AquaRES follow-up committee meeting 24/06/14

Participants

Follow-up committee

- Dimitri Brosens, Belgian Biodiversity Platform/Research Institute for Nature and Forrest (INBO)
- Koen Martens, Royal Belgian Institute of Natural Sciences (RBINS)/Freshwater Animal Diversity Assessment (FADA)
- Wouter Addink, Naturalis/Species 2000
- Saskia Van Gaever, Federal Public Service Health, Food Chain Safety and Environment - Marine Environment
- Maaike Vancauwenberghe, Belspo/BRAIN-Axis 6
- Ward Appeltans, Intergovernmental Oceanographic Commission of UNESCO/Ocean Biogeographic Information System (OBIS)
- Geoff Boxshall, The Natural History Museum/WoRMS Steering committee
- Wouter Los, University of Amsterdam/LifeWatch
- Anton Vandeputte, biodiversity.aq (through Skype)
- Nabil Youdjou, biodiversity.aq

Excused:

- Peter Schalk, Naturalis/Species 2000/GBIF Governing board

Project partners

- Bruno Danis, Université Libre de Bruxelles
- Leen Vandepitte, Flanders Marine Institute (VLIZ)
- Michel Kapel, RBINS
- Aaike De Wever, RBINS/coordinator AquaRES

Excused:

- Bart Vanhoorne, VLIZ

Introduction

In this report I will focus on the questions, discussion topics and recommendations. The topics discussed are regrouped under the originally proposed headings and do not necessarily represent the order in which they were discussed during the meeting. Recommendations that can/should be thoroughly explored or implemented on short term are indicated in **bold**.

Presentations

The World Register of Marine Species (WoRMS) and the Aphia database system: development, status and functionalities (Leen Vandepitte)

Q: How are invasive species data managed in Aphia
A: Information is stored as species in combination with the region in which it is invasive. The data itself is delivered through Shyama Pagad, member of the Invasive Species Specialist Group (ISSG). The Aphia database is used as a tool to store and manage the data. For each introduced species, information on the location of its introduction is stored, in combination with e.g. its level of invasiveness.

Q: How are global and regional checklists in Aphia cross referenced
A: Species names are only stored once in the Aphia database, regional (and thematic) checklists represent a selection of names from the global one. Aphia has an emailing alert system, where editors receive a 'weekly digest', listing all the changes and additions that have been made within their group. This includes changes and additions they have made themselves, or that were made by their fellow-editors. In addition, this weekly digest lists unchecked information, offering the editors the opportunity to easily validate small chunks of data in the database.

Remark (Geoff): The editor network is an important bottleneck for initiatives such as WoRMS, FADA, etc. as they heavily rely on expert input, which can be time-consuming. Because of this, we should try to facilitate their work as much as possible.

The Register of Antarctic Marine Species: development and status (Bruno Danis)

Presentation of the biodiversity.aq ecosystem including the Microbial Antarctic Resource System (mars.biodiversity.aq)

Q: How does MARS relate to GenBank
A: MARS links to GenBank, sequence data can be cached, but MARS not acting as repository for this type of data.

The Freshwater Animal Diversity Assessment: origin, rationale and recent developments (Koen Martens, Aaike De Wever)

Q: How far off are you from having complete lists (and how far will you get by including groups already in WoRMS)
A: Currently almost 50k species are included in the FADA database, an additional 30k are in the pipeline (following last years project on freshwater macro invertebrates.

Considering the larger groups from WoRMS we could import over 8000 additional animal species into FADA [(6579-2814) Copepods + 880 Decapoda +698 Isopoda + 1299 Mollusca-Gastropoda + 1573 Amphipoda], but would still require a substantial effort to reach a comparable level of completeness than WoRMS.

Discussion topics

Data exchange among databases and planned improvements on the databases

Geoff stressed the **importance of consulting editor groups**. While some of the editors may be overlapping, the group of editors they represent may be quite different.

Koen: Specific overlapping groups may require data harmonisation (example: joining Ostracod datasets as done by Dave Horne)

Link up with already available external datasets (e.g. as also done for WoRMS): incl. fish, birds, reptiles.

Further inclusion of data from CoL could be achieved by **using the CoL DwCA webservice** (http://www.catalogueoflife.org/DCA_Export/webservice/ or the CoL DwCA manual download: http://www.catalogueoflife.org/DCA_Export), while respecting the 3-level credits required when using CoL.

Currently, parasites are not included in the FADA database. Geoff recommended doing so, indicating that they may represent half of the biodiversity, and that such datasets are extremely useful esp.

when linked to the host (e.g. Plathyhelminthes host db). Marine parasites are included in WoRMS, but not (yet) in a systematic way.

Developments to improve data entry and validation by experts and possibilities for citation and valorisation of entered data

Ward recommends to **use the cache for providing feedback in input interface** (e.g. for checking names already in other dbs). If not done directly in the input interface, we definitely aim to produce a reporting facility on the cache that does this.

Several members of the follow-up committee appreciated the intention to reduce the work for the editors by letting them choose what system works best for them. While this does not necessarily means that the databases need to merge into one system, they stressed the fact that the data exchange should be transparent and the user can easily consult data independent from where it is physically stored. Wouter Los used the analogy to the banks to illustrate this.

Geoff brought to our attention that tools for checking data in Excel templates were produced in by Fauna Europea, as the FADA template was inspired by the Fauna Europea ones, such tools are potentially very valuable for facilitating the validation for FADA. RBINS will **explore the possibilities of existing Fauna Europea tools**.

Ward indicated that imposing a format will always face resistance/be tricky, that there is always work to be done on the files received, and that one should sometimes just take what exists (even in Word or on paper) and work with the material provided. Once the data are integrated small edits can easily be managed through an on-line editing facility. Geoff confirms that the WoRMS system works very well for him. On the other hand, the option to perform edits off-line in an Excel file (which we can use for updating the database) could be considers as an advantage by certain editors. **FADA will consult its editors to see what options they prefer and need for editing/updating their checklists.** Here we need to distinguish between bulk uploads (=mostly the first upload ever, or a major revision) and the possibilities to make little changes through an online interface. As was mentioned in the discussion, it seems inevitable to have an online edit interface at some point.

Wouter Los mentioned that semantic annotation of data could be an interesting avenue to add value to data. This could include notes field not requiring editor validation, tags and could help to document and track data and the way it is used cfr. twitter hashtags.

Geographic regions to be used (for freshwater data). Currently FADA only includes distribution information at the level of faunistic regions, options for providing distribution at a finer scale will be explored. Koen mentioned that for most groups, the consensus at the FADA workshop in Bruges was

that original point data are the best source of information and that providing country level data was the most feasible option. To accommodate the difference in country size, the TDWG region classification was proposed.

Ward suggested that the best option is to link to a geographic gazetteer and provide the data at the most detailed scale (as already the case for WoRMS; using Marine Regions (www.marineregions.org)). Such a gazetteer could be used to determine to which exclusive economic zone (EEZ) a certain place name belongs. He also suggested that indicating the difference between marine continental shelf species and species occurring in deeper zones could be useful. Apparently this will be largely covered by the "neritic flag" in WoRMS, which will be documented as part of the traits information.

In terms of citation and valorisation of data, the FADA checklist generation tool will be improved and a similar tool will be developed for WoRMS and RAMS. Such lists can be used to generate a checklist publication for e.g. ZooTaxa, EJT,… thus helping to valorise the work on the checklist and increasing the chance that the data will correctly be cited. In addition to this, one of the participants also pointed out the value of record level citations (as is the case for WoRMS). This option will need to be explored for FADA, but for now, users are requested to cite the entire checklist.

Data exchange and tools targeting regional and national initiatives: potential, possibilities for wider uptake of data products?

It is important to remember that the databases included in AquaRES have a global focus. National and regional initiatives should be made aware of this and be pointed to the right locations and contact persons. We should also make the distinction between our editor group and our user group. The first aim of AquaRES is to make the system more easy-to-use for our editors, so they can upload data more easily. The users (end-users) should primarily be pointed towards the initiatives where AquaRES will feed into, such as e.g. LifeWatch.

Geoff recommended us to inform public bodies about the available tools (example: JNCC used taxon matching tools for checking species names)

Saskia mentioned that tools and data products to support reporting for the Marine Strategy Framework Directive (MSFD) would be very valuable. These report represent an assessment of state of the marine environment incl. species lists, invasives. The last report was produced in 2012, and the next one is due in 2018. In this particular case, EMODnet could be a relevant resource as it brings together information from different systems which can be helpful for MSFD work. This highlights the need to point end users to systems where AquaRES will feed into.

Wouter Los suggested to get in touch with the Eionet (http://www.eionet.europa.eu) contact user group through a visit to Copenhagen. This would help to disseminate news on developed tools and databases to a wide group of relevant potential users.

Editors should be pointed to PESI and other regional initiatives and check how they can extract information from there, from a technical point of view. It was stressed here again that the focus of the project is on the development of tools to assist editors in data entry, and not the data entry in itself.

Data exchange and tools targeting international initiatives: streamlining exchange, towards improved data quality, uptake of developed tools, requirements from potential users.

Here we introduced our plans to automate the creation of Darwin Core Archive files that meet the needs of as much international initiatives as possible. In addition, FADA also aims to automate the data exchange to Catalogue of Life by exposing the data in their recommended format.

Wouter Los mentioned that a tool for running cross-checking of checklists is available from EUBrazilOpenBio (I guess this one at eubrazilopenbio.eu). Wouter Addink added that this is the tool that has also been used by Catalogue of Life (and the Chinese Academy of Science). We will **explore these existing cross-mapping tools**.

For the overlapping groups, we aim to contact the editor (groups) and provide them with a cross-checking list where relevant, to help us to discuss the approach in how to exchange data/give priority to certain data, etc.

Nabil suggested that it would be worthwhile to also develop record level web services for FADA. On the other hand, he mentioned that such services should be kept perfectly stable (which is even something which is hard for big players such as GBIF). In terms of identifiers, he mentioned that assigning DOIs could be a stabilising factor. **Nabil and the RBINS team will further continue the discussion on FADA web services to explore which options are feasible.**

Koen asked whether harvesting taxonomic data directly from xml-publications is already in place for WoRMS or other Global Species Directories (GSDs). We are aware of some work going on in the framework of Pro-iBiosphere and through Plazi, but not yet at the level of GSDs. WoRMS is currently looking into this.

Wouter Addink mentioned that it might be interesting to link the Marine species identification portal (http://species-identification.org/) to the AquaCache.

Other

Koen mentioned that the "North Sea 2015" meeting in September 2015 could be a potentially interesting meeting to which we could link our editor workshop. **The option to link to this meeting will be further explored.**