11. Going dynamic: a vision for the Biogeographic Atlas of the Southern Ocean (BASO)

Bruno Danis[1], Claude De Broyer[2], Philippe Koubbi[3], Anton P. Van de Putte[2]

[1] Marine Biology Laboratory, Université Libre de Bruxelles, Brussels, Belgium
[2] Royal Belgian Institute of Natural Sciences, Brussels, Belgium
[3] Sorbonne Universités. UMR BOREA 7208, Université Pierre et Marie Curie, Muséum national d'Histoire naturelle, Paris, France

1. Context

Biogeographic information is of primary importance for monitoring biodiversity and detecting impacts of environmental changes, developing the ecoregionalisation of the Southern Ocean, designing marine protected areas, conducting comparative biogeographic analyses, and discovering marine biodiversity hotspots. Stressed by fast-paced change, and strong environmental pressure, the importance of having high quality data openly available online has become part of the Antarctic researchers culture. As a result, more and more biodiversity primary data is made available through dedicated information systems, which take the form of web or data portals, offering free and open access to highly needed information.

The multi-authored «Biogeographic Atlas of the Southern Ocean» (BASO), builds upon information networks and their collaborative communities, to provide an up-to-date synthesis of the Antarctic biogeographic knowledge. The first version of the BASO is taking the form of a «classic» printout version. To take full advantage of the existing informatics tools, in particular biodiversity informatics and web tools, a dynamic approach has been envisaged from the beginning of the project to follow-up on the initial, static version. This dynamic version will give birth to a new interoperable system, building customized biogeographic synthesis pages on-the-fly, using the content of relevant information systems as it improves over time.

It is important to stress that only by ensuring community-driven developments, scientific independence, and high levels of data interoperability will this initiative meet the success it deserves and involve a wide range of stakeholders.

2. Overall scope of BASO

Adopting the same approach, the dynamic, digital version will be made available through a dedicated portal. This dynamic BASO will build its content from a blend of authoritative sources, including of course data assembled in the framework of the static version, but also taking advantage of a dynamic approach, allowing re-generation of maps including the latest available taxonomic and biogeographic data from various sources. The dynamic BASO could also include advanced functionalities, such as the capacity to build tailor-made maps, the possibility to tweak the modelisation parameters, the inclusion of data available from different databases (Register of Antarctic Marine Species (RAMS), SCAR-MarBIN, ANTABIF, Antarctic Field Guides, GenBank,…), allowing the end-users to provide feedback on data or models, eventually turning the BASO into a community-driven platform entailed to publish and share maps generated using the most accurate and up-to-date information.

Hence, the dynamic version will complete and expand the hard copy version, and offer in addition – using appropriate technical solution such as webservices and APIs (Application Programing Interfaces) - visualisation tools as well as environmental/model datasets in appropriate formats directly useful for Geographical Information Systems (GIS) users. This approach would be particularly useful in the framework of habitat modelling and prediction of potential climate-related changes. Habitat modelling allows predicting species potential distribution according to the species relation to a set of environmental variables. The species modelling is not possible for all of the considered species, but the identification of areas of major biodiversity change is nevertheless achievable. Two workshops identified in 2010 and 2011 the methodology that will be used for the Atlas. (see Koubbi *et al.* 2010, 2011; Loots *et al.* 2007; Pinkerton *et al.* 2010).

In designing the dynamic version of BASO, special care should be devoted to ease the access to data, documenting the content (using metadata from the Global Change Master Directory, SERF format, or GBIF Ecological Markup Language (EML)) and using widely accepted Geospatial standards (Open Geospatial Consortium (OGC)) to ensure interoperability and direct usability of the data in as many contexts as possible.

To the best extent feasible, emerging datatypes (tracking, genetics, microbial, ROV imagery,...) should be anticipated in the design of the data and workflows for the BASO.

3. Stakeholders

A dynamic version of the BASO will probably attract a wide variety of users from various horizons, but will also be useful for connected initiatives which will integrate the relevant data and workflows generated in the framework of the BASO, in a multidirectional fashion.

Regarding the users, there are high chances that the primary end-users for the BASO data portal will be the Antarctic researchers (biogeographers, taxonomists, ecologists, molecular biologists, conservationists, modellers,…).

However, BASO typically represents a cross-disciplinary exercise. This is especially relevant in the context of the new SCAR research programmes, which have been specifically designed to promote cross-linked approaches. Other stakeholders include conservationists, policy makers (Antarctic Treaty System: Committee on Environmental Protection (CEP), Commission for the Conservation of the Antarctic Marine Living Resources (CCAMLR),…), as well as students, the education & outreach community or even the general public. Networked initiatives such as ANTABIF, OBIS, GBIF, SOOS, the Antarctic Environments Portal will also be incorporated in the multidirectional data flows.

4. Requirements

Even if it is still early to anticipate all the fine-scale requirements and functionalities of the dynamic version of BASO at this stage, a set of general considerations can nevertheless be proposed. BASO should be considered as a tool to facilitate discovery, generation and access to both factual and predictive biogeographic data, including the possibility to offer access to additional sources of information at various levels of biological complexity.

As a development platform, BASO will be open to both data end-users and data publishers. Future managers of the BASO are expected to be aware of the needs of data end-users and the wishes and needs of data publishers. To the best extent possible, the wishes of the relevant communities regarding BASO functionalities should be expressed in the flexibility and scalability of the tool itself. Based on an iterative developmental process, feedback can be expected from the users community in order to match their needs and the functionalities of the BASO.
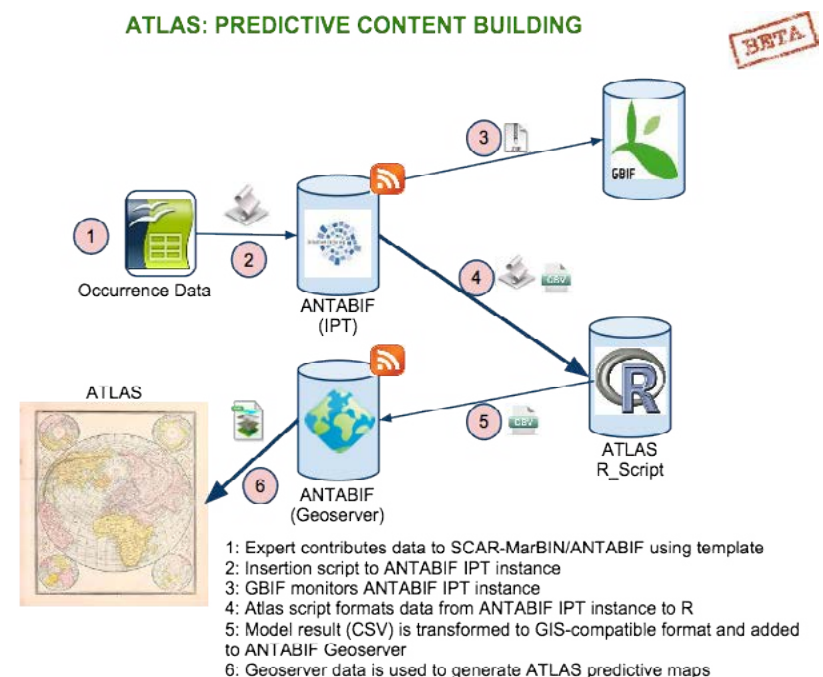

Figure 1 Wireframe describing the general concept for the Dynamic version of the Biogeographic Atlas content building, and associated workflows

A proposed general workflow to generate new predictive maps is described in Fig 1. New occurrence data is published and documented by data providers through the ANTABIF Integrated Publishing Toolkit (IPT). New data is exposed to ANTABIF, GBIF and OBIS for harvesting, and also to the BASO R server, which processes the data and incorporates it in newly generated predictive maps. The maps are published using the ANTABIF geoserver, and made available and searchable through the BASO webportal.

A set of basic requirements, which should be met during the development of the dynamic version of the BASO are described below:

Data consistency between BASO instances and ANTABIF. The BASO will include the possibility to acquire contextual data from the ANTABIF central index. The process of synchronizing data between BASO and ANTABIF should be more or less automated, depending of the type of data to be synchronised:

1. occurrence data - format Darwin Core (DwC) table, all data kept by ANTABIF, extending to all fields available, as defined on the relevant GBIF resource webpage.
2. species lists - format DwC table for species lists extension, all data kept by ANTABIF, not only the indexed data.

3. metadata - in GBIF EML and GCMD DIF formats, again not only indexed data but resource-level metadata.
4. maps - generated from the BASO R-server are posted on the ANTABIF geoserver using OGC-compliant standards.
5. multimedia metadata - according to the MRTG schema for the multimedia metadata format, all data, not only the index.

Free & Open source solutions. To the best extent possible, the BASO system should be based on layers of open source solutions such as PHP, MySQL, Apache. Open source application development frameworks or libraries such as JQuery (javascript), CakePHP, that streamline and simplify both code development and updates should also be used. Web standards (HTML5, CSS3, JQuery) should also be considered. Easy maintenance of the application is an important criteria.

Compliant with the ANTABIF Informatics Infrastructure / Architecture. To ensure the interoperability of the BASO with other ANTABIF tools such as IPT, AFG,..., BASO should be developed in a manner that is compliant with ANTABIF informatics infrastructure and architecture. The interoperability of the BASO with other ANTABIF informatics tools should be aiming at optimizing current development, and ensuring the same standards are adopted and that resulting data flows are kept simple and consistent (i.e. not affecting the data quality of the data system as a whole).

Robust. The BASO system should be thoroughly tested to ensure robustness on two fronts: (1) the BASO should be able to cope (without crashing) with a wide range of stress situations, and it should perform reasonably under important workload (queries and data discovery). User experience should be as intuitive and friendly as possible; any errors should be trapped and not seen by users, but systematically addressed as priority requirements or issues; (2) it should also allow the management of dependencies on external services (for example, external data calls).

Scalable. Scalability is an essential feature, as the BASO should be able to cope with rapidly increasing data masses at least equivalent to that available from the current ANTABIF content, in a fashion that does not impact the end-user's experience (delays). The BASO should be adaptive, in order to allow the inclusion of major data loads in the future.

Modular. The BASO should be built in a modular fashion, which ensures that component functionalities can be adjusted, added, improved, removed or exchanged without impacting the application as a whole.

Extensible. The BASO development and implementation should follow design principles that will enable future growth, both in terms of the addition of new functionality or through modification of existing functionalities. In addition, BASO should provide the possibility to consume services and functions that are created elsewhere. Examples might include taxonomic or biogeographic data sources, mapping functions or web map/feature services.

Intuitive and friendly end-user interface. The interface for end-users should be intuitive and easy to use, without requiring extensive technical assistance, support or off line user guides.

Documentation and code maintenance. A clear and standard practice for the maintenance and documentation of code should be set in place, agreed and communicated as a best practice.

Open collaborative development. The development of the application should be consistent with open source, allowing the core code and any modules developed to be improved over time.

Community-driven. By including opportunities for user feedback, the system should evolve by iterations and respond to user needs, with a focus on the SCAR users community. User interactions can also happen by using stored queries. The implication of the community is a core requirement for the BASO, and the objective is to ensure that the tool has enough buy-in for it to be allowed to evolve in the future.

Discoverable through search engines. The BASO should have appropriate html metadata tags that allow the tool itself to be discovered through search engines. Metadata made available to search engines should aim primarily at making the BASO data contents easily discoverable through e.g. Google.

Availability of APIs and Web Services. The design and creation of Application Programming Interfaces (APIs) that expose BASO functions should be explored to allow third parties to use BASO functionalities. Web services should also be explored.

6. Perspectives

The Biogeographic Atlas of the Southern Ocean (BASO) is an unprecedented effort to reconcile vast amounts of knowledge on a single interface. This interface is first envisioned as a hard copy, with the potential to become a dynamic, online information system.

Built in a modular fashion, the dynamic BASO should be able to cope with fast-evolving data flows, biodiversity informatics technologies and user needs. Constantly up-to-date, the system could not only play an important role for research purposes (representing a powerful integration tool) but also for conservationists or policy makers. Emerging initiatives, such as the Antarctic Environments Portal (www.environments.aq), or the Southern Ocean Observing System (www.soos.aq), will immediately benefit from the modelling power of the BASO workflows.

In addition, these workflows can be further shared with other integrative infrastructures, such as the Lifewatch initiative (www.lifewatch.eu).

A hundred and forty contributors have worked on the static version of the project so far, and will be directly or indirectly associated to the dynamic version of the BASO. There are high hopes that this interest and commitment will grow over time, and that other communities will join the project, and further broaden its scope. During IPY, the polar data science community was humbled by how the rapid and dramatic changes in the polar regions increased the expectations for an efficient access to rigorous, scrupulously documented data. A next priority for the BASO community is to identify mechanisms for collaborative international funding that supports the necessarily global aspects of such interdisciplinary projects. The BASO addresses deeply interdisciplinary problems. As such, as a key component of the polar scientific data ecosystem, it requires continued organic adaptation by technology, people, organizations, and society.

The latest information about the developments of the dynamic version of BASO can be found on the following url: http://atlas.biodiversity.aq

References

Koubbi, P., Moteki, M., Duhamel, G., Goarant, A., Hulley, P.-A., O'Driscoll, R., Ishimaru, T., *et al.*, 2011. Ecoregionalization of myctophid fish in the Indian sector of the Southern Ocean: Results from generalized dissimilarity models. *Deep Sea Research Part II: Topical Studies in Oceanography*, **58(1-2)**, 170–180. doi:10.1016/j.dsr2.2010.09.007

Koubbi, P., Ozouf-Costaz, C., Goarant, A., Moteki, M., Hulley, P.-A., Causse, R., Dettaï, A., *et al.*, 2010. Estimating the biodiversity of the East Antarctic shelf and oceanic zone for ecoregionalisation: Example of the ichthyofauna of the CEAMARC (Collaborative East Antarctic Marine Census) CAML surveys. *Polar Science*, **4(2),** 18. doi:10.1016/j.polar.2010.04.012

Loots, C., Koubbi, P., Duhamel, G., 2007. Habitat modelling of *Electrona antarctica* (Myctophidae, Pisces) in Kerguelen by generalized additive models and geographic information systems. *Polar Biology*, **30**, 951–959.

Pinkerton, M.H., Smith, A.N.H., Raymond, B., Hosie, G.W., Sharp, B., Leathwick, J.R., Bradford-Grieve, J.M., 2010. Spatial and seasonal distribution of adult *Oithona similis* in the Southern Ocean: Predictions using boosted regression trees. *Deep Sea Research I*, **57(4)**, 469–485. doi:10.1016/j.dsr.2009.12.010