Distributed database system of the New Atlas of Amphibians and Reptiles in Europe: the NA2RE project

Neftalí Sillero[1,*], Marco Amaro Oliveira[2,3], Pedro Sousa[3], Fátima Sousa[3], Luís Gonçalves-Seco[1,3]

Abstract. The Societas Europaea Herpetologica (SEH) decided in 2006 through its Mapping Committee to implement the New Atlas of Amphibians and Reptiles of Europe (NA2RE: http://na2re.ismai.pt) as a chorological database system. Initially designed to be a system of distributed databases, NA2RE quickly evolved to a Spatial Data Infrastructure, a system of geographically distributed systems. Each individual system has a national focus and is implemented in an online network, accessible through standard interfaces, thus allowing for interoperable communication and sharing of spatial-temporal data amongst one another. A Web interface facilitates the access of the user to all participating data systems as if it were one single virtual integrated data-source. Upon user request, the Web interface searches all distributed data-sources for the requested data, integrating the answers in an always updated and interactive map. This infrastructure implements methods for fast actualisation of national observation records, as well as for the use of a common taxonomy and systematics. Using this approach, data duplication is avoided, national systems are maintained in their own countries, and national organisations are responsible for their own data curation and management. The database could be built with different representation levels and resolution levels of data, and filtered according to species conservation matters. We present the first prototype of NA2RE, composed of the last data compilation performed by the SEH (Sillero et al., 2014). This system is implemented using only open source software: PostgreSQL database with PostGIS extension, Geoserver, and OpenLayers.

Keywords: chorological Atlas, Europe, herpetology, Spatial Data Infrastructures, spatio-temporal data.

Introduction

Species conservation is not possible without knowledge on biology and distribution. In order to achieve this purpose, chorological databases are essential (Sillero, Celaya and Martín-Alfageme, 2005; Loureiro and Sillero, 2010). The presentation of chorological information has evolved from basic species lists (e.g. Padial and De la Riva, 2004; Padial, 2006), to complex Internet-based information visualisation methods. Species lists only provide information about the presence of species in parti-

cular areas, offering vague information about the distribution of the species (Padial and De la Riva, 2004; Padial, 2006). More recently, complete information about species chorology is offered by a large number of projects with focus on either local, regional or national levels (e.g. Sindaco et al., 2006; Salvi and Bombi, 2010; Sos et al., 2012); some compilations already cross national borders, such as the last compilation for European herps (Sillero et al., 2014), the Atlas of Western Palearctic reptiles (Sindaco and Jeremcenko, 2008), or the project Biogr (www.bio-gr.eu). Also, the Fauna Europaea project (www.faunaeur.org) provides data in an interactive online system, but includes only presence/absence data per country. Besides species and geographical area of interest, two distinctive characteristics can be found in these initiatives: how the observational data is stored (data curation) and how the resulting knowledge is transmitted or presented (scientific visualisation). Regarding data curation, little has been done: the main concern of

1 - Centro de Investigação em Ciências Geo-Espaciais (CICGE), Observatório Astronómico Prof. Manuel de Barros, Alameda do Monte da Virgem, 4430-146 Vila Nova de Gaia, Portugal
2 - INESC TEC – INESC Technology and Science (formerly INESC Porto), Rua Dr Roberto Frias 378, 4200-465 Porto, Portugal
3 - Instituto Superior da Maia (ISMAI), Av. Carlos Oliveira Campos, 4475-690 Avioso S. Pedro, Portugal
*Corresponding author;
e-mail: neftali.sillero@gmail.com

most projects is to simply store the observational data, using different ways such as ASCII text files, simple spreadsheets, or databases, but mostly with no curation concerns (one exception can be the Portuguese database: Loureiro and Sillero, 2010). A distinction also exists regarding scientific visualisation of the information: most commonly, static maps are printed in books, but there are some examples of web applications allowing dynamic visualisation of chorological data (e.g. Netherlands, Switzerland, and Spanish systems: http://telmee.nl/; http://lepus.unine.ch/; and www.herpetologica.es/programas/siare, respectively).

Although end users' attention is usually focused on the final results, giving greater importance to the species maps, databases are the most important part of a chorological project (Sillero, Celaya, and Martín-Alfageme, 2005; Loureiro and Sillero, 2010). However, these databases are frequently ignored by their managers upon fulfilling their goals. Therefore, maps are constantly out of date, whereas the database should be a lasting, permanent and upgradable product, even if it is not online. Databases should allow periodic production of map compilations (published as a book, for example) in a relatively fast and simple way, but should also become a repository and tool for subsequent studies.

Chorological databases can be prone to numerous and varied errors, mainly in species misidentification and erroneous locality data (Sillero, Celaya, and Martín-Alfageme, 2005). Errors in species identification are almost impossible to correct, mainly a posteriori, when individual vouchers or photographs are not collected. Actually, almost all records on chorological databases only include information about the species, location, date, and collector. Similarly, geographical errors are difficult to correct but this depends on how data are collected and stored. In fact, incorrect species locations are the main source of error in chorological databases (Sillero, Celaya, and Martín-Alfageme, 2005; Loureiro and Sillero, 2010).

The only way of avoiding them is to build secure and reliable systems where the introduction of any type of errors is reduced as much as possible through automatic data validation. For example, erroneous observational data locations can be eliminated if coordinates are recorded by GPS enabled devices.

Problems increase exponentially when compiling data from many sources with different owners, where formats, data, spatial resolution and geographical coordinates are disparate. This is of special importance for continental atlases (e.g. Atlas of amphibians and reptiles of Europe, Sillero et al., 2014). In these cases, the main problem is how to simplify the implementation process of an integrated representation of observational data owned and managed by third persons. This is of special relevance when the data owner is a government, as any government will not allow the control of its chorological database by another country or a foreign institution. Additionally, representing species distribution data from different owners (e.g. countries) may require the duplication of the data, by the creation of a new database compiling all the data. Therefore, duplication errors (i.e. data is collected in the original databases but not in the compilation and vice-versa) can be generated if all databases are not synchronised at the same time or they are not accessed in an interoperable way. Therefore, the best found solution is to implement a distributed (geo-referenced) system of systems, where a Web application (a system) facilitates access to the data in geographically distributed national systems, as if it were one virtual integrated database. Upon user request, the Web interface searches all distributed data-sources for the requested data, integrating the answers in a constantly updated and interactive map. This infrastructure can implement methods for fast actualisation of national observation records, as well as for the use of a common taxonomy and systematics. Using this approach, data duplication is avoided, national systems are maintained in their own countries, and national organisations are responsible for

their own data curation. The database could be built with different representation (abstraction) and resolution levels of data, and filtered following species conservation matters.

The main objective of this work is to present the first prototype of the NA2RE system (the New Atlas of Amphibians and Reptiles of Europe), composed by 30 national or personal databases (currently allocated in virtual machines distributed on three physical servers; for a complete list of contributing institutions and persons see NA2RE website and Sillero et al., 2014). The project considers the following requirements:

1. Only Free/Open Source software should be used for the implementation of the system. Hence, the system is less dependent on funding requirements, allowing its distribution to countries with different developing levels.

2. The system should be managed using geographical information technologies, in order to minimise geographical errors in records.

3. Means for easy and fast data actualisation should be implemented, remaining under the control of the database owners.

4. The taxonomy and systematics of the species should be easily modified and actualised under the control of the Societas Europaea Herpetologica (SEH).

NA2RE system

NA2RE includes data from all European countries and national territories, following political limits together with biogeographical ones (see Sillero et al., 2014). However, as explained below, this system can be applied to any region and taxon. NA2RE is, above all, an infrastructure built upon a system of systems approach known as Maier's criteria (Maier, 1998), i.e. with independence of operation and management of each element, evolutionary development, emergent behaviour (e.g. unpredictable properties may arise and influence the whole

system of systems) and geographical distribution of elements. In NA2RE, support for data source access is implemented following Open Geospatial Consortium (OGC) and ISO 19100 set of standards (www.opengeospatial. org/standards/is). Each independent system (i.e. the national systems) is implemented using only Free and Open Source Software (FOSS), for this purpose: 1) Ubuntu Server, 2) PostgreSQL database system (with PostGIS extension to support spatial data objects and operations); 3) Apache HTTP server; and 4) GeoServer Open source implementation of OGC standards for data access in order to process, edit, and share georeferenced data, and it is used as a bridge between the data and the website to allow the display of database observations.

The intrinsic nature of the system provides that each independent element may request and access data from any other element, thus interoperating among them. This characteristic is basal to the implementation of NA2RE, where the Web map interface seamlessly integrates data provided by the independent national systems. The NA2RE Web map interface participates in the overall system as yet one more element (or node) in the afore mentioned infrastructure, with similar characteristics to other independent elements, also using only FOSS: 1) Ubuntu Server; 2) Apache HTTP server; and 3) Openlayer javascript library. This web map interface will act as a medium to request and integrate data from all independent elements (in each country), to filter and present with colours the different taxa (family, genus, and species), and to represent the results in NA2RE web page (http://na2re.ismai.pt). Upon user request for a specific taxa, the Web interface searches all distributed elements for the requested data, selects the national systems with records of the requested taxa, and send back the data, integrating the answers in an interactive map (fig. 1).

Each database includes several related tables, storing taxonomical information (family, genus, species) as well as records' coordinates (fig. 2). Together with these data, a history table
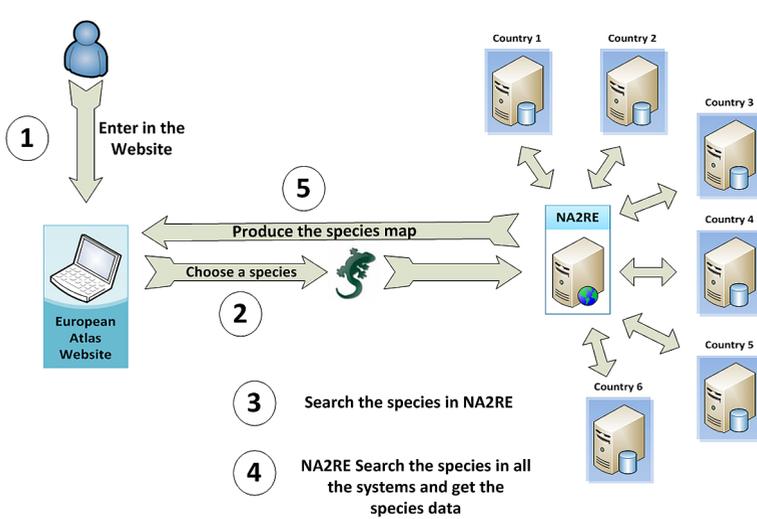
Figure 1. Flow-chart of the Spatial Data Infrastructure of the NA2RE project, the New Atlas of Amphibians and Reptiles of Europe, a system of geographically distributed systems.
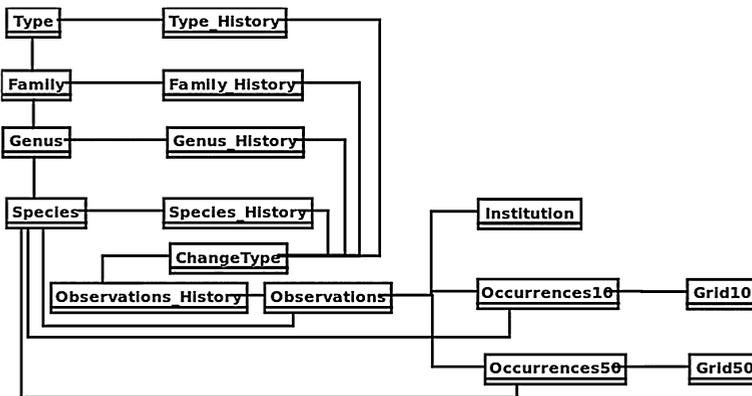

Figure 2. Abbreviated structure of the database of NA2RE. The database contains several related tables, with the objective of reducing physical space and increasing computing performance.

is linked in order to track any change in the tables. In a near future, year data should also be included. This simple database structure is very effective, reducing physical space and increasing computing performance.

Description of the NA2RE website and interactive map

The NA2RE website has a first page of presentation, where the project is briefly explained, and where the Atlas page can be accessed. This page is divided in three parts (fig. 3):

1) The left part is a table of contents and shows the list of available layers, composed by an extensible tree containing all the taxa considered in the project. The first division of the tree includes the groups of amphibians and reptiles. After this, the tree shows the different taxa by family, genus, and species categories, following a hierarchical structure. The data to be represented in the map include the chosen taxa and all the following taxa below this level on the tree. For example, if the user chooses the genus *Bufo*, all the observations of all the species belonging to this genus will be added
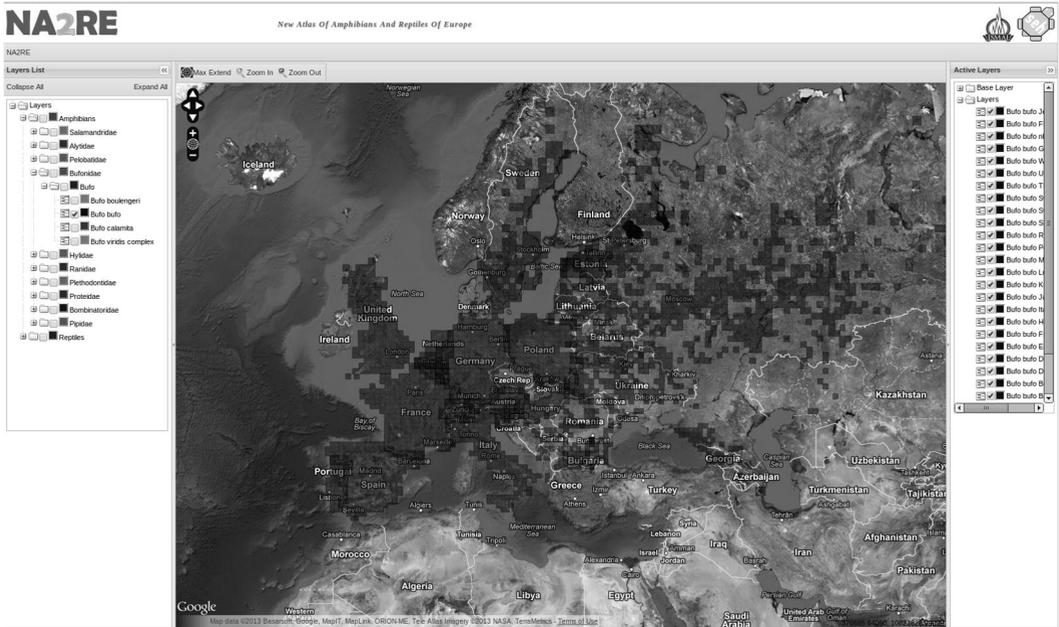
Figure 3. The web interface of NA2RE (http://na2re.ismai.pt). Upon user request, the web interface searches all distributed data-sources for the requested data, integrating the answers in an always updated and interactive map. See main text for a detailed description of the web interface.

to the map. Taxa records are represented with transparent colours, allowing layer overlapping. Therefore, it is possible to choose several taxa simultaneously, each one represented by its own layer and with the chosen colour. There are two buttons close to the taxon names: the first button is to turn on/off the taxa layer in the map and the other one to choose the representation colour (initial colours are randomly defined). The tree in the table of contents is automatically created from the family, genus, and species records and stored in the database. Therefore, the tree changes automatically when taxonomical modifications are introduced. For this purpose, the NA2RE system provides a web application that allows modifications in a fast way. Here, changes in taxon names, and addition or deletion of species are possible. All client databases of the NA2RE system should fit this taxa list in order to be introduced in the system.

2) The central part of NA2RE interface corresponds to the Atlas map. Once the user turns on the map of a specific taxon and chooses a

Figure 4. Table with taxa records for a specific 50 × 50 km UTM square of the NA2RE interactive web map. This table is opened upon user request on a specific UTM square.

colour for it, the NA2RE system connects to all the independent distributed elements, and represents the selected map. Typical zoom and pan functions are also available. The map is zoomed directly to Europe by default. Clicking on a particular cell of any grid (layers on the map), will pop-up a table showing all records represented by that square (fig. 4). For example, if the user chooses the genus *Bufo*, selecting a cell of the

resulting grid layer will show the species of the genus *Bufo* present in that cell.

3) In the right part of the interface, there is a tree including all layers presented in the map: the background layer and each national or personal database with data for the represented taxa (all databases included in NA2RE are listed in its website and in Sillero et al., 2014).

As stated above, each national system, besides being geographically distributed and operationally independent, also has managerial independence. Thus, each country or institution is able to curate its own observational data and maintain its own system. These administrative tools are not currently implemented, but they will be developed in the next version of NA2RE system. For this purpose, each national system will include a simple but secure web application which will support these managerial tasks. This web application, after passing through a login process, will provide to the administrator user a set of tools to edit the taxonomical list of species and to perform all the required tasks in order to update in a simple way the database and the interface systems. Tools and mechanisms are also planned for the automation of some tasks, presently requiring some human interference, such as for validating the results of an updated process in the list of species before setting a national system as updated. When the system is actualised, two versions of the systems will work at the same time, the actualised version and the old one. Therefore, the actualised version of the system will only work after validation and when considered as correct. In this way, the service is never interrupted.

Advantages of NA2RE

NA2RE presents several advantages: 1) no duplication of records as there is no central database joining all the records from different databases; 2) chorological databases are under complete control of their owners; 3) updated or most recent available species data are quickly accessed and visualised; 4) chorological data is organised hierarchically, allowing the representation of all the taxa below a specific level; 5) easy actualisation of species' taxonomy and systematics; and 6) several taxa can be mapped at the same time. Furthermore, this system is independent of the study area and taxa represented. It can be used easily for mapping other groups of flora and fauna in other regions of the world. However, the implementation of the system is not a simple task when compared to non-distributed systems. Although at first it seemed easier to gather all data under only one database, issues associated with data ownership, updates and duplication proved to be more relevant than the apparent complexity of the distributed system.

Further developments

As stated in the Objectives, the 30 databases incorporated to the system are allocated to three servers. Our initial intention was only to prove that the system works properly. Therefore, the next steps of NA2RE project will be to develop administrative tools and to implement each national or personal database under the owners' control in their place of origin. All NA2RE systems will follow the INSPIRE European directive. We hope that very soon, all European herpetological institutions will join NA2RE. New versions of NA2RE should allow the introduction of chorological data directly through a web application.

References

Loureiro, A., Sillero, N. (2010): Metodologia. In: Atlas dos anfíbios e répteis de Portugal, p. 66-74. Loureiro, A., Ferrand, N., Carrertero, M.A., Paulo, O., Eds, Esfera do Caos, Lisboa.

Maier, M.W. (1998): Architecting principles for systems-of-systems. Systems Engineering **1**: 267-284.

Padial, J.M. (2006): Commented distributional list of the reptiles of Mauritania (West Africa). Graellsia **2**: 159-178.

Padial, J.M., De la Riva, I. (2004): Annotated checklist of the amphibians of Mauritania (West Africa). Rev. Esp. Herpetol. **18**: 89-99.

Salvi, D., Bombi, P. (2010): Reptiles of Sardinia: updating the knowledge on their distribution. Acta Herpetologica **2**: 161-177.

Sillero, N., Celaya, L., Martín-Alfageme, S. (2005): Using GIS to make an atlas: A proposal to collect, store, map and analyse chorological data for herpetofauna. Rev. Esp. Herpetol. **19**: 87-101.

Sillero, N., Campos, J., Bonardi, A., Corti, C., Creemers, R., Crochet, P.A., Crnobrnja Isailović, J., Denoël, D., Ficetola, G.F., Gonçalves, G., Kuzmin, S., Lymberakis, P., de Pous, P., Rodríguez, A., Sindaco, R., Speybroeck, J., Toxopeus, B., Vieites, D.R., Vences, M. (2014): Updated distribution and biogeography of amphibians and reptiles of Europe based on a compilation of countrywide mapping studies. Amphibia-Reptilia **35**: 1-31.

Sindaco, R., Doria, G., Razzetti, E., Bernini, F., Eds (2006): Atlante degli Anfibi e dei Rettili d'Italia/Atlas of Italian Amphibians and Reptiles. Societas Herpetologica Italica. Edizioni Polistampa, Firenze, Italy.

Sindaco, R., Jeremcenko, V.K. (2008): The Reptiles of the Western Palearctic. Belvedere, Latina, Italy.

Sos, T., Kecskés, A., Hegyeli, Z., Marosi, B. (2012): New data on the distribution of *Darevskia pontica* (Lantz and Cyrén, 1919) (Reptilia: Lacertidae) in Romania: filling a significant gap. Acta Herpetologica **7**: 175-180.