



AntaBIF.BE  
BIODIVERSITY.AQ

# AntaBIF - mid-term report

June 2011

## Background

The AntaBIF (Antarctic Biodiversity Information Facility), funded by the Belgian Science Policy Office ([www.belspo.be](http://www.belspo.be)), is building a dedicated Antarctic biodiversity data portal giving access to a distributed network of contributing database, according to the principles of the Global Biodiversity Information Facility ([www.gbif.org](http://www.gbif.org)), in the framework of the International Year of Biodiversity 2010.

This report is to be considered as the mid-term report for the project and as the second issue of the new tri-monthly reports, intended to give access to the latest news about AntaBIF, including the latest datasets published, advancements of the websites, community activities, communication, and ongoing developments. These reports will be sent in priority to the members of the AntaBIF [ISC](#), and will be posted in the [shared repository](#).

We are inspired by the Antarctic Treaty of 1959, and consider that Antarctic biodiversity information should be available freely and openly in a collaborative spirit. Efforts are directed towards changing the attitude of the scientific community towards making the data discoverable, open, linked, useful, and safe.

## Antarctic Biodiversity

Polar regions experience faster rates of climate change than any other region in the world, as noted in 2007 by the Intergovernmental Panel on Climate Change (IPCC) and subsequent reports. Because of the growing realization of the potential impact of global environmental change on the Antarctic ecosystems and the services they provide, it is crucial to establish robust and comprehensive baseline information on the Antarctic biodiversity as a sound benchmark against which future change can reliably be assessed.

SCAR-MarBIN and AntaBIF answer the urgent need to provide access to scientifically sound information on the current state of Antarctic communities and their diversity.

SCAR-MarBIN ([www.scarmarbin.be](http://www.scarmarbin.be)), started in 2005 to build the first comprehensive assessment of Antarctic Marine Biodiversity, to better understand the actual diversity and status of Antarctic marine life. These activities took place in the context of the International Polar Year (IPY), and of the Census of Antarctic Marine Life (CAML). SCAR-MarBIN was funded by the Belgian Science Policy office (BELSPO) until September 2009, and its core activities are now supported through a partnership schema including Australia, Germany, The Netherlands, and the ArcOD consortium. Further networking activities are developing thanks to a new BELPO-funded project, the Antarctic Biodiversity Information Facility (AntaBIF, [www.biodiversity.aq](http://www.biodiversity.aq)), which expands SCAR-MarBIN beyond the marine realm, by providing access to all Antarctic Biodiversity information, through a single webportal, through to a tight collaboration with the Australian Antarctic Data Center (AADC).

## Latest News

The AntaBIF team has just been expanded, and is welcoming a new Science Officer, Dr Anton Vandeputte. As of mid-June 2011, Anton will be in charge of Networking and Data Management-related tasks, improving the content of the Antarctic Biodiversity databases.

## Websites

A series of website are online, giving the AntaBIF project some webpresence, while the data portal is being developed. The main website is available on two urls: [www.biodiversity.aq](http://www.biodiversity.aq) and [www.AntaBIF.be](http://www.AntaBIF.be).

<http://www.biodiversity.aq> or <http://www.AntaBIF.be> : 510 visits since March 2011, 66% of new visits, shopwindow, semi-static website, providing basic information on AntaBIF, as well as a dynamic link to the AntaBIF blog. Accesspoint to the latest news and resources available from the network.

<http://share.biodiversity.aq> : 97 visits since April 2011, filesharing webspace (downloaded bandwidth: 7,25 GB), offers free and open access to original datasets, documents, R-code, GIS layers and other relevant resources.

<http://afg.biodiversity.aq>: 3,795 visits since September 2010, 75% of new visits, interactive Antarctic Field Guides (see below, "Side Projects" sections for further details).

[gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=AntaBIF&MetadataType=0](http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=AntaBIF&MetadataType=0): AntaBIF metadata portal, hosted by NASA. Immediately interoperable with other contexts using the Global Change Master Directory (AAD, OBIS, EBA, IPY,...). Will constitute the metadata backbone of the AntaBIF ecosystem.

<http://scratchpads.biodiversity.aq>: AntaBIF Scratchpad website (social networking application that enable communities of researchers to manage, share and publish taxonomic data online). (in test, not monitored yet)

<http://ogc.biodiversity.aq>: AntaBIF geoserver, allows to serve geospatial layers in many formats in a dynamic way. The geoserver will give access to the occurrence data as well as environmental layers such as those produced in the framework of the Biodiversity Atlas of the Southern Ocean (see below, "Side Projects" for further details). (in test, not monitored yet)

<http://www.princesselisabeth.be>: reserved domain name for a portal dedicated for the Princess Elisabeth context of the AntaBIF databases. Development Framework ([Mahara](#)) in test.

## Networking

Since its inception, SCAR-MarBIN has been seeking to federate as many partners as possible, at the national and international levels, and to mobilize the energy and competences of all the potential players. This approach is taken further by the AntaBIF project. SCAR-MarBIN has built its networks on trust and commitments from its partners and community. The data and information published through the different web platforms, as well as the technology used are now part of the public domain, in the spirit of the Antarctic Treaty (Art. III.1.c). AntaBIF follows this path.

SCAR-MarBIN and AntaBIF are deeply involved in many networks which share similar objectives and long-term vision on raw data access and publication, and adopting a collaborative approach to improve Antarctic science and to help conservationists and environmental managers. A few examples of these open networks include:

OBIS: **Ocean Biogeographic Information System**, ([www.iobis.org](http://www.iobis.org)). SCAR-MarBIN is the Antarctic Node of OBIS. OBIS is the information component of the Census of Marine Life.

GBIF: **Global Biodiversity Information Facility**, ([www.gbif.org](http://www.gbif.org)). AntaBIF and SCAR-MarBIN compose the Antarctic Node of GBIF, publishing all available data. SCAR is the Associate Partner of GBIF for Antarctic biodiversity information.

SCAR: **Scientific Committee on Antarctic Research**, ([www.scar.org](http://www.scar.org)). SCAR-MarBIN is an Action Group of SCAR. SCAR provides a direct connection with the Antarctic scientific community, and provides data and strategic advice to SCAR-MarBIN and AntaBIF. AntaBIF is embedded as the data management component of the upcoming SCAR Biology programmes (AntETR and AntECO, see below).

CoML: the **Census of Marine Life** ([www.coml.org](http://www.coml.org)). Project officially ended in 2010, SCAR-MarBIN played the role of data component for the Antarctic project of the Census, CAML.

CAML: the **Census of Antarctic Marine Life** ([www.caml.aq](http://www.caml.aq)). The Antarctic project of the Census of Marine Life. SCAR-MarBIN is the sister project of CAML, as its data management arm. Still a lot of data to be published, and various sub-networks still kept alive through ongoing project, such as the Biogeographic Atlas of the Southern Ocean (see details below).

GEO-BON: the **Group on Earth Observations** - Biodiversity Observation Network ([www.earthobservations.org/geobon.shtml](http://www.earthobservations.org/geobon.shtml)). AntaBIF is invited to play the role of Antarctic Node for GEO-BON.

Royal Belgian Institute of Natural Sciences

EoL: the **Encyclopedia of Life** ([www.eol.org](http://www.eol.org)). SCAR-MarBIN provides material directly to EoL (taxonomy, pictures, distribution) on over 5,000 species, through the webportal, and also through the new Antarctic Field Guides ([afg.biodiversity.aq](http://afg.biodiversity.aq)).

SOOS: the **Southern Ocean Observing Systems** ([www.scar.org/soos](http://www.scar.org/soos)). SCAR-MarBIN will play the role of biodiversity node for SOOS, starting in January 2011.

AntETR/AntECO: Antarctic Ecosystems: Adaptation, Thresholds and Resilience/State of Antarctic Ecosystems. A proposal for a new SCAR biology-program presented by the Standing Scientific Group on Life Sciences (SSG-LS), Expected duration: 10 years (2012-2022). AntaBIF is the Data Management component for both programmes.

SCADM: the **Standing Committee on Antarctic Data Management** ([scadm.scar.org](http://scadm.scar.org)). AntaBIF project manager acts as Deputy Chief Officer.

It is envisioned that many new combinations will arise in the future, along with new biodiversity informatics technologies and standards being developed at an exponential speed. The multiplicity of contexts in which the data is made available is a very healthy sign, and will accelerate far-seeing shifts in science, conservation and management.

## Science

AntaBIF constantly seeks direct contact with the scientific community, in order to ensure that the current developments answer the needs of the community it is serving and are inline with best available technologies and standards. It allows determining the best options, and choose priorities for future developments. Scientific publication is used as a means to maintain credibility and ensures that SCAR-MarBIN and AntaBIF are cited in the influential literature and embedded in strategic documents on Antarctic Research (eg SCAR strategic plan 2011-2016, and SCAR AntETR programme 2012-2022). To date, SCAR-MarBIN or derived products are **mentioned in >80 publications** (as found through a [Google scholar search](#)).

Some publications are completely based on data available through the webportals. A recent example is the following book chapter:

Bluhm, B, D. Watts, and F. Huettmann (2010). Free Database Availability, Metadata and the Internet: An Example of Two High Latitude Components of the Census of Marine Life. Chapter 13, pp. 233 – 244. In: S. Cushman and F. Huettmann. Spatial Complexity, Informatics and Wildlife Conservation. Springer Tokyo, Japan. Pp. 233-244

Recent publications include:

Danis B. & Griffiths HJ., 2009. Polar science: bid for freely accessible biodiversity archive. Nature (<http://www.nature.com/nature/journal/v458/n7240/full/458830b.html>)

Griffiths HJ., Danis B. and Clarke A., 2010. Quantifying Antarctic marine biodiversity: The SCAR-MarBIN data portal. DSRII ([doi:10.1016/j.dsr2.2010.10.008](https://doi.org/10.1016/j.dsr2.2010.10.008))

De Broyer C., Danis B. and 64 taxonomic editors, 2010. How many species in the Southern Ocean? Towards a dynamic inventory of the Antarctic marine species. DSRII ([doi:10.1016/j.dsr2.2010.10.007](https://doi.org/10.1016/j.dsr2.2010.10.007))

Parsons M et al., accepted. "Data for modern, complex, interdisciplinary science—another dimension of the Fourth Paradigm", Journal of Information Science.

Gutt J et al., submitted. "How to use species distribution modelling for ecological predictions in the Antarctic, a cross-disciplinary concept". Biological Reviews.

Danis B & Griffiths HJ, in prep. "Sharing biodiversity data during the IPY: a longterm vision to shift the norms" in di Prisco G, Walton D, Kallenborn R (Eds), "From Pole to Pole: Polar Environmental Research during the International Polar Year 2007 – 2009". Springer book series.

## Side Projects, data products

In parallel with the networks maintenance and development activities, a series of projects are coordinated and contribute to the visibility of the SCAR-MarBIN and AntaBIF projects. These projects are mostly taken over by dedicated communities of contributors and play a very important role in motivating the scientific community, who sees an added value when sharing their data. Some of these projects are detailed below:

### 1. The Antarctic Field Guides (AFG):

The objective of the AFG project is to develop a community-driven, online identification aid, which takes the form of an interactive, customizable Antarctic Field Guide ([afg.biodiversity.aq](http://afg.biodiversity.aq)). The field guide offers access to dynamically-generated information (custom-built from the content of RAMS, SCAR-MarBIN, GBIF and AFG custom databases), and allows the user to create his own field guide to the Antarctic wildlife, in function of the region he/she is interested in, or the taxa he/she wants to focus on. As part of a community effort, a substantial amount of information has already been made available by Prof A. Clarke and Dr S. Schiaparelli. The front-end (public website) of the project was developed by Vizzuality ([www.vizzuality.com](http://www.vizzuality.com)).

The new SCAR Expert Group on Birds and Marine Mammals (EG-BAMM) is progressively contributing to the AFG with the best available material for what is expected to be one of the most attractive content of the AFG for the general public.

An example of Field Guide for the Antarctic Sponges can be visualized here: <http://afg.biodiversity.aq/guides/144164-a-field-guide-to-antarctic-sponges>

*The Antarctic Field Guides is an Open Source project.*

### 2. The Biogeographic Atlas of the Southern Ocean:

The Biogeographic Atlas will take the form of a collection of maps and synthetic texts presenting the key biogeographic patterns, and their causal processes, of the main benthic and pelagic/nektonic taxa and assemblages/communities in the Southern Ocean south of 40°S. There will be a hard copy edition in large format (somewhat like the [Antarctic Map Folio Series in the sixties, e.g. Hedgpeth 1969 Folio 11](#)) as well as a dynamic online version built on-the-fly from the content of the databases. On the hard copy version synthetic texts will face the selected maps. Illustrative maps will be the key elements of the Atlas (which is not intended to be a Treatise of the Southern Ocean biogeography). As for the AFG, the front-end (public website) of the project will also be outsourced. A strong focus on predictive mapping will be included, in order to bring a modern approach to the classic Hedgpeth volume, using an array of modelling techniques (GLM, BRT, OpenModeller, RandomForest,...), which will require strong collaboration between taxonomic experts and modelers.

So far a total of over 60 specialists (taxonomists, ecologists, modelers, developers) have accepted to contribute data and expertise to this project, which will feed the SCAR-MarBIN/AntaBIF data systems. In the long term, it is hoped that this approach can be expanded to non-marine realms.

*The Southern Ocean Biogeographic Atlas is an Open Source project.*

## Communication

Apart from its websites, reports and scientific publications, AntaBIF communicates through various channels, including direct communication, participation in meetings, interviews, social networks, blogging and micro-blogging. A few relevant links include:

AntaBIF blog: <http://antarcticbiodiversity.blogspot.com/>, 70 posts since May 2009, 5000 pages view

AntaBIF Facebook group: biodiversity.aq, <http://www.facebook.com/group.php?gid=203309517162>, 127 members, up since December 2009

AntaBIF twitter channel: username: @biodiversityaq, 300 tweets, 70 followers, up since December 2009

### Meetings attended/organized

Royal Belgian Institute of Natural Sciences

2011:

MPA/Biogeographic Atlas workshop (Brest, France). [co-convening]

Biosystematics conference 2011 (Berlin, Germany). [presentation]

IOC's International Oceanographic Data and Information Exchange (IODE) Programme 50th anniversary conference (Liège, Belgium). [participation]

SCAR Crosslinkages workshop (Ottawa, Canada). [presentation]

2010:

AntaBIF kickoff workshop (Buenos Aires, Argentina) [organization]

Polar Macroscope Synthesis workshop (Brussels, Belgium) [co-organization]

PIC workshop (Paris, France) [presentation]

IPY Science conference (Oslo, Norway) [presentation].

SCAR Open Science conference (Buenos Aires, Argentina) [keynote presentation]

SCAR Life Science Standing Scientific Group meeting (Buenos Aires, Argentina). [presentation]

SCADM Meeting (Buenos Aires, Argentina). [presentation]

CAML SSC workshop (Buenos Aires, Argentina). [presentation]

EG-BAMM kickoff workshop (Buenos Aires, Argentina). [presentation]

OBIS Steering Group meeting (Oostende, Belgium). [presentation]

Belgian IPY symposium (Brussels, Belgium). [presentation]

Positive Visions for Biodiversity conference (Brussels, Belgique) [participation]

OLA-CAML workshop (Buenos Aires, Argentine) [participation]

Census of Marine Life 2010 : a decade of Discovery (London, UK). [participation]

### **Meetings planned for 2011**

SCADM workshop (Palma de Mallorca, Spain), 7-9 September [co-organization]

World Conference on Marine Biodiversity (Aberdeen, Scotland), 26-30 September [2 presentations]

SCAR-MarBIN workshop (Aberdeen, Scotland), 30 September [organization]

AntaBIF workshop (Aberdeen, Scotland), 1 October [organization]

## **Technical developments**

AntaBIF IT developer, Ir Nabil Youdjou, is in charge of the practical exploration and implementation of the new IT architecture for AntaBIF. The idea underlying this architecture is to offer an added value to the user and partners, by integrating the best available resources and standards, such as those developed in the frame of [SCAR-MarBIN V2](#), building upon webservice-oriented, dynamic, technologies. Careful attention is also taken to make the best possible use of GBIF informatics tools.

Using this approach, the networks will be interoperable with many entities, allowing the publication of the data in many different contexts, in a distributed fashion.

The AntaBIF technical architecture is organized around the following components: data harvesting, metadata, occurrence, taxonomy, genetics, search engine. Details are given below:

**AntaBIF technological ecosystems:**

The technological ecosystem is largely based on the progress made in the framework of the [SCAR-MarBIN V2 data portal](#). It is based on 100% Open Source solutions, and integrates the latest GBIF informatics components (Harvesting and Indexing Toolkit (HIT) and Integrated Publishing Toolkit (IPT)). Details are given in the table below:

Item	Technological solution
Design patterns	MVC (Model-View-Controller)/ORM (Object-Relational-Mapping)
Framework	Rails (ActiveRecord) and YUI (User Interface Library)
Search engine	Full text (Elasticsearch-Lucene)
Language	Ruby
Database	PostgreSQL
GIS server	Geoserver
Spatial database	PostGIS
Mapping client	OpenLayers
Web services	RESTish (all resources)
Protocols	DIF (Data Interchange Format), DarwinCore, DarwinCore archive, Tapir
GBIF tools	HIT (Harvesting and Indexing Toolkit), IPT (Integrated Publishing Toolkit)
OS	FreeBSD
Hosting	BBPF (ULB/VUB joint IT Center)
Metadata systems	GCMD (Global Change Master Directory, mirrored), through API

**AntaBIF data harvesting component [COMPLETE]**

To build the content of its databases, AntaBIF will be using a customized version of [GBIF's HIT](#) (Harvesting and Indexing Toolkit). Most of the work so far on this component has been directed towards a technical customisation of the HIT, improving the coherence and integration at the level of occurrences data.

Technically, AntaBIF and HIT will be managed by the same database management system (DBMS), which takes the form of a PostgreSQL relational database. Another advantage of this approach is to induce less transformation on the original data than in the actual SCAR-MarBIN data model, which will reduce potential data corruption. Another customization to the GBIF tool include the possibility to create new datasets even if they haven't yet been registered in the [Global Biodiversity Resources Discovery System](#) (GBRDS).

DONE: Debugging and testing phase using AntaBIF principal data providers (using DiGIR and DWC-A protocols) is now completed.

**AntaBIF metadata component [IN PROGRESS]**

AntaBIF metadata system will mirror the relevant [Global Change Master Directory](#) (GCMD) content, using an API. NASA has recently agreed to setup a special access for AntaBIF, to ensure an efficient interoperability between the GCMD and AntaBIF metadata systems.

DONE: So far, the HIT metadata model has been implemented and first batches of data have been harvested. 4 data providers were considered for the testing/debugging phase, including ("ScarMarbin", "Australian Antarctic Data Centre", "Antarctic Bacteria diversity", "PANGAEA - Publishing Network for Geoscientific and Environmental Data").

**7,200** datasets have been harvested (196 in SCAR-MarBIN).

TO DO: The development of the GCMD model in AntaBIF database still needs to be done.

### **AntaBIF Occurrences component [INSTALLED AND USED]**

HIT harvests the raw occurrences, which are integrated in the AntaBIF database. The taxonomy as well as the geospatial data scopes extractors have been developed.

DONE: So far, **2,659,392** records have been harvested and integrated in the database. Among these records, **914,566** are georeferenced (below -50S latitude). Details are given in the table below:

<b>Data Provider</b>	<b>All_records</b>	<b>Georeferenced</b>	<b>% Georeferenced</b>
SCAR-MarBIN	284911	284823	99,97
Australian Antarctic Data Center	357212	356433	99,78
British Antarctic Survey	135137	134107	99,24
PANGAEA (Alfred Wegener Institut)	1882132	1882132	100
<b>TOTAL</b>	<b>2659392</b>	<b>2657495</b>	<b>99,93</b>

TO DO: The occurrences model needs to be developed as well as the ETL (Extract, Transform, Load) tools for contexts and scopes extraction (taxonomy candidates, temporal data, geospatial data).

### **AntaBIF taxonomic component (Register of Antarctic Biodiversity, RAB) [IN PROGRESS]**

The approach used to build the taxonomic component of AntaBIF is based on the occurrence model, generating a list of candidate names used to retrieve taxonomic information retrieval. This approach allows to avoid duplication of efforts ungoing under other initiatives such as RAMS, CoL, FADA. The AntaBIF taxonomic component is capable of interacting with the latter initiatives. The AntaBIF taxonomic component (RAB) is designed as a pointer to other taxonomic data providers, including direct access to a minimal amount of locally stored.

DONE: Client web-services for Rams, Col and FADA have been developed, building the first verions of RAS, which includes a total of **30,472** scientific names (17,184 in RAMS), **37,376** vernacular names (432 in RAMS), and **66,557** (10,000 in SCAR-MarBIN) literature sources references.

TO DO: The Taxonomic candidates extractor tool from occurrences as well as client web-services for other taxonomic data providers still need to be developped.

### **AntaBIF genetics component [IN PROGRESS]**

The objective of this component is to integrate the DNA-related information related to taxonomy, and vice-versa (obtaining taxonomic information based on DNA data).

DONE: So far, a client web-services prototype has been developped, retrieving DNA sequences in Blast format from GenBank. The preliminary test are promising and it is foreseen that DNA information should be integrated into the AntaBIF database.

TO DO: The implementation of the client web-services to retrieve the DNA data form NCBI, related to a RAB taxon as well as local DNA database for the taxa for which only DNA-based taxonomy is available from GenBank tools need to be developped.

**AntaBIF search engine [IN PROGRESS]**

The AntaBIF search engine will be capable of searching full text content in all AntaBIF databases, within any context (taxonomy, DNA, spatial, temporal, metadata), offering the user a very flexible access to data.

DONE: An index server (Elasticsearch node) has been installed, the design for the search model has been developed , and the implementation of scripts to create, update, index objects, based on the database context is in test phase. Preliminary testing on retrieving taxonomic information has been carried out as well.