernacular Names

The common names of taxa can be stored in the database. For each vernacular, the language is documented according to the ISO 639-3 standard and a reference can be added for each vernacular. Within one language, a preferred vernacular can be selected.

3.1.6. Notes

Notes contain all relevant information which does not fit in any of the other modules in a structured way. Notes are free text, but some level of organization is added by assigning a “note type”, e.g., “etymology” or “morphology”. The search interface allows users to query these notes, either based on the note type or a specific text string. It is, however, recommended that editors use notes as the last option for storing information.

3.1.7. Sources

The “sources” module is linked to several other modules: taxonomy, distribution, attributes, specimens, identification keys, common names and notes. Each piece of information added to the Aphia database should be backed up with a source. This allows users to trace information to its original source and provides a strong basis for good quality data. Until 2014, a source was entered as free text, allowing considerable variation in the format of the entered sources.

In response to a request from the experts, this single source-field was atomized into six separate fields, allowing better management of the references. The fields are DOI, author(s), year, journal, title and suffix (volume, pages...). These selected fields represent a trade-off between the complexity of atomizing reference strings and avoiding overloading editors with too much work. The system can also store the ZooBank reference LSIDs, which are used as the globally unique identifier for ZooBank registration entries [46].

A “journal importer” tool can help editors to semi-automatically import (newly) published names, based on a ZooBank Reference LSID or a ZooKeys DOI. This tool is explained in more detail under Section 3.2.

Both the free text field and the atomized fields are currently used in parallel, allowing for a transition period for the tedious task to re-format the free text to purely structured information. Editors are recommended to also upload PDFs of the references they add to the database. Aphia offers an online request system for stored PDFs, allowing users to retrieve a publication that might otherwise be hard to find.

3.1.8. Links

All links in Aphia are so-called deep links, pointing directly to a specific web page within a site. The deep links are based on a linkage between the AphiaID and the corresponding ID in another data system, e.g., Encyclopedia of Life (EoL) [5], AlgaeBase [15], ITIS [14] and the molecular databases hosted at the European Bioinformatics Institute of the European Molecular Biology Laboratory (EMBL-EBI) [47]. Through these links, Aphia users can easily gain access to relevant, non-taxonomic information. Through web services and based on these links, species related information can be retrieved from a variety of other available databases. The majority of deep links is generated automatically, whereas only a small percentage is added manually by the editors. Links are also used to display images which are hosted outside Aphia, as in the case of large specimen collections, e.g., the holotype specimen image collection of Mollusca of the Museum National d’Histoire Naturelle (Paris).

3.1.9. Identification keys

Aphia offers the possibility of creating polytomous identification keys for any taxonomic group at any chosen level. The system behind the Aphia keys is based on the NeMysKey, which was originally developed by Deprez [48] (see also [49]). At VLIZ, this NeMysKey was transformed from its original ASP software to PHP, to be compatible with the Aphia platform.

A polytomous or multi-state identification key has the advantage that an overview of all possible characteristics is available at any time during the identification process, and allows users to start with the most obvious characteristics. Based on this, the remaining set of characteristics relevant for the

further identification is narrowed down, leading to the actual identification (see also [48,50]). All listed characteristics can be defined and illustrated. The keys can also easily be updated when new species are described within a group.

The first online available key in Aphia was for Bryozoa: Key to the Caulibugula of the world. Other keys for the Bryozoa soon followed. In addition to the five Bryozoan keys, Aphia also harbours 11 Nematode identification keys, at either the species or genus level [51]. An overview of the existing Aphia ID-keys is available at [52]. Compiling these identification keys is not always straightforward and can be time-consuming, as also pointed out by Hagedorn *et al.* [50]. Keys need a high level of maintenance: because each newly described species should be added to the existing key. Ideally, keys should be compiled at a global rather than a regional level. Such online keys help scientists to share their knowledge on species identification and can be faster than paper publications as well as being more readily accessible.

In the long term, Aphia will look into the possibilities of importing existing keys developed in e.g., the DELTA software [53] and making them available online.

3.1.10. Images

Although images are not classified as core information for Aphia, their potential and importance in relation to outreach and communication cannot be underestimated. Images may reveal information on e.g., life style, size, color, habitat and feeding method of a species. Collectively, they can illustrate the diversity of life, help users verify field observations, and can be used in lectures and presentations.

When uploading images or video, the system automatically reads and displays embedded camera capture metadata and it allows the user to add minimal metadata such as a title, author, email and the name of the species depicted. The latter is offered as a drop-down list from Aphia, avoiding spelling errors or adding names not yet documented in the system. By default, images are stored under the Creative Commons license CC BY-NC-SA 3.0. The image type can be defined; images can be digital photos, microscopic images, scans of drawings, text, *etc.* The use of the Aphia image gallery—in the framework of the Canadian Register of Marine Species (CaRMS) has been described extensively [54].

Images do receive a quality assurance indication, based on whether it has been uploaded or validated by an editor or not. Editor-approved images are automatically displayed on the taxon pages, whereas unapproved images are only available through the general image gallery.

3.1.11. Internal Database Management

Unique to Aphia is the use of so-called “contexts” at the database management level. These contexts allow an easy management of specific content—be it global, regional or thematically driven—and form the backbone for the creation of distinct portals, through which editors can maintain their taxonomic register. For an overview of the existing portals we refer to Table S1. The advantage of these contexts is that all information only needs to be added once to the database; one entry can then receive multiple contexts and can thus be displayed at several locations. Based on the contexts, several statistics can be calculated, the most common ones being the total number of (accepted) species and taxa already available in a Species Register and an indication of the editor activity and general progress towards completion.

Three main categories of Species Databases can be distinguished based on the use of a context (Figure 2): Global Species Databases (GSD), Regional Species Databases (RSD) and Thematic Species Databases (TSD). GSDs are based on the higher classification of taxa. Editors can take responsibility for a specific taxonomic group—at any level of the hierarchy: e.g., Phylum, Class or Order—and maintain these through a dedicated taxonomic portal. In RSDs, taxa are grouped based on their known geographic occurrence. Editors take responsibility for a specific geographical region—e.g., Canada—and they add relevant distributions to the database. Based on this distribution information, the taxa are grouped and displayed separately in a dedicated regional portal, e.g., the Canadian Register of Marine Species [55]. For more information on how a Regional Register can be compiled, we refer to Nozères *et al.* [56]. For TSDs, taxa are grouped based on a particular characteristic. So far, three such portals exist within Aphia: the World Register of Deep-Sea Species (WoRDSS) [57], the World Register of Introduced Marine Species (WRIMS) [39] and the IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae (HAB) [58]. In addition, the content of externally hosted and managed species databases also receives a context, mainly intended for data management purposes.

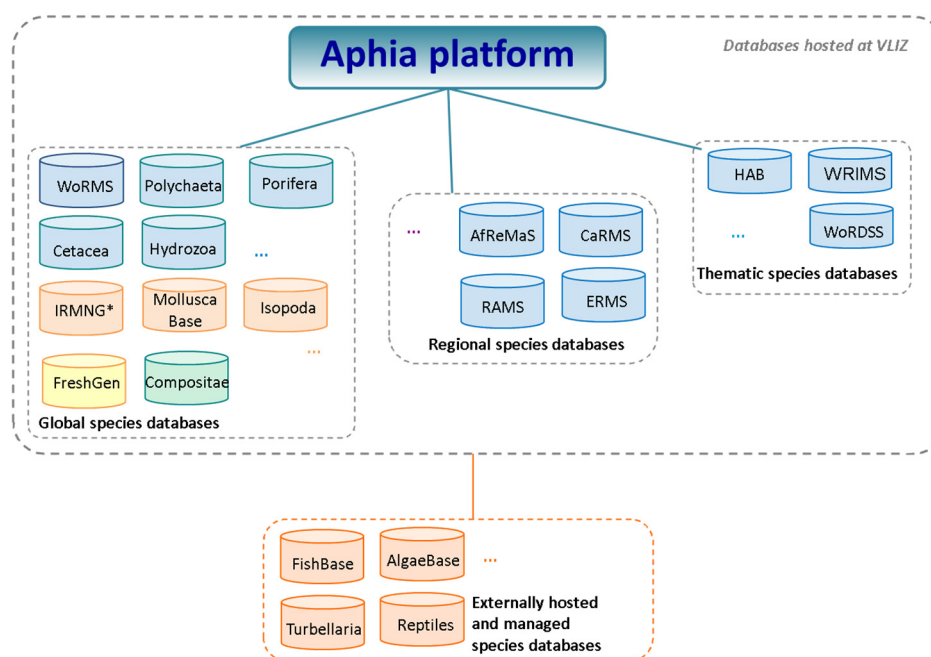


Figure 2. Conceptual representation of the relationships between the Aphia platform and its component databases. Blue: species database with focus on marine environment; Orange: species database dealing with all environments (marine-fresh-terrestrial); Yellow: species database with focus on freshwater environment; Green: species database with focus on terrestrial environment. WoRMS: World Register of Marine Species; IRMNG: Interim Register for Marine and Non-marine Genera; AfReMaS: African Register of Marine Species; CaRMS: Canadian Register of Marine Species; RAMS: Register of Antarctic Marine Species; ERMS: European Register of Marine Species; HAB: IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae; WRIMS: World Register of Introduced Marine Species; WoRDSS: World Register of Deep-Sea Species. *: IRMNG is focused on genera, not species.

In the administration part of Aphia, the editing rights of each person are defined. This includes which taxonomic group(s) each expert can edit—either as a taxonomic or thematic editor—and whether or not he or she will receive a weekly update of the changes and additions that have been made to the group(s) under their responsibility. Taxonomic editors can edit at all levels, whereas thematic editors can only edit non-taxonomy related information (e.g., distributions, traits, specimens, images, notes...). Upon login, all actions are registered in the database and an edit history is shown on the website, allowing both users and editors to verify who last edited a specific piece of information. Through the “advanced search” option, one can e.g., look for taxa that have been added by a specific person in a defined time period. This becomes very useful when several experts share responsibility for a single group. To accommodate the need for general control or follow-up of the actions taken by different editors, a weekly email alert is sent out, which is personalized and gives an overview of the latest changes and additions made by everyone involved in the management of a particular taxon group.

Back-ups of the database are taken on a daily basis (see also [29]) for disaster recovery. In addition, monthly archives of Aphia are generated, as DarwinCore Archive files. These monthly archives are taken for management and comparison purposes and are currently limited to the taxonomic information of Aphia.

One of the strengths of Aphia is that it has a permanent host institute—the Flanders Marine Institute (VLIZ)—dedicated to keeping the Aphia database and its related tools, functionalities and web services up-and-running 24/7 regardless of external project funding. However, project funding is vital, as this makes it possible to have a dedicated Data Management Team (DMT) at VLIZ, who can give full and continuous support to the editors and the large user community, and who can invest time in optimizing existing tools and creating new tools and functionalities that meet the needs of both the editors and the user communities. The daily support of the DMT is very diverse, ranging from guiding editors in the use of the online editing interface to filtering user questions, uploading large amounts of new information provided by experts directly into the database and following up on editor actions such as requesting deletions. Bulk information delivered to the DMT can be very diverse in format, ranging from Excel templates which can easily be uploaded, to Word documents or PDFs which need to be re-structured into easily uploadable Excel or Access files. Examples of the former are reference lists, distribution data, host-parasite links (traits), vernaculars and type locality data. These generally need little work to process, although a standard quality check is always performed and feedback is given to the editor prior to upload. Examples of the latter include published monographs which can fill gaps but for which the editor lacks time to fully process them and make them available. A recent example is the update of the Kinorhyncha in WoRMS, based on published work by the responsible editor. In essence, the DMT is willing to deal with any kind of data format, as long as the involved editor is prepared to give guidance and feedback during this process. All this support is handled through a dedicated email account, to which the whole DMT has access, thus ensuring a continuous follow-up of all incoming requests. In addition, a lot of technical support is provided to institutes and organizations making use of available web services.

3.2. Online Editing Environment

The online editing environment allows editors to access the database from anywhere at any time, provided they have an internet-connection. In addition to user-friendly interfaces and an editing manual [59], the online interface also offers the experts a number of useful editing tools, making their work as time-efficient as possible. One of these tools is the option to create a checklist of their specific group within Aphia, which can serve as the basis for a checklist publication. Recently, a “journal importer” tool has been developed, allowing editors to semi-automatically import (newly) published names coming from ZooBank and ZooKeys. The tool is based on a ZooKeys DOI or a ZooBank Reference LSID. When adding this, the tool will help the editor to import both the source and the taxon (or taxa) in a guided manner. The tool also automatically lists the latest ZooKeys publications, again simplifying the search and import of new taxa. A demonstration of the tool in action is available at [60].

Due to both single online and bulk offline uploads of information, a number of sources in Aphia are duplicated. Managing these duplicates has recently become a lot easier, as there is now a tool that lists possible duplicates and allows editors to indicate which source should be retained. For the source that will be deleted, all information is first transferred, ensuring that no data are lost. Next to this de-duplication of sources, a level of automatization has been developed for entering new sources: by adding a DOI or a ZooBank LSID, the system is able to automatically complete the atomized fields. Currently, the following services are used: CrossRef, ReFindit and FreeCite. In all cases, a double-check of the information is advised to verify the data.

When editors add information to WoRMS, in many cases, they go through a species list for a single location or a list of geographic locations for a single species. To make the data entry easier in these cases, a “rapid distribution entry” tool has been developed. This tool allows the system to not only “remember” the taxon or the geographic location for the next entries but also the reference and any other information linked to the record. This saves a great deal of time, as fewer selections need to be made before uploading to the database.

3.3. User Tools

Users can easily access the online portals built on the Aphia database and its user-relevant tools and functionalities. The most well-known portal at this time is the World Register of Marine Species (WoRMS) [30], but all the tools and services can run on any taxonomic list made available through the Aphia database.

Users have several search options, providing the opportunity to browse through the Aphia content. Searches can be taxon, distribution or attribute driven, e.g., list all currently accepted species within a specific taxon group and/or environment, or list all taxa within a specific geographic region. A separate search can be performed via the literature module, where the system gives an overview of all taxa and distribution information that has been retrieved from a specific source. As a user you can also check whether a specific specimen has already been documented in the database, either e.g., by identifier, museum, specimen type or preservation method. A visual representation of a taxon is—and will remain—an important feature. Through the photo gallery, images and drawings of taxa are being made available to the users, and the system offers an easy search option.

In the online interface, the editing-rights of each expert are represented as “taxonomic editor”, “thematic editor” or “other”. Information from these three levels—so-called “quality indicators”—is visually distinct, allowing users to distinguish between the different levels of reliability of any information. In general, information from taxonomic and thematic editors is categorized as trustworthy, while “other” information which has come from a variety of sources (mostly through bulk uploads in the past) and has not (yet) been reviewed by an expert, could potentially contain errors.

The Aphia Taxon Match Tool has been online since 2008 and is a freely accessible service that automatically matches your taxon list with the World Register of Marine Species (WoRMS) [61]. This tool is based on the following components: the TAXAMATCH fuzzy matching algorithm created by Tony Rees and some additional libraries. When uploading a file to the Taxon Match Tool, the final output can be defined by the user, and can contain the AphiaID, LSID, TSN (Taxonomic Serial Number of the Integrated Taxonomic Information System ITIS [14]), the accepted taxon name, authority, full classification, quality status, environment information and the citation of the individual taxon page [62].

A range of web services is already available on the Aphia database, specifically for the World Register of Marine Species [63]. These machine-to-machine services use Web Services Description Language (WSDL) and the Simple Object Access Protocol (SOAP). They allow users to e.g., consult WoRMS and perform taxonomic quality control by matching their own taxon list with the standard register of WoRMS or to retrieve common names and resolve the correct identity of unaccepted names, including their full higher classification. As the database is constantly being updated, the use of web services makes it possible to access the most recent and up-to-date information. The available web services or APIs have been published or implemented in a variety of online platforms, such as BioCatalogue [64], BiodiversityCatalogue [65], ProgrammableWeb [66], rOpenSci package [67] and gCube [68].

4. Feedback

The Aphia data platform has a multitude of users, ranging from the daily involved experts over data managers to the occasional visitor. This melting pot of users—all with their different needs and ideas about how taxonomy should be presented and what information should be available through Aphia—provides an enormous amount of feedback, both related to the content and the technical functioning of Aphia and its related species portals.

The Data Management Team acts as a filter for all this feedback, making sure that editors are not overwhelmed by questions and comments by the many users. All input related to tools and functionalities is thoroughly analyzed before any new development, and an extensive testing-phase is built in before releasing new features. This ensures that the tools provide the maximum possible value for the whole of Aphia. The majority of the editor and user oriented tools have been developed at the request of the editor and user communities. Editor-driven examples include the checklist publication tool, the rapid distribution entry tool, the quality indicators, the ability to easily document and keep track of the original name combinations, the journal importer and a tool to identify and delete similar sources. Some tools developed with user-input are the taxon match, RSS-feeds, advanced search, image search and new web services.

It is these dynamics—the interactions between editors, users and the data management team—that make it possible for the Aphia infrastructure to be flexible and to keep up with the needs and wishes of both communities.

Looking forward, it is expected that more such tools will be developed, some specifically related to the attributes data that have recently been generated.

5. Relationship with other Systems/Projects

How Aphia relates to other projects and systems can roughly be divided into four categories: (1) Aphia as a data supplier, (2) Aphia as a data integrator, (3) Aphia as a quality control tool and (4) Aphia as a platform for data rescue and an infrastructure for hosting orphaned taxonomic databases (Figure 3).

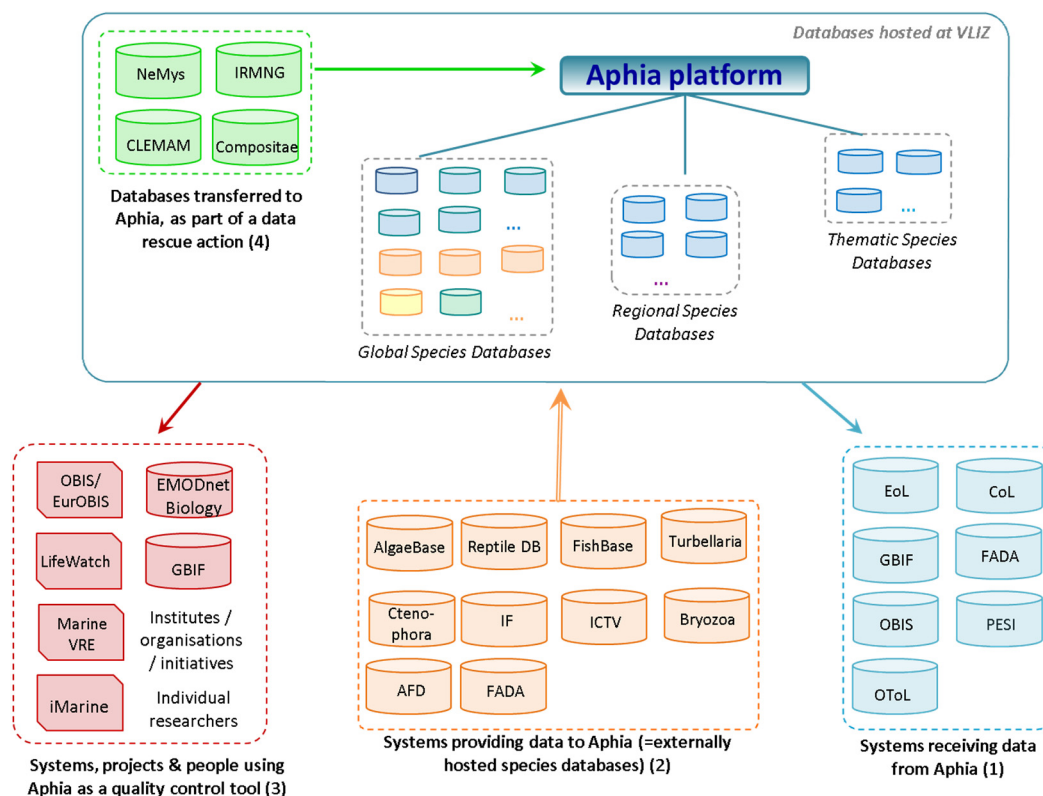


Figure 3. Relationships of Aphia with other data systems and projects. Four kinds of categories can be identified: (1) as a data supplier to other systems, (2) as an integrator of relevant data from other systems, (3) as a quality control tool for the (taxonomic) content of other systems, projects and initiatives or (4) as an infrastructure to host and rescue databases at the risk of being lost. NeMys: World List of Free-living Marine Nematodes; IRMNG: Interim Register for Marine & Non-marine Genera; CLEMAM: Checklist of European Marine Molluscs; Compositae: Global Compositae Checklist; OBIS: Ocean Biogeographic Information System; EurOBIS: European node of OBIS; Marine VRE: Marine Virtual Research Environment; GBIF: Global Biodiversity Information Facility; Turbellaria: Turbellarian Taxonomic Database; Ctenophora: Phylum Ctenophora: list of all valid species names; IF: Index Fungorum; ICTV: International Committee on Taxonomy of Viruses; Bryozoa: Recent & Fossil Bryozoa; AFD: Australian Faunal Directory; FADA: Freshwater Animal Diversity Assessment; EoL: Encyclopedia of Life; CoL: Catalogue of Life; PESI: Pan-European Species directories Infrastructure; OTOL: Open Tree of Life.

5.1. *Aphia as Data Supplier*

From within Aphia, numerous Global Species Databases are delivered to other initiatives, including the Encyclopedia of Life (EoL) [5], the Catalogue of Life (CoL) [13], the Global Biodiversity Information Facility (GBIF) [69,70] and the Pan-European Species directories Infrastructure (PESI) [71]. Recently, Aphia also started delivering data to the Open Tree of Life project (OToL) [72,73], allowing OToL to integrate the available taxonomic information within Aphia in the reconstruction of phylogenetic relationships. As both Aphia and the receiving systems are dynamic, database dumps are created automatically on a monthly basis and these are currently being picked up by CoL, EoL and GBIF, in DarwinCore-Archive files. These dumps contain taxonomy, synonyms, distributions, vernaculars, notes, sources and images. As these receiving systems are also displaying the Aphia data on their website—and are thus redistributing the data—a negotiated Memorandum of Understanding is usually in place, clearly stating what can and cannot be done with the data and explaining how both parties can benefit from each other's work. In the future, data dumps containing taxonomy and distributions will also be sent to the Freshwater Animal Diversity Assessment (FADA) [37], as part of the intensified collaboration between WoRMS and FADA in the framework of AquaRES—Aquatic Species Register Exchange and Services [74]. The AquaRES project aims to manage mixed aquatic groups (marine & freshwater) in only one system and set up a data exchange mechanism that will allow both registers to still display complete lists for their environments.

Users can also request monthly downloads of Aphia content, which can be used for institutional or personal purposes related to the quality control of their data systems, without re-distribution of the data in order to avoid different and possibly outdated versions of the database to be in circulation simultaneously.

The Ocean Biogeographic Information System [17] has recently requested an extract of the type locality data for the marine species in Aphia. This extract is currently being prepared and will allow OBIS-users to get an indication of where and when marine species were originally discovered.

Aphia provides RDF output via HTML content negotiation and via its LSID resolver. RDF—the Resource Description Framework—is a standard model for data interchange on the World Wide Web and is seen as the most important standard for machine-readable data exchange, including the semantic web.

5.2. *Aphia as Data Integrator*

Aphia can also take in data from taxonomic databases which are hosted and maintained elsewhere. Examples of such species lists are FishBase [16], AlgaeBase [15], the Reptile Database [75], Index Fungorum (IF) [76], the Freshwater Animal Diversity Assessment (FADA) [37], Recent & Fossil Bryozoa [77], International Committee on Taxonomy of Viruses (ICTV) [78], the Turbellarian Taxonomic Database [79] and Phylum Ctenophora: list of all valid species names [80]. Aphia—and its corresponding editor network—does not want to duplicate efforts: if a well-maintained and globally accepted species list already exists, agreements are sought so that their content can also be stored in Aphia and displayed through the different relevant portals, while the daily management remains the responsibility of the original host institute. For both AlgaeBase and FishBase, their respective host institutes are currently developing web services, which will speed up a semi-automated synchronization

process. During 2015, collaboration between Aphia and the Australian Faunal Directory (AFD)—an online catalogue of taxonomic and biological information on all animal species known to occur within Australia [81]—will further be discussed. AFD contains very detailed type locality information which could be shared with Aphia and be displayed on the relevant species pages.

5.3. Aphia as Quality Control Tool

The information in Aphia is highly used as a quality control tool for other databases and within several projects and initiatives.

The available web services (see 3.3 User tools) and monthly downloads (see 4.1 Aphia as data supplier) are used by a variety of organizations and people to improve and cross-check the quality of their own species lists. For WoRMS, an overview of its users is available online [82].

Within the Belgian LifeWatch e-lab, the approach is somewhat different. There, the Aphia taxon match service is offered in a Marine Virtual Environment (Marine VRE—<http://marine.lifewatch.eu>), together with other taxonomy and geography related services. The Marine VRE offers access to existing virtual labs—where data and web services can be combined to enable easy retrieval of data from different data systems, required for addressing complex, ecologically relevant questions.

A similar approach is used within the iMarine project, a Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources [83].

Aphia also provides the taxonomic backbone for several initiatives, such as (Eur)OBIS, GBIF and EMODnet Biology. Here, the content of Aphia is used to cross-check the taxonomic names available in each of these data systems. The idea is that each name in those systems is either linked to a name in Aphia, or an explanation is provided as to why a name cannot be added to Aphia. The latter can e.g. concern non-sense, unpublished or non-taxonomic names.

Not only the taxonomy, but also the distribution information in Aphia can be used in quality control purposes. As an extra functionality in Aphia, the occurrence information of OBIS can be combined with the distribution maps in Aphia, which are based on literature and expert input. The visual combination of Aphia and OBIS data can serve as a quality control for both systems: any major deviation between the Aphia and OBIS information could point to a data gap in Aphia or could indicate possible errors in OBIS. Both outcomes can help improve the content and quality of the two data systems.

5.4. Aphia as a Platform for Data Rescue and Infrastructure for Hosting Other Taxonomic Databases

Given its stable situation and extensive structure for the capture of content, the Aphia platform is also able to ingest databases at risk of disappearance. The use of the Aphia infrastructure is offered to taxonomic initiatives in need of a permanent host institute and continuous support of a data management team, to protect their data from becoming lost for the scientific community. In the past, several offline Global Species Databases—only available as Excel sheets or Access databases—have been integrated into Aphia, thereby contributing to the World Register of Marine Species. Some examples are the World Database of Proseriata and Kalyptorhynchia [84], the World Nemertea Database (not yet online) and MASDEA, the Marine Species Database for Eastern Africa [26] which is now continued as AfReMaS, the African Register of Marine Species [85].

More recently, three existing online databases were integrated into the Aphia infrastructure, being NeMys—the World List of Free-living Marine Nematodes—[51], CLEMAM—the Checklist of European Marine Mollusca [86]—and IRMNG, the Interim Register for Marine and Non-marine Genera [87].

Reasons for migrating to the Aphia platform can be multiple, e.g., the fact that the host institute can no longer maintain the database and its online services, or it is felt that Aphia would provide a more stable platform for continuing management of the database.

In collaboration with the original manager and host institute, an import and future management plan is designed. The structure of the original database is compared to Aphia and all available fields are mapped accordingly. The data import is followed by a quality control and feedback to the original manager, ensuring all information is captured and interpreted correctly. Once this has been done, the management of the data is left to the original provider—with support of the Data Management Team—and in most cases, a more elaborate editor network is put together, allowing an even better and more up-to-date management of the content.

Although the content of Aphia has mostly been marine—with a slight extension to the freshwater and terrestrial environment for some groups—the system is capable of dealing with purely non-marine taxa. Early 2015, the manager of the World List of Compositae has requested to make use of the Aphia infrastructure for the future hosting and maintenance of this database. The data transfer itself will take place during 2016.

6. Aphia—What Can the Future Bring?

Aphia is an elaborate infrastructure capable of dealing with many kinds of content and of putting this content to use in multiple ways. The strength of Aphia partly lies in its capability to link with other data systems and to offer tools that can combine content of different nature and systems, and the fact that there is continuous support available from a Data Management Team (DMT).

Immediate applications of these combinations are seen in coupling traits and taxonomy from the World Register of Marine Species (WoRMS)—hosted in Aphia—with distribution information coming from biogeographic databases, such as (Eur)OBIS. This would allow comparing of the distribution of taxa across e.g., depth gradients, and the exploration of whether this is related to any other traits, such as feeding method or body size. If OBIS would tap into the Aphia attributes for WoRMS as well as its taxonomy, more targeted data selections could be made, leading to more accurate datasets for additional analyses. Examples of combining specific traits with taxonomic and biogeographic data are given in the discussion of Costello *et al.* [43]. Currently, the EMODnet Biology Portal has implemented this combination and allows to search for species and their distribution across Europe, based on human-defined traits (Figure 4). Of particular importance within EMODnet Biology is the distribution in space and time of policy-relevant species, e.g., in the framework of the Marine Strategy Framework Directive (MSFD) and the assessment of Good Environmental Status (GES) [88] of the regional areas within Europe. To assist in this process, the different regional MSFD related indicators have been uploaded as traits in the Aphia database. Linking this to EurOBIS now allows easy grouping and distribution data retrieval for these species.

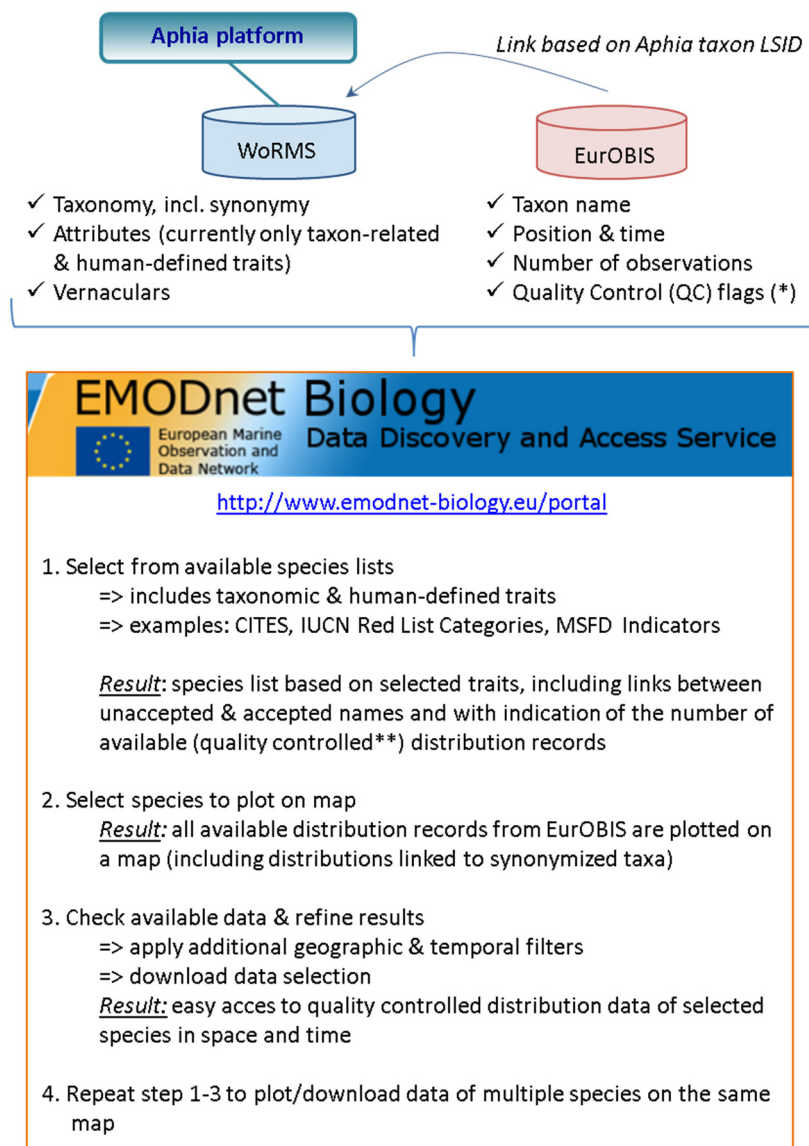


Figure 4. Schematic representation of how WoRMS and EurOBIS are brought together into the EMODnet Biology Data Portal, allowing a combination of taxonomy, traits and distribution information. EMODnet Biology can be seen as a “layer” on top of the EurOBIS database, and is pulling additional taxonomic and attribute data—based on the Aphia taxon LSID—from WoRMS, which is hosted in the Aphia platform. This unique combination allows searching for distribution data based on traits, benefiting from the strengths of both systems (e.g., quality control on the EurOBIS data and expert-validated information in WoRMS). *: the quality control procedures on EurOBIS data are described by Vandepitte *et al.* [89]; **: for EMODnet Biology, each data record has to comply to a minimum quality, related to taxonomy and geography before it is made available through the data portal [89].

Mapping expert-validated species distribution information from Aphia together with occurrence information from biogeographic databases such as OBIS can help identifying gaps in both systems. When mapping these occurrences on top of each other, clear gaps in literature-information can become clear, when OBIS data cover an area not yet documented in the literature or by experts in Aphia.

Conversely, Aphia distribution areas not covered by OBIS can indicate gap in OBIS, stimulating OBIS to actively seek out data from those areas. Such a reciprocal comparison could potentially also uncover errors in OBIS, for example as a consequence of erroneous species identification. A random check by Vandepitte *et al.* [89] revealed a potential identification error in OBIS, which was confirmed by the taxonomic expert for that particular species. Based on this, the error in OBIS could be annotated. Although this was a random, small-scale check, the potential of a large-scale, structured comparison of both systems is enormous. It would not only prove to be an excellent quality control tool, but it would also be an important way of feeding both systems with more data, thereby allowing a more accurate estimate of geographical coverage of species and thus support better scientific analyses related to species distribution patterns. Each system however has its limitations which need to be taken into account. In general, taxonomic databases might be incomplete or have a backlog in keeping up with recent taxonomic findings whereas biogeographic databases can e.g., struggle with sampling bias in the available data or the quality and completeness of the data e.g., [89–91]. Users need to be aware of these limitations, so they can take them into account when putting the data to use.

Based on the available—expert-validated—information on taxonomy, attributes and biogeography, scientists can create data products. Within the European EMODnet Biology project, the need for four categories of priority products was described: (1) species distribution maps and trends; (2) species sensitivity and vulnerability maps; (3) species attributes and (4) biodiversity indices. Ideally, these products should serve different user communities including scientists, policy makers, practitioners and the public [44]. All products however relate to taxonomy, distribution and/or attributes and need high quality data to ensure reliability.

Looking towards the future, Aphia could play an important role in making available online identification keys with a global coverage and offering the scientific community easy ways of accessing the keys and of keeping them updated. The keys can be maintained through the online platform. Numerous keys already exist, but mainly in paper format and they are not always easy to get hold of. Many of these paper keys have never been updated, or updates are scattered and need to be consolidated. The online tool offers experts one location to store all the relevant information, including online identification key as well as relevant publications via the “sources” module. Ideally, mostly global keys would be made available, allowing the world-wide scientific community to make use of this and avoid confusion in identifying similar species from different regions or misidentifying a newly arrived alien species. Scientists describing new species could alert the manager of the relevant keys, and—together—they can add the new species to the key.

From a broader perspective, the Aphia infrastructure can serve as a general platform for the management of taxonomic databases, either marine or non-marine. The potential of this has been demonstrated, as a number of external databases have already been successfully incorporated into the Aphia structure. So far, such data rescue actions have been upon request. There has been no general advertising of this service, as it does involve a certain amount of manual work from both the original data manager and the Aphia Data Management Team to bring such an effort to a successful conclusion, especially if further technical support was previously in place towards third parties. The Data Management Team maintains an active watch for existing databases that are in danger of being lost, so that timely action can be taken if necessary.

Acknowledgments

The authors would like to thank the WoRMS Steering Committee and the current and former members of the WoRMS Editorial Board for their continuous efforts in feeding the database, their feedback on new developments and features and their many suggestions for the improvement of Aphia in general, be it structure- or content-wise or in relation to the online editing interface.

Over the years, many people have temporarily been part of the Data Management team or the IT-staff at the Flanders Marine Institute (VLIZ) and have contributed to the content of Aphia or have made developments to improve the system. In addition, a special thank you to our many users who have provided valuable feedback over the years, which has also led to the improvement of both content and technical functioning of the Aphia infrastructure.

Special thanks to Geoff Boxshall (Natural History Museum, London, UK) for proof-reading the final version of the manuscript and improving the language and the three anonymous referees for their constructive feedback on an earlier version of the manuscript.

The creation and continuous further development of Aphia is possible through numerous projects: MarBEF Network of Excellence “Marine Biodiversity Ecosystem Functioning” (EU FP6, GOCE-CT-2003-505446; 2004–2009), Pan-European Species directories Infrastructure (PESI, EU-FP7, RI-223806; 2009–2011), EMODnet Biology I (MARE/2008/03 -Lot 4 Biology—Contract nr SI2.531562; 2009–2011), 4D4Life (INFRA-2008-1.2.2—Scientific Data Infrastructures—Grant Agreement 238988; 2009–2012), EMODnet Biology II (MARE/2012/10—Lot 5 Biology—Contract No. SI2.657705; 2012–2016), LifeWatch (2012–2016), AquaRES (BR/132/A6/AQUARES; 2014–2016).

Next to structured project funding, a number of *ad hoc* funds have also contributed to the further development of Aphia and its portals: Alfred P. Sloan Foundation (through the Census of Marine Life, Memorial University, Canada), the Global Biodiversity Information Facility (GBIF), the Department of Fisheries and Oceans (DFO, Canada), the Marine Environmental Data and Information Network (MEDIN, UK) and the Coordinating Research on the North Atlantic (CORONA) project of the National Science Foundation (NSF, USA).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Canhos, V.P.; Souza, S.; Giovanni, R.; Canhos, A.L. Global biodiversity informatics: Setting the scene for a “new world” of ecological modeling. *Biodivers. Inf.* **2004**, *1*, doi:10.17161/bi.v1i0.3.
2. Costello, M.J. Developing species information systems: The European Register of Marine Species (ERMS). *Oceanography* **2000**, *13*, 48–55.
3. Gewin, V. Taxonomy: All living things, online. *Nature* **2002**, *418*, 362–363.
4. Polaszek, A. A universal register for animal names. *Nature* **2005**, *437*, 477, doi:10.1038/437477a.
5. Parr, C.S.; Wilson, N.; Leary, P.; Schulz, K.S.; Lans, K.; Walley, L.; Hammock, J.A.; Goddard, A.; Rice, J.; Studer, M.; Holmes, J.T.G.; *et al.* The Encyclopedia of Life v2: Providing Global Access to Knowledge about Life on Earth. *Biodivers. Data J.* **2014**, *2*, e1079.

6. Sugden, A.; Pennisi, E. Diversity digitized. *Science* **2000**, *289*, 2305.
7. Bisby, F.A. The quiet revolution: Biodiversity informatics and the internet. *Science* **2000**, *289*, 2309–2312.
8. Hardisty, A.; Roberts, D. The Biodiversity Informatics Community (2013). A decadal view of biodiversity informatics: Challenges and priorities. *BMC Ecol.* **2013**, *13*, doi:10.1186/1472-6785-13-16.
9. Sarkar, I.N. Biodiversity informatics: Organizing and linking information across the spectrum of life. *Brief. Bioinf.* **2007**, *8*, 347–357.
10. Schalk, P.H. Management of marine natural resources through by biodiversity informatics. *Mar. Policy* **1998**, *22*, 269–280.
11. Page, R.D.M. Biodiversity informatics: The challenge of linking data and the role of shared identifiers; *Brief. Bioinf.* **2008**, *9*, 345–354.
12. Patterson, D.J.; Cooper, J.; Kirk, P.M.; Pyle, R.L.; Remsen, D.P. Names are key to the big new biology. *Trends Ecol. Evol.* **2010**, *25*, 686–691.
13. Catalogue of Life. Available online: <http://www.catalogueoflife.org> (accessed on 1 July 2015).
14. Integrated Taxonomic Information System (ITIS). Available online: <http://www.itis.gov> (accessed on 1 July 2015).
15. AlgaeBase. Available online: <http://www.algaebase.org> (accessed on 3 July 2015).
16. FishBase. Available online: <http://www.fishbase.org> (April 2015).
17. Ocean Biogeographic Information System. Available online: <http://www.iobis.org> (accessed on 2 July 2015).
18. LifeWatch Regional Portal Home Page. Available online: <http://www.lifewatch.be> (accessed on 1 July 2015).
19. *Taxonomic Information System for the Belgian Coastal Area (TISBE)*; Flanders Marine Institute: Oostende, Belgium, 2004.
20. Vanden Berghe, E.; Mees, J.; Seys, J.; Cattrijsse, A.; Bulckaen, B. TISBE: Taxonomic Information System for the Belgian Continental Shelf. In *VLIZ Young Scientists' Day, Brugge, Belgium, 13 March 2002: Book of Abstracts*, VLIZ Special Publication 7; Mees, J., Seys, J., Eds.; Vlaams Instituut voor de Zee: Oostende, Belgium, 2002; p. 50.
21. Appeltans, W.; Vanden Berghe, E.; Mees, J. A taxonomic and biogeographic information system of marine species in the southern North Sea developed by Flanders Marine Institute. In *Proceedings of the Ocean Biodiversity Informatics*; Hamburg, Germany, 29 November–1 December 2004; p. 54.
22. Costello, M.J.; Emblow, C.; White, R.J. European register of marine species: A check-list of the marine species in Europe and a bibliography of guides to their identification. In *Collection Patrimoines Naturels*, 50; Muséum National d'Histoire Naturelle: Paris, France, 2001; p. 463.
23. Boxshall, G. Updating, expanding ERMS, the European Register of Marine Species. Available online: <http://www.marbef.org/modules.php?name=People&module=ref&refid=66315> (accessed on 19 November 2015)
24. Vanden Berghe, E. The European Register of Marine Species: Tool for data integration. In *Proceedings of the ICES 2006 Annual Science Conference*, Maastricht, The Netherlands, 17–26 September 2006.

25. Cuvelier, D.; Claus, S.; Appeltans, W.; Vanhoorne, B.; Vanden Berghe, E.; Costello, M.J. European Register of Marine Species (ERMS)—Plans turning into reality! *MarBEF Newsl.* **2006**, *4*, 14–15.
26. Vanden Berghe, E. MASDEA: Marine species database for Eastern Africa. *Indian J. Mar. Sci.* **2005**, *34*, 128–135.
27. Appeltans, W.; Costello, M.J.; Vanhoorne, B.; Decock, W.; Vandepitte, L.; Hernandez, F.; Mees, J.; Vanden Berghe, E. Aphia for a World Register of Marine Species (WoRMS). In *VLIZ Young Scientists' Day, Brugge, Belgium, 29 February 2008: Book of Abstracts*, VLIZ Special Publication 40; Mees, J., Seys, J., Eds.; Vlaams Instituut voor de Zee: Oostende, Belgium, 2008; p. 27.
28. Appeltans, W.; Decock, W.; Vanhoorne, B.; Hernandez, F.; Bouchet, P.; Boxshall, G.; Fauchald, K.; Gordon, D.P.; Poore, G.C.B.; van Soest, R.; *et al.* The World Register of Marine Species: An authoritative, open-access web-resource for all marine species. In *Proceedings of the Future of the 21st Century Ocean: Marine Sciences and European Research Infrastructures, an International Symposium*, Brest, France, 28 June–1 July 2011.
29. Costello, M.J.; Bouchet, P.; Boxshall, G.; Fauchald, K.; Gordon, D.; Hoeksema, B.W.; Poore, G.C.B.; van Soest, R.W.M.; Stöhr, S.; Walter, T.C.; *et al.* Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS ONE* **2013**, *8*, doi:10.1371/journal.pone.0051629.
30. WoRMS Editorial Board. World Register of Marine Species. Available online: <http://www.marinespecies.org> (accessed on 3 July 2015).
31. SchemaSpy analysis of the Aphia Database. Available online: <http://www.marinespecies.org/structure/> (accessed on 31 August 2015).
32. Lapage, S.P.; Sneath, P.H.A.; Lessel, E.F.; Skerman, V.B.D.; Seeliger, H.P.R.; Clark, W.A., Eds.; In *International Code of Nomenclature of Bacteria: Bacteriological Code, 1990 Revision*; ASM Press: Washington, DC, USA, 1992; p. 232.
33. International Commission on Zoological Nomenclature. *International Code of Zoological Nomenclature, Adopted by the International Union of Biological Sciences*, 4th ed.; International Trust for Zoological Nomenclature: London, UK, 1999; p. 306.
34. King, A.M.Q.; Adams, M.J.; Carstens, E.B.; Lefkowitz, E.J. *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses*; Elsevier Academic Press: San Diego, CA, USA, 2012.
35. McNeill, J.; Barrie, F.R.; Buck, W.R.; Demoulin, V.; Greuter, W.; Hawksworth, D.L.; Herendeen, P.S.; Knapp, S.; Marhold, K.; Prado, J.; *et al.* *International Code of Nomenclature for Algae, Fungi, and Plants (Melbourne Code). Regnum Vegetabile 154*; Koeltz Scientific Books: Königstein, Federal Republic of Germany, 2012.
36. Claus, S.; de Hauwere, N.; Vanhoorne, B.; Deckers, P.; Souza Dias, F.; Hernandez, F.; Mees, J. Marine Regions: Towards a global standard for georeferenced marine names and boundaries. *Mar. Geod.* **2014**, *37*, 99–125.
37. Balian, E.V.; Segers, H.; Lévêque C.; Martens, K. An introduction to the Freshwater Animal Diversity Assessment (FADA) project. *Hydrobiologia* **2008**, *595*, 3–8.
38. Kroh, A.; Mooi, R. World Echinoidea Database. Available online: <http://www.marinespecies.org/echinoidea> (accessed on 13 October 2015).

39. Pagad, S.; Hayes, K.; Katsanevakis, S.; Costello, M.J. World Register of Introduced Marine Species (WRIMS), 2015. Available online: <http://www.marinespecies.org/introduced> (accessed on 3 July 2015).
40. Baird, D.J.; Rubach, M.N.; van den Brink, P.J. Trait-based ecological risk assessment (TERA): The new frontier? *Integr. Environ. Assess. Manag.* **2008**, *4*, doi:10.1897/IEAM_2007-063.1.
41. Green, J.L.; Bohannon, B.J.M.; Whitaker, R.J. Microbial Biogeography: From taxonomy to traits. *Science* **2008**, *320*, 1039–1043.
42. Menezes, S.; Baird, D.J.; Soares, A.M.V.M. Beyond taxonomy: A review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. *J. Appl. Ecol.* **2010**, *47*, 711–719.
43. Costello, M.J.; Claus, S.; Dekeyser, S.; Vandepitte, L.; O Tuama, E.; Lear, D.; Tyler-Waters, H. Biological and ecological traits of marine species. *Peer J.* **2015**, *3*, doi:10.7717/peerj.1201.
44. Claus, S.; Vandepitte, L.; Hernandez, F.; Mees, J. *Report of the Biological Data Products Workshop of the European Marine Observation and Data Network (EMODnet)*; Vlaams Instituut voor de Zee (VLIZ): Oostende, Belgium, 2010; p. 41.
45. Massó i Alemán, S.; Bourgeois, C.; Appeltans, W.; Vanhoorne, B.; de Hauwere, N.; Stoffelen, P.; Heughebaert, A.; Dahdouh-Guebas, F. The “Mangrove Reference Database and Herbarium”. *Plant Ecol. Evol.* **2010**, *143*, 225–232.
46. Pyle, R.L.; Michel, E. ZooBank: Developing a nomenclatural tool for unifying 250 years of biological information. *Zootaxa* **2008**, *1950*, 39–50.
47. European Bioinformatics Institute. Available online: <http://www.ebi.ac.uk> (accessed on 3 July 2015).
48. Deprez, T. Taxonomy and biogeography of the Mysida (Peracarida, Crustacea): A global approach through the biological information system NeMys. Ph.D. Thesis, Universiteit Gent, Gent, Belgium, 2006.
49. Deprez, T.; Merckx, B.; Vincx, M. Online identification of Mysida through NeMys. In *VLIZ Young Scientists' Day, Brugge, Belgium, 25 February 2005: Book of Abstracts*, VLIZ Special Publication 20; Mees, J., Seys, J., Eds.; Vlaams Instituut voor de Zee: Oostende, Belgium, 2005; p. 31.
50. Hagedorn, G.; Rambold, G.; Martellos, S. Types of identification keys. In *Tools for Identifying Biodiversity: Progress and Problems*; Nimis P.L., Vignes Lebbe, R., Eds.; Edizioni Università di Trieste: Trieste, Italy, 2010; pp. 59–64.
51. Vanaverbeke, J.; Bezerra, T.N.; Braeckman, U.; de Groote, A.; de Meester, N.; Deprez, T.; Derycke, S.; Gilarte, P.; Guilini, K.; Hauquier, F.; *et al.* NeMys: World Database of Free-Living Marine Nematodes. Available online: <http://nemys.ugent.be> (accessed on 3 July 2015).
52. WoRMS ID Keys. Available online: <http://www.marinespecies.org/aphia.php?p=idkeys> (accessed on 9 September 2015).
53. Dallwitz, M.J.; Paine, T.A.; Zurcher, E.J. 1993 onwards. User's guide to the DELTA System: A General System for Processing Taxonomic Descriptions; 4th ed. Available online: <http://delta-intkey.com/www/uguide.htm> (accessed on 21 August 2015).
54. Kennedy, M.K.; Nozères, C.; Miller, R.; Vanhoorne, B.; Appeltans, W. The Canadian Register of Marine Species Photo Gallery: A User's Guide, Version 1. Available online: <http://www.marinespecies.org/imis.php?module=ref&refid=206045> (accessed on 19 November 2015).

55. Kennedy, M.K., van Guelpen, L., Pohle, G., Bajona, L., Eds.. Canadian Register of Marine Species. World Wide Web electronic publication, version 1.0/2010; Available online: <http://www.marinespecies.org/carms/> (accessed on 1 July 2015).
56. Nozères, C.; Vandepitte, L.; Appeltans, W.; Kennedy, M. *Best Practice Guidelines in the Development and Maintenance of Regional Marine Species Checklists*, version 1.0; Global Biodiversity Information Facility: Copenhagen, Denmark, 2012; p. 27.
57. Glover, A.G.; Higgs, N.; Horton, T. World Register of Deep-Sea Species; Available online: <http://www.marinespecies.org/deepsea> (accessed on 6 July 2015).
58. Moestrup, Ø., Akselmann, R., Fraga, S., Hansen, G., Hoppenrath, M., Iwataki, M., Komárek, J., Larsen, J., Lundholm, N., Zingone, A., Eds. IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae. Available online: <http://www.marinespecies.org/hab> (accessed on 6 July 2015).
59. WoRMS Instructions to Editors. Available online: <http://www.marinespecies.org/tutorial.php> (accessed on 6 July 2015).
60. WoRMS Photogallery. Available online: <http://www.marinespecies.org/photogallery.php> (accessed on 6 July 2015).
61. Vanhoorne, B.; Deneudt, K.; Appeltans, W.; Hernandez, F.; Mees, J. The Aphia Taxon Match Tool, an online quality control tool for checking your biological data against the World and European Registers of Marine Species. In *Proceedings of the International Marine Data and Information Systems Conference IMDIS-2008*, Athens, Greece, 31 March–2 April 2008; Sissy, I., Maillard, C., Tosello, V.. Eds.; 2008; p. 145.
62. WoRMS Taxon Match. Available online: <http://www.marinespecies.org/aphia.php?p=match> (accessed on 6 July 2015).
63. Deneudt, K.; Vanhoorne, B.; Appeltans, W.; Hernandez, F. Talking to the WoRMS: What can VLIZ web services do for you? In *Book of abstracts - VLIZ Young Marine Scientists' Day. Brugge, Belgium, 24 February 2012*; Mees, J., Seys, J., Eds.; VLIZ Special Publication: Brugge, Belgium, 2012; Volume 55, p. 23.
64. Bhagat, J.; Tanoh, F.; Nzuobontane, E.; Laurent, T.; Orlowski, J.; Roos, M.; Wolstencroft, K.; Aleksejevs, S.; Stevens, R.; Pettifer, S.; *et al.* BioCatalogue: A universal catalogue of web services for the life sciences. *Nucl. Acids Res.* **2010**, *38*, 689–694.
65. BiodiversityCatalogue: Providing a Curated Catalogue of Biodiversity Web Services. Available online: <https://www.biodiversitycatalogue.org/> (accessed on 9 September 2015).
66. ProgrammableWeb. Available online: <http://www.programmableweb.com/> (accessed on 9 September 2015).
67. Transforming Science through Open Data Home Page. Available online: <https://ropensci.org/> (accessed on 23 August 2015).
68. gCube Framework Home Page. Available online: <https://www.gcube-system.org/> (accessed on 23 August 2015).
69. Edwards, J.L.; Lane, M.A.; Nielsen, E. Interoperability of biodiversity databases: Biodiversity information on every desktop. *Science* **2000**, *289*, 2312–2314.
70. Global Biodiversity Information Facility Home Page. Available online: <http://www.gbif.org> (accessed on 23 August 2015).

71. EU-NOMEN—Pan-European Species directories Infrastructure. Available online: <http://www.eunomen.eu> (accessed on 23 August 2015).
72. Open Tree of Life. Available online: <https://tree.opentreeoflife.org> (access on 12 October 2015)
73. Hinchliff, C.E.; Smith, S.A.; Allman, J.F.; Chaudhary, R.; Coghill, L.M.; Crandall, K.A.; Deng, J.; Drew, B.T.; Gazis, R.; Gude, K.; *et al.* Synthesis of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, doi:10.1073/pnas.1423041112.
74. AquaRES. Available online: <http://odnature.naturalsciences.be/aquares/> (accessed on 23 August 2015).
75. Uetz, P., Hošek, J., Eds. The Reptile Database. 2015. Available online: <http://www.reptile-database.org> (accessed on 2 July 2015).
76. Index Fungorum. Index Fungorum Partnership: Landcare Research-NZ and RBG Kew: Mycology. Available online: <http://www.indexfungorum.org> (accessed on 30 June 2015).
77. Bock, P. World list of Recent and Fossil Bryozoa. Available online: <http://www.bryozoa.net> (accessed on 30 June 2015).
78. International Committee on Taxonomy of Viruses. Available online: <http://www.ictvonline.org> (accessed on 30 June 2015).
79. Tyler S.; Schilling, S.; Hooze, M.; Bush, L.F. (compiler), (2006–2013) Turbellarian taxonomic database; Version 1.7. Available online: <http://turbellaria.umaine.edu> (accessed on 7 July 2015).
80. Mills, C.E. Phylum Ctenophora: List of all Valid Species Names. Available online: <http://faculty.washington.edu/cemills/Ctenolist.html> (accessed on 1 July 2015).
81. Australian Biological Resources Study. Australian Faunal Directory. Available online: <http://www.environment.gov.au/biodiversity/abrs/online-resources/fauna/afd/index.html> (accessed on 03 July 2015).
82. Users. Available online: <http://www.marinespecies.org/users.php> (accessed on 8 July 2015).
83. iMarine—Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources. Available online: <http://www.i-marine.eu/Pages/Home.aspx> (accessed on 8 July 2015).
84. Artois, T.; Schockaert, E. World database of Proseriata and Kalyptorhynchia. Available online: <http://www.marinespecies.org/rhabditophora> (accessed on 6 July 2015).
85. Odido, M., Appeltans, W., BelHassen, M., Mussai, P., Nsiangango, S.E., Vandepitte, L., Wambiji, N., Zamouri, N. Jiddou, A.M., Eds. African Register of Marine Species. Available online: <http://www.marinespecies.org/afremas> (accessed on 6 July 2015).
86. Le Renard, J.; Gofas, S. CLEMAM—Check List of European Marine Mollusca. Available online: <http://www.somali.asso.fr/clemam/index.clemam.html> (accessed on 10 July 2015).
87. Rees, T. (compiler), 2006-present. The Interim Register of Marine and Nonmarine Genera (IRMNG). Available online: <http://www.obis.org.au/irmng> (accessed on 2 July 2015).
88. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive). Available online: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32008L0056> (accessed on 11 October 2015).

89. Vandepitte, L.; Bosch, S.; Tyberghein, L.; Waumans, F.; Vanhoorne, B.; Hernandez, F.; de Clerck, O.; Mees, J. Fishing for data and sorting the catch: Assessing the data quality, completeness and fitness for use of data in marine biogeographic databases. *Database* **2015**, doi:10.1093/database/bau125.
90. Hjarding, A.; Tolley, K.A.; Burgess, N.D. Red list assessments of East African chameleons: A case study of why we need experts. *Oryx* **2014**, *49*, doi:10.1017/s0030605313001427.
91. Meyer, C.; Kreft, H.; Guralnick, R.; Jetz, W. Global priorities for an effective information basis of biodiversity distributions. *Nat. Commun.* **2015**, *6*, doi:10.1038/ncomms9221.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).