

# mARS Standard Operating Procedure (SOP)

August 2014

## About

This SOP details how you can upload (meta)data to the microbial Antarctic Resource System (mARS). To ensure this procedure only has to be carried out once, the mARS team has devoted special care to following widely-used standards for biodiversity data, as promoted by the [Global Biodiversity Information Facility](#) and the [Genomics Standards Consortium](#). In this particular case, this SOP is built around two main types of standards, namely [DarwinCore](#), and [MIxS](#) (where [MiMARKS](#) has been designated), ensuring maximal interoperability with internationally-recognized data and metadata repositories.

There are three different “entry-levels” from which you can participate in mARS (described below in 0.1). If you have sequence data set(s) that you would like reflected in mARS, we are requiring that (i) the data is accessible in a public repository (Genbank, IMG-M, or other web-accessible site) and (ii) two additional metadata files “MiMARKS” to communicate sample-associated environmental data ([Yilmaz et al. 2011](#)) and “Microbial sequence set spreadsheet” be attached, which summarizes sequence-related information and allows us to access and represent sequence data you have published in open repositories (eg. GenBank).

To be clear from the outset, mARS will NOT store sequence data; it is intended that the sequence data produced by your project is submitted to a public repository. The [vision](#) for mARS is to provide forum for project data discovery and exchange in the Antarctic scientific community and beyond. (For more information on topics of interest, click the links associated with blue/underlined words).

To upload your data, simply follow the following procedure step-by-step. If you have any difficulties, don't hesitate to [get in touch with us](#).



## 0. Before you start

## 1. Describe your project in the IPT

## 2. Provide environmental data on mARS

## 3. Provide sequence data on mARS

## 4. You're done!!

The diagram above provides an overview of the project work flow, and steps that are further described in this SOP.



## 0. Before you start

1. Send an email to request a username and password from the [IPT administrator](#) of ipt.biodiversity.aq.
2. Get a google login if you don't already have one - to copy the two template files that you need, things work best if you have a google/gmail login.
3. Make a copy or download the [MiMarks Googlesheet](#) from the RDP MiMarks Googlesheet that we have a link to here (**click on “Make copy” or “from the “File” menu**). If you cannot access Google Documents, get in touch with us.
4. Make a copy of the [Microbial Sequence Set](#) from the mARS Googlesheet (**click on “Make copy” from the “File” menu**). If you cannot access Google Documents, get in touch with us.



## 1. Describe your project in the IPT.biodiversity.aq

1. Login to the [IPT](#) using your credentials: the “Manage Resources” tab will appear at the top of the page.
2. Use the form at the bottom of the “Manage Resource” page to create a new resource. Provide a unique “shortname” for your dataset.
3. Under “Type” select “Metadata only” or “Other”; at this point, the IPT is not designed to take environmental or molecular diversity information at this point - so we will handle the data outside of GBIF and serve the data and links to sequence in mARS.
4. Click the “Create” button. You will arrive on the Resource Management page.
5. Click on the “Edit” button in the Metadata section on the left and fill in the details for the different metadata sections. A detailed instructions are available from [IPT quick reference guide](#). *Hint*: mention your grant number in the “Project Data” section, to allow us to link your resource to relevant projects in the GCMD/AMD.

### Hints for filling out the IPT forms:

“ \* ” signifies fields that need to be filled out

There's a navigation bar to the right with various forms that require information

Basic Metadata sheet: under “Title” enter a descriptive title for your project.

Temporal Coverage sheet: you can either enter specific sample seasons or dates if you have few samples, or if a long term project, then enter the complete date range.

Keywords: You can enter “n/a” in the Thesaurus Vocabulary box. Enter Keywords that can be used to find your project. This will be searchable in GBIF among other search engines.

Collection data: These terms don't apply to the mARS data sets perfectly; though you can still enter relevant terms for your data. This data is particularly relevant if you do have archived samples either in your lab or in the DNA archive in New Zealand. Enter Not Applicable where appropriate.

External Links: This is a good place to put the URL for a project website if you have one.

Additional Metadata: If your data set is registered in other data bases (e.g. Antarctic Master Directory; GenBank or the Short Read Archive, BCO-DMO) then enter that information here.

## 2. Publish and register your data

1. Back in the IPT website, from your Resource Management page, click on the “Publish” button in the “Published release” section on the left of the page. Do not worry when you see a warning message “Source data or Darwin Core mappings missing. No data archive generated. Please note that this is perfectly normal if you are publishing a metadata-only resource.”
2. By default, your resource’s visibility is set to “*Private*”. To allow your resource to become visible on the IPT for all users, click on the “*Public*” button in the “Visibility” section.
3. Send an email to one of the mARS data administrators to request final “registration” of your data set. This allows your metadata to be discoverable through online biodiversity information networks. Your data set receives a stable, unique identifier on the web.

## 3. Prepare your MiMarks spreadsheet

1. In the [MiMarks GoogleSheet](#) you’ve created in step 0, fill in your environmental metadata details using the “Google Documents” interface, following the instructions available from the MiMarks GoogleSheet [documentation at RDP](#). Example files are available from the mARS website.
2. In the header for each column that will hold your sequence set data, list the unique identifier of your sequence set.
3. Once you are finished, download your spreadsheet as a CSV (Comma-separated Values) file on your computer.

## 4. Upload your MiMarks CSV(s) in mARS [Still in DEV]

For now just mail us your Mimarks and Microbial Sequence set CSVs

1. Login the mARS webportal, using the same credentials as those you have used for the IPT.
2. On the mARS webportal, select “upload my data” from the XXX menu.
3. Follow the steps you will be guided through.
4. Once you have followed the process, click on “Submit”. Your data will be checked manually for consistency, and then indexed in the mARS database.

## 5. Prepare your Microbial Sequence Set spreadsheet

1. In the [Microbial Sequence Set GoogleSheet](#) you’ve created in step 0, fill **all** the fields (replace the examples available from the GoogleSheet)
2. Once you are finished, download your spreadsheet as a CSV (Comma-separated Values) file on your computer.

## 6. Upload your Microbial Sequence Set CSV(s) in mARS [Still in DEV]

For now just mail us your Mimarks and Microbial Sequence set CSVs

1. Login the mARS webportal, using the same credentials as those you have used for the IPT.
2. On the mARS webportal, select “upload my data” from the XXX menu.
3. Follow the steps you will be guided through.
4. Once you have followed the process, click on “Submit”. Your data will be checked manually for consistency, and then indexed in the mARS database.

## 7. You're done, go shopping or bar hopping!

Get in touch with us when you've gone through the process.

Do you need a GCMD/AMD identifier? Send us an email, and we'll sort that out for you.

## 8. You realize you made a mistake, or want to update the IPT

1. You can edit any field. Log into the IPT, go to manage resources overview, change the fields that you want to change. Go to the Published Release box on the left panel and hit “publish” again.
2. You must email the mARS administrator to notify them of your update.

## References/links:

Yilmaz P., R. Kottman, D. Field, R. Knight, J. R. Cole, L. Amaral-Zettler et al. (2011). The “Minimum Information about a MARKer gene Sequence” (MIMARKS) checklist: Capturing contextual data about marker gene sequences and introducing MIxS, a unified standard for sequence checklist development including environmental data. *Nat Biotechnol.* <http://www.nature.com/nbt/journal/v29/n5/pdf/nbt.1823.pdf>

DarwinCore: <https://code.google.com/p/gbif-providertoolkit/wiki/DarwinCore>

MiMarks Wiki: <http://wiki.gensc.org/index.php?title=MIMARKS>

MiMarks GoogleSheet documentation at RDP: <http://rdp.cme.msu.edu/misc/googleSheetsHelp.jsp>

IPT quick reference guide: [https://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes#Quick\\_Reference\\_Guide](https://code.google.com/p/gbif-providertoolkit/wiki/IPT2ManualNotes#Quick_Reference_Guide)

mARS white paper: <http://marswhitepaper.blogspot.be>

mARS FAQs: <http://mars.biodiversity.aq/howto>

Global Change Master Directory: <http://gcmd.nasa.gov>

Antarctic Master Directory: <http://gcmd.gsfc.nasa.gov/KeywordSearch/Home.do?Portal=amd&MetadataType=0>