2.1. Data and mapping

Anton P. Van de Putte[1], Huw J. Griffiths[2], Ben Raymond[3][4][5] & Bruno Danis[6]

[1] Royal Belgian Institute of Natural Sciences, Operational Directorate Natural Environment, Brussels, Belgium
[2] British Antarctic Survey, Cambridge, United Kingdom
[3] Australian Antarctic Division, Department of the Environment, Kingston, Australia
[4] Antarctic Climate & Ecosystems Cooperative Research Centre, University of Tasmania, Hobart, Australia
[5] Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Australia
[6] Marine Biology Laboratory, Université Libre de Bruxelles, Brussels, Belgium

1. Introduction

The data used in this Atlas were primarily drawn from SCAR-MarBIN (Scientific Committee on Antarctic Research—Marine Biodiversity Information Network, www.scarmarbin.be) and its successor project ANTABIF (the Antarctic Biodiversity Information Facility, www.biodiversity.aq). These two initiatives, both official SCAR products, are responsible for managing Antarctic biodiversity data, including that from the ambitious Census of Antarctic Marine Life programme (CAML; Danis *et al*. 2013), linked to the International Polar Year 2007-09. The complete expert-validated database, including records from beyond the SCAR-MarBIN primary area of interest up to 40°S (Map 1), represents 1.07 million occurrence records for 9064 validated species from ~434,000 distinct sampling stations. For more information on the data coverage see Chapter 2.2.
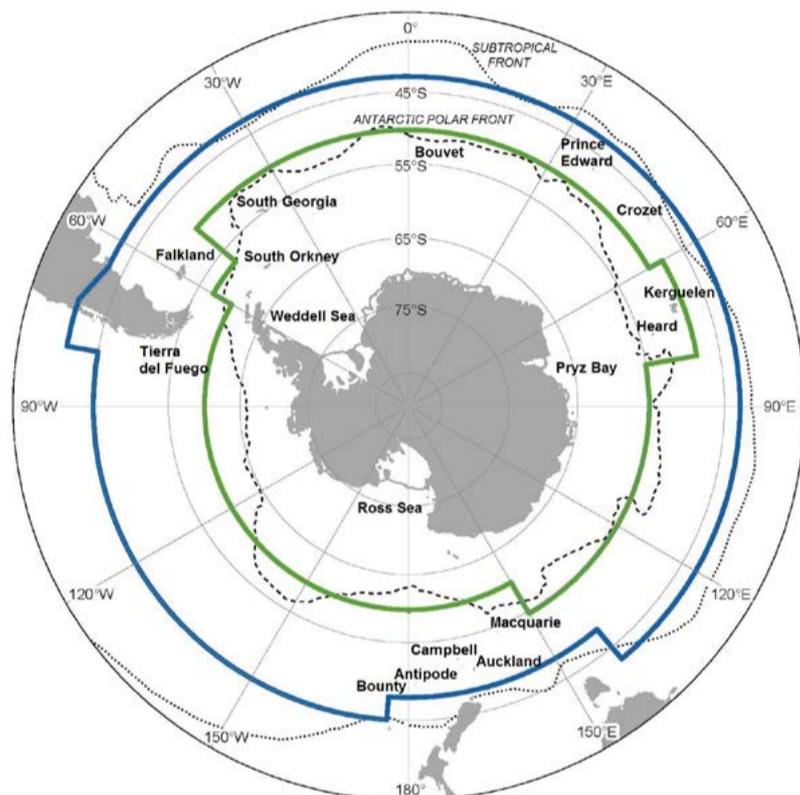
Data and mapping Map 1 SCAR-MarBIN / RAMS area of interest. The most inner line close to the Antarctic Polar Front indicates the operational (average) northern limit of the Antarctic primary area of interest. The outer line close to the sub-Tropical Front indicates the operational (average) northern limit of the sub-Antarctic area of interst (see De Broyer & Danis 2011: Table 1 for details).

2. History

The origins of biodiversity.aq can be traced back to the CAML, a 5-year project which aimed at assessing the nature, distribution and abundance of the Southern Ocean biodiversity (see Chapter 1.3). During the International Polar Year (IPY) 2007/09 CAML focused the attention of the public on the ice-bound oceans of Antarctica. SCAR-MarBIN was created as a companion-project of CAML, in order to manage its biodiversity data and make this openly accessible to all (Danis *et al*. 2013). During the course of IPY and CAML, international expeditions collected an unprecedented mass of information on the biogeography of Antarctic marine species, including numerous species new to science (e.g. Brandt *et al*. 2007).

SCAR-MarBIN was funded by the Belgian Science Policy office (BELSPO) until September 2009. It was succeeded by another BELSPO-funded project, the Antarctic Biodiversity Information Facility (ANTABIF, www.biodiversity.aq). Whereas SCAR-MarBIN focused on marine life in Antarctica, the new biodiversity.aq portal builds on the success of SCAR-MarBIN and ensures its continuation by creating an overarching network that provides access to data from both the marine and the terrestrial realms. Data made available through SCAR-MarBIN or biodiversity.aq are freely and openly available, complying with SCAR data policies. This honors the idea within the Antarctic scientific community that primary biodiversity data should be made publicly available as soon as possible after it has been collected, in the spirit of ar-

ticle III. 1c of the Antarctic Treaty (which states that *"Scientific observations and results from Antarctica shall be exchanged and made freely available."*). Following the model of SCAR-MarBIN, biodiversity.aq aggregates Antarctic biodiversity data from various providers, such as the German PANGAEA and the Australian Antarctic Data Centre, and redistributes these data to larger projects such as the Ocean Biogeographic Information System (OBIS) and the Global Biodiversity Information Facility (GBIF). Biodiversity.aq also offers a data hosting and publishing service to nations or research institutes that lack such facilities (ipt.biodiversity.aq). All publicly available biodiversity data can be searched and retrieved through this data platform.

3. The biodiversity.aq ecosystem

Biodiversity.aq is a modular data ecosystem that integrates different biodiversity tools including the Biogeographic Atlas of the Southern Ocean. The main gateway is www.biodiversity.aq, and more specialised functions are hosted under various subdomains: IPT, Data, Atlas, AFG and mARS (Fig. 1).

The first subdomain ipt.biodiversity.aq is dedicated to the integrated publishing toolbox (IPT) developed by GBIF for publication of primary biodiversity data and metadata. This important new feature allows users to standardise, document and share their primary biodiversity data using a simple web interface. It is a one-stop shop for opening up data, which can be crucial in the face of the extremely rapid environmental change that is currently challenging Antarctic ecosystems.

Data that is published through the IPT can be discovered through global biodiversity portals such as GBIF or OBIS, enhancing the visibility of the research. However there is also a dedicated Antarctic biodiversity data portal: data.biodiversity.aq. This portal builds onto SCAR-MarBIN, which offers access to marine biodiversity data for the Southern Ocean, by extending the data access to include terrestrial biodiversity information as well.

Data.biodiversity.aq offers an intuitive search engine that allows users to explore Antarctic occurrence data, taxonomic data and metadata. Occurrence data can be explored through the online mapping system, which in the future will also include various environmental layers. Data can be downloaded free of charge in common data formats such as .xls and .txt and also as data layers for Geographical Information Systems (GIS).
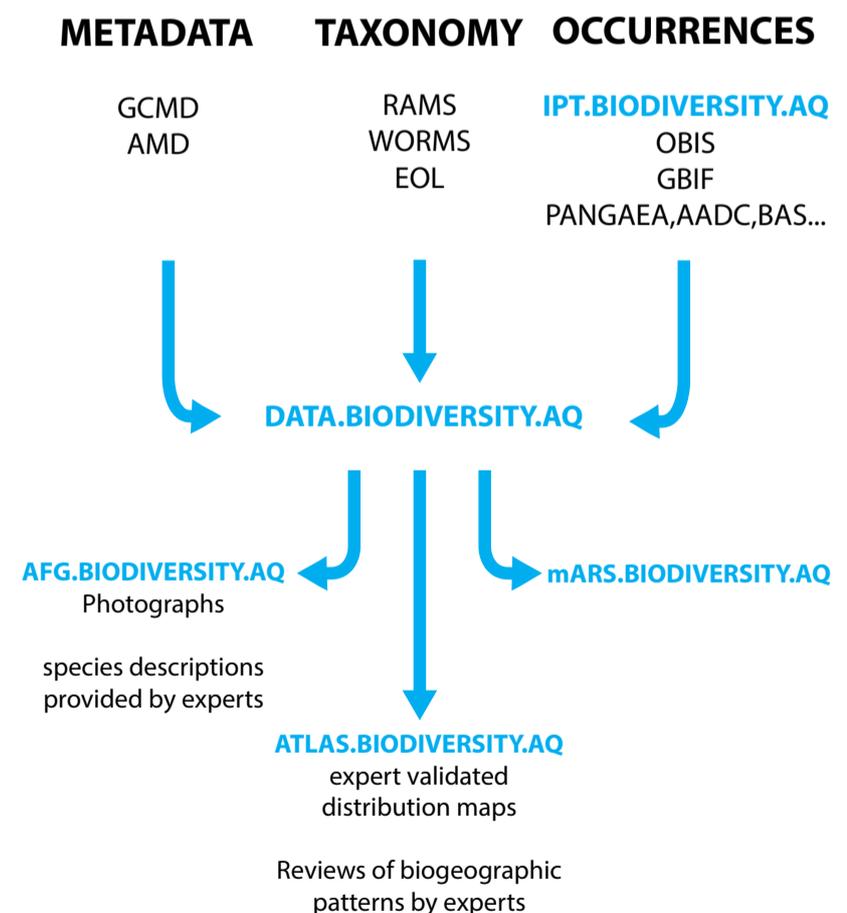

Figure 1 The Biodiversity.aq data ecosystem

Other tools for the exploration of Antarctic biodiversity information are the Antarctic Field Guides, afg.biodiversity.aq. This portal aggregates information from various resources in order to provide a digital tool for the identification of Antarctic species. The field guides present information on taxonomy and distribution in combination with high quality photographs and descriptions of (morphological) features that will aid in the identification of specimens in the field. Users can compile their own personal field guide based on taxonomic or geographic areas of interest. Personalised field guides can be downloaded as PDF files or shared online with other users.

The Microbial Antarctic Resources System (mARS) is a project under development that is envisioned as an information system dedicated to facilitate the discovery, access and analysis of geo-referenced molecular Antarctic microbial diversity data. The scope of diversity will encompass all free-living and host-associated viruses, bacteria, archaea, and singled-celled eukaryotes. mARS focuses on past, present and future works. It is an open, community-driven platform that will allow scientists to publish, document, analyse and share their data and metadata with the broad community for science, conservation and management purposes, in the spirit of the Antarctic Treaty.

4. Datatypes

The main focus of SCAR-MarBIN and biodiversity.aq is primary biodiversity data, which constitutes "digital text or multimedia data records detailing facts about the occurrence of an organism." Secondary data constitutes a wide range of ecological information on sampling sites and organisms found there (GBIF 2009).

Broadly there are three types of primary biodiversity data: occurrence data, taxonomic data, and metadata (Fig. 1). Occurrence data provide information on the times and locations at which an organism has been observed. Taxonomic data offer information on the identity of organisms. Metadata is descriptive information that provides the data user with additional details about the data such as collection procedures, and thus a means of verifying the appropriateness of the dataset for the desired usage. These three main data types were the focus for the Atlas.

The metadata of datasets published through biodiversity.aq are made available through the IPT, but also the Antarctic Master Directory (AMD), which is a component of NASA's Global Change Master Directory metadata system (GCMD). The AMD provides metadata on Earth science data sets pertaining to Antarctica and the Southern Ocean. It was crested through a collaboration between the Standing Committee on Antarctic Data Management (SCADM), and NASA's GCMD staff and acts as a resource for discovering and describing internationally held Antarctic data.

The Register of Antarctic Marine Species (RAMS; De Broyer & Danis 2011) covers the taxonomic component of the Atlas. RAMS was compiled and published thanks to a collective effort of a board of 64 taxonomic editors. It built on earlier work by Clarke & Johnston (2003), and provides an accurate list of more 8300 valid species with an up-to-date systematic classification comprising more than 18,250 taxon names (Fig. 2). RAMS is the taxonomic backbone of biodiversity.aq and also feeds in to global biodiversity initiatives such as the World Register of Marine Species (WORMS), OBIS and GBIF.
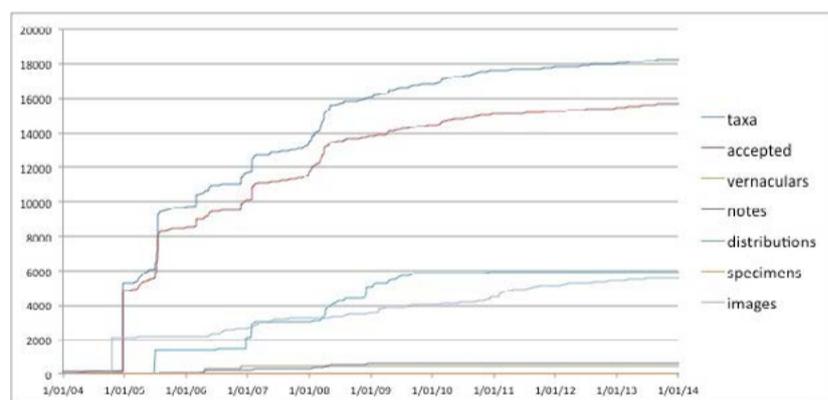

Figure 2. Number of records available through the Register of Antarctic Marine Species.

The occurrence data used in the Atlas generally fall into one of three categories: (i) validated occurrence records of Antarctic marine species that were previously available via SCAR-MarBIN, (ii) newly-provided records that were obtained for the Atlas project from other sources, and which have been made publicly available through biodiversity.aq, or (iii) data from other sources which have not been incorporated into biodiversity.aq. Data in the latter category include records provided by the Commission for the Conservation of Antarctic Marine Living Resources (CCAMLR), the International Whaling Commission (IWC), and PANGAEA (pangaea.de).

Newly provided metadata and data was submitted by the expert contributors to biodiversity.aq using a template that followed the Darwin Core 2.0 data standards, and also incorporated additional fields from the OBIS data system. This approach ensures long-term interoperability with global information systems, including OBIS, GBIF and the AMD.

5. Data cleaning

The data cleaning process relied on input of the experts of each chapter. All changes that were made by the data managers during the data cleaning process were confirmed by the principal authors of the respective chapters.

A common first step in data cleaning was to check the data-type constraints for all variables. In the provided template, each variable was associated with a particular data type (e.g. text for species names, or numeric values for latitudes and longitudes). Where possible, incorrect data types were transformed to the correct ones. The taxonomic information of the records was validated using RAMS through the taxon match tool on the WoRMS website. For samples lacking geographic coordinates but possessing verbatim localities, approximate localities were determined using SCAR's and other gazetteers. In cases where records were georeferenced using degrees minutes seconds or degrees decimal seconds, these values were converted to decimal degrees. Plots of the data points were used to identify obvious outliers (e.g. points over land, or in the wrong hemisphere) and errors due to swapping of latitude and longitude values, or in the sign of decimal degrees.

6. Mapping

Maps were produced in ArcGIS or Matlab using a standardised template. The maps have a south pole centred stereographic projection with a central meridian of 0° of longitude. The most northerly extent of the maps is generally 35° south and includes the southern parts of South America, Australia and New Zealand. The Antarctic coastline data used in the maps was sourced from the SCAR Antarctic Digital Database. Average frontal positions were taken from Sokolov & Rintoul (2009a, b), for more information see Chapter 4: Environmental Setting.
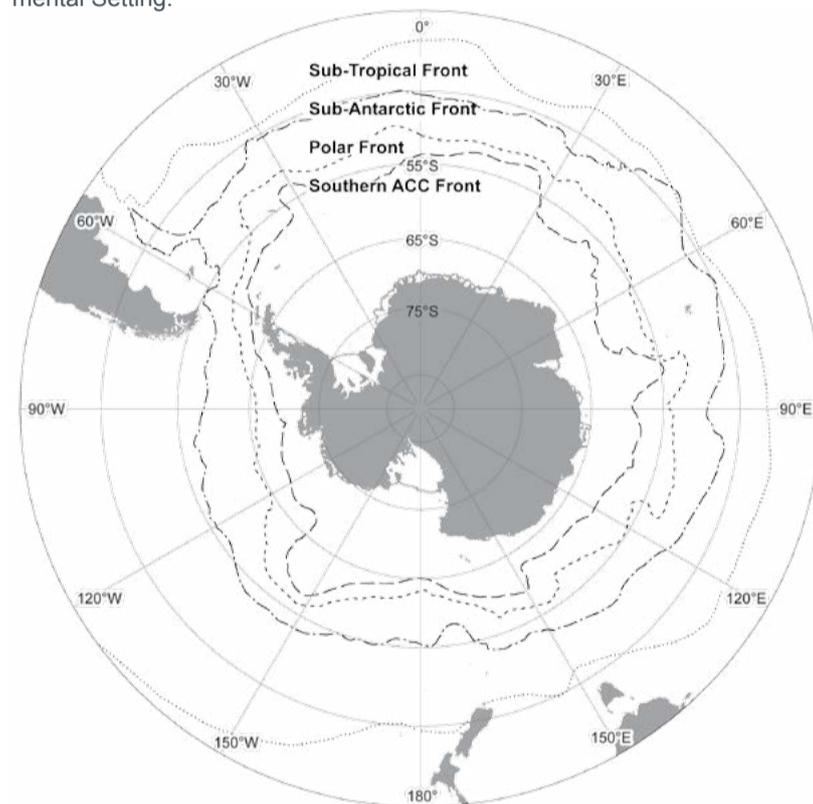
Data Map 2 Fronts in the Southern Ocean. The long-term sea surface temperatures between the fronts are as follows, from south to north (approximate ranges, see details in Chapter 4): Between the continent and the Southern ACC Front (SACCF): -1.6 to +3°C; Between the SACCF and the Polar Front: - 0.3 to +6°C; Between the Polar Front and the Sub-Antarctic Front: + 1.2° to 12°C; Between the Sub-Antarctic Front and the (extreme northern limit) of the Sub-Tropical Front: 5.7 to 24°C

References

Brandt A., Gooday A.J., Brandão S.N., Brix S., Brökeland W., Cedhagen T., Choudhury M. *et al.*, 2007. First insights into the biodiversity and biogeography of the Southern Ocean deep sea. *Nature*, **447**, 307–311. doi:10.1038/nature05827

Clarke A., Johnston N.M., 2003. . Antarctic Marine Benthic Diversity. *Oceanography and Marine Biology Annual Review* **41**, 47–114.

Danis B., Van de Putte A.P., Renaudier S., Griffiths H.J., 2013. Connecting biodiversity data during the IPY: the path towards e-polar science. In: Verde C., di Prisco G. (eds.). *Adaptation and Evolution in Marine Environments*, **2**, 21-32.

De Broyer C., Danis B., 2011. How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. *Deep-Sea Research II*: *Topical Studies in Oceanography* **58**, 5-17. doi:10.1016/j.dsr2.2010.10.007 http://dx.doi.org/10.1016/

GBIF, 2009. GBIF Work Programme 2009-2010. Copenhagen: Global Biodiversity Information Facility, 59pp. Accessible at http://imsgbif.gbif.org/CMS_NEW/get_file.php?FILE=976c6c301 5d7d2505d5ac1d45357c8

Sokolov, S., Rintoul S.R., 2009. The circumpolar structure and distribution of the Antarctic Circumpolar fronts: 1. Mean circumpolar Paths. *Journal of Geophysical Research*, **114**, C11018, doi:10.1029/2008JC005108.

Sokolov, S., S. R. Rintoul. 2009b. The Circumpolar structure and distribution of the Antarctic Circumpolar Fronts: 2. Variability and Relationship to Sea Surface Height. *Journal of Geophysical Research*, **114**:C11019, doi:10.1029/2008JC005248.