An Ecologist's Guide to BIIGLE
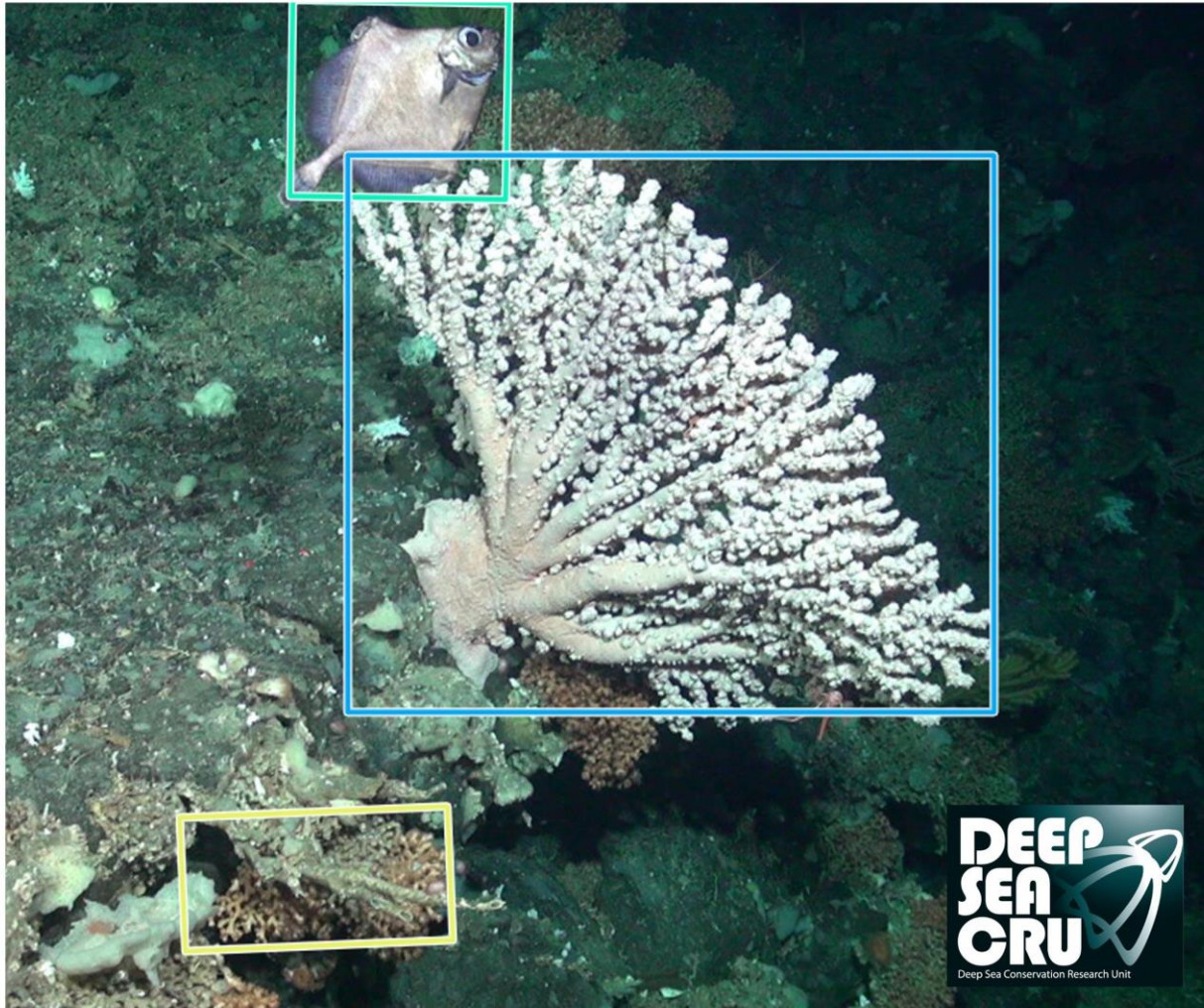
IMPORTANT INFORMATION: We intend for this to be a living document that others can contribute to. The manual is hosted on GitHub in our BIIGLE-resources repository. If you see something that you feel could be made clearer or needs changing, please let us know by flagging an issue on GitHub. This can be done by going to our BIIGLE-resources repository, clicking 'Issues' (top left), and then logging a 'New issue' (green button on the right). This will allow us to update and improve this resource as we intend to do so annually. A PDF version of this manual is available for download on Zenodo, here.

# Table of contents

AN ECOLOGIST'S GUIDE TO BIIGLE

KL Howell, AEH Bridges, J Davies, A Parimbelli & N Piechaud   17/03/2023

1. Introduction

Hello and welcome to our BIIGLE manual. This document is intended to help ecologists get started using BIIGLE to annotate their image and video data. The manual was originally written to support new members of the Deep-Sea Conservation Research Unit as well as undergraduate students at the University of Plymouth. It is a collection of information that we have found helpful to document in our experience of setting up and using BIIGLE. It is not intended as a replacement for the official BIIGLE manual available here https://biigle.de/manual.

It is fair to say our current instruction manual is bias toward image data (although video is included), and only provides instructions for setting up using Amazon Web Service as the host for your imagery. If you can provide instruction for setting up on other cloud services, please contribute. But first things first, you need to create a BIIGLE account. To create an account, click on the "sign up" button in the top right corner of the website (https://biigle.de/) homepage. Fill in the module with your details and affiliation, choose a password and, after agreeing to the privacy notice and the terms of use, click on sign up. To log in, click on the green "login" button and insert the email used to sign up, and the chosen password.
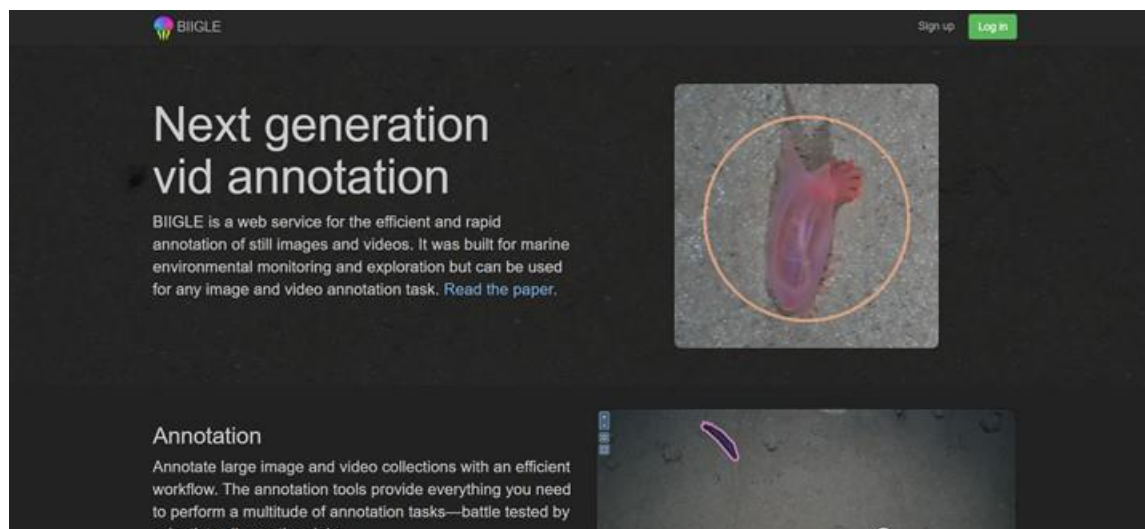
Happy annotating!

2. Setting up on BIIGLE

Getting set up on BIIGLE (and other annotation softwares) first requires that you host your imagery data, there are three options for this: via BIIGLE (upload files on homepage); local instances and remote volumes. See the BIIGLE manual for more details. Many institutes use a cloud-based storage solution to host their remote volumes. This is not the same as a google drive or equivalent where you might store or share data. It is a formal repository that enables your images to each have a unique URL. There are many types of cloud-based storage solutions that you may wish to consider, key ones we are aware of include Amazon Web Services (AWS), and Microsoft Azure. Each uses its own terminology to refer to its storage

'containers'. AWS calls its containers 'Buckets', Microsoft Azure calls them 'Blobs'; they are the same thing.

Below we describe the process for setting up and working with AWS. We invite others to contribute similar text for other cloud-storage options.

2.1 How to get set up with an Amazon Web Services remote server

Create an Amazon Web Services account.

On the AWS management console type "S3" into the find services search box.
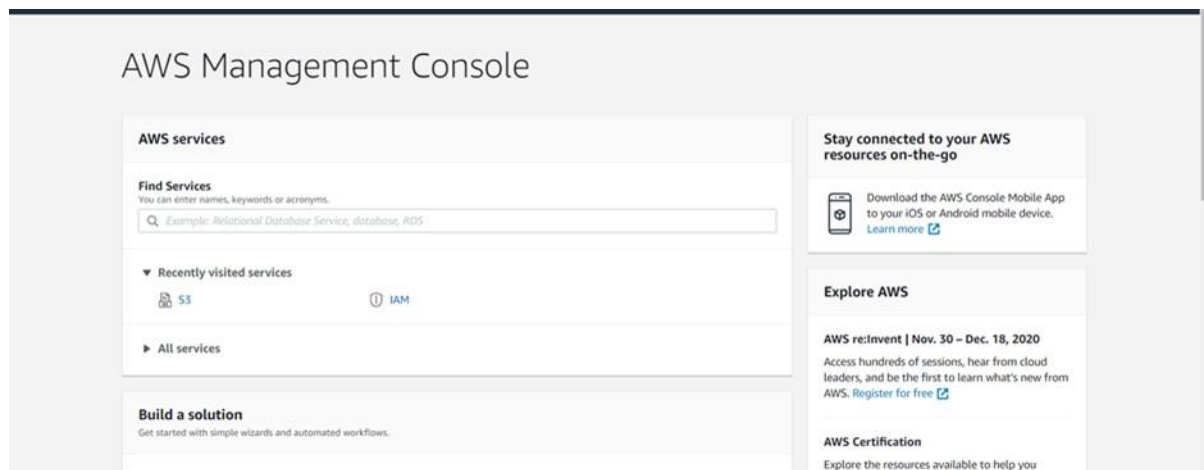
Select "S3"

Before you can upload data to Amazon S3, you must create a bucket in one of the AWS regions to store your data.

Click on "create bucket"

Name your bucket in a sensible way. BIIGLE recommends you use their personal random string generator (available here https://biigle.de/manual/tutorials/volumes/remote-volumes) to do this as this is hard to guess and keeps your data safer. But you can use whatever name you like.

Select your nearest region

Uncheck the "Block all public access" box as BIIGLE needs to see your folder.

Leave all other settings as default and click on "create bucket"

Once you have created your bucket, click on it to take you to this screen
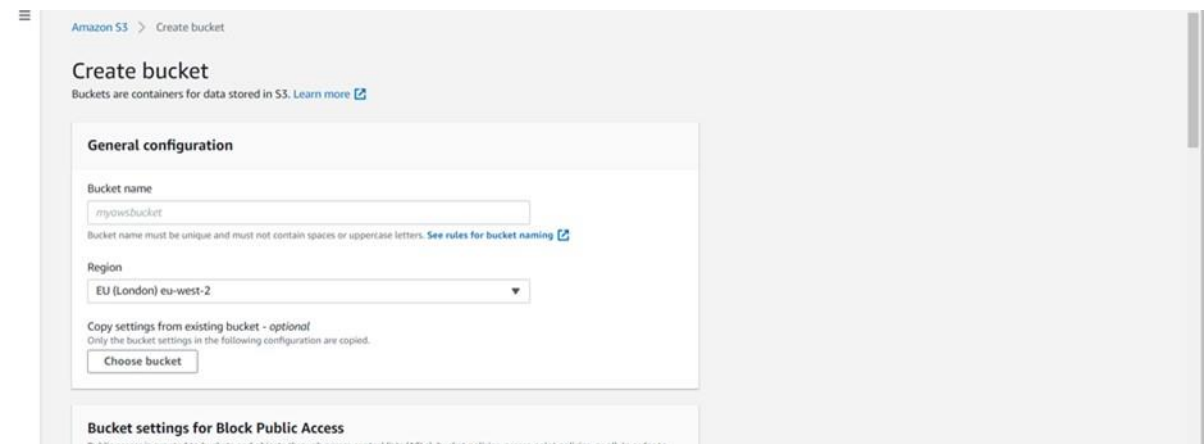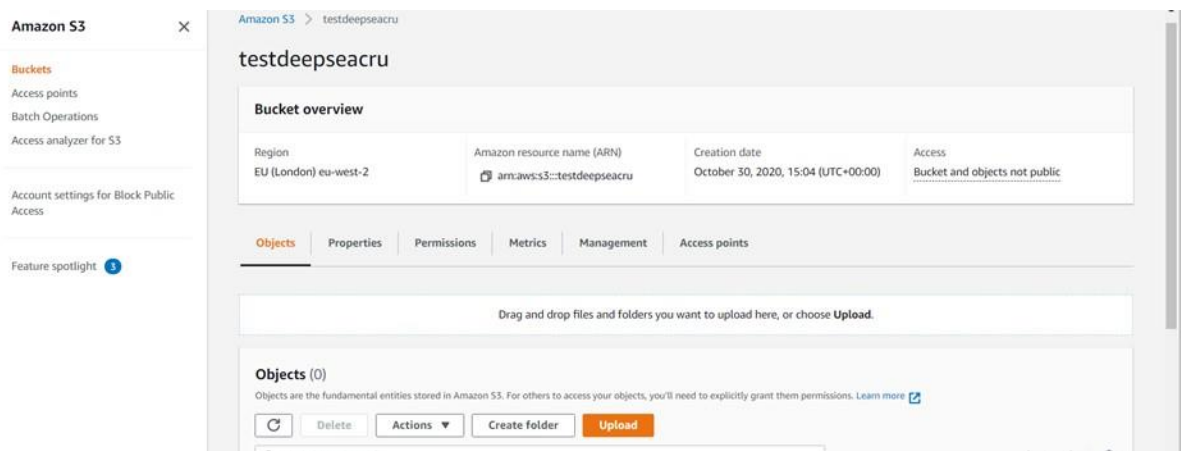

This next section may not be necessary but is included for completeness.

Select the permissions tab

Scroll down to Cross-origin resource sharing (CORS) and click on edit.

The cross-origin policy is a security mechanism of web browsers that prevents malicious third parties from extracting sensitive information from your web pages. This includes cases like loading files from remote sources in BIIGLE. Although those files can be displayed, BIIGLE cannot access the raw data in the browser which is a requirement for some features of the image annotation tool. Cross-Origin Resource Sharing (CORS) is a mechanism to manually configure exceptions for the cross-origin policy. With a correct CORS configuration, BIIGLE can process files from remote sources just like regular files.

To set up CORS for the files of your remote source (in this example we are using AWS), you have to update the configuration of the webserver that serves the files.

Copy and paste the script below into the text box. Then save changes.

```
[
    {
        "AllowedHeaders": [
```
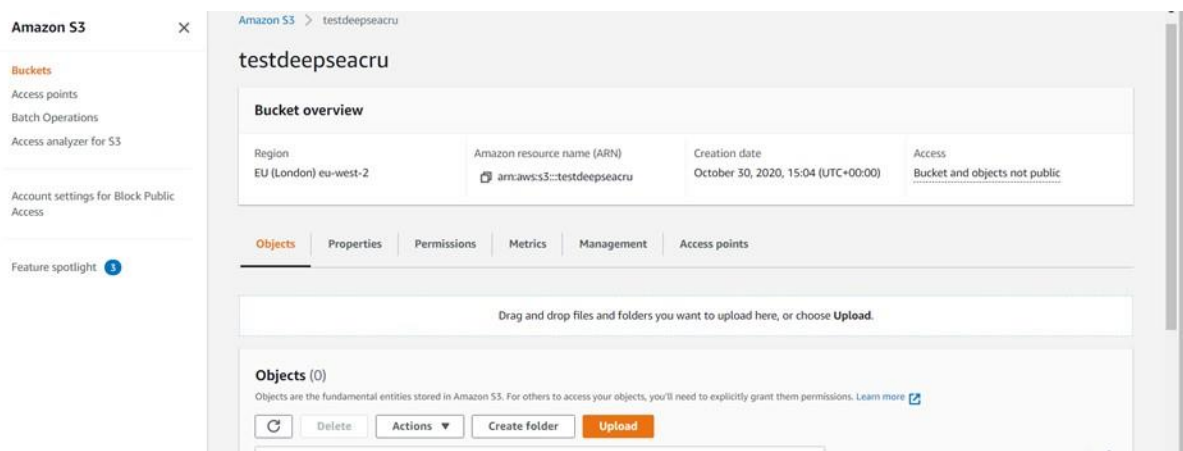
```
        "*"
    ],
    "AllowedMethods": [
        "GET"
    ],
    "AllowedOrigins": [
        "*"
    ],
    "ExposeHeaders": []
  }
]
```

Continue from here if skipping the previous section

The following section is repeated under the 'Setting up your BIIGLE volumes' section. We include here as it is important you test if your AWS bucket is set up correctly before moving on.

To test that your setup is working go back to the "Objects" tab.

This is where you can upload files, so create a new folder, for example called test.

Click on your folder and upload a couple of images.

Click "upload", click "add files", add your files, scroll down to the access control list and check the "everyone (public access) read check box. A new check box will appear to confirm that you understand the implications of making your files public access, just check it, then scroll down to the bottom of the page and click"upload". Once uploaded, you can check that the files will be accessible by BIIGLE by clicking on the object URL, if it has been uploaded correctly with public settings enabled, the image or video will open. If the file does not open, check the settings.

Now return to your folder and click on a newly uploaded image. In the overview tab you will see the object URL, you will need to copy this but only the bit up to and including your folder name. This is what BIIGLE will need to link up to.

Now log into BIIGLE and go to your dashboard.
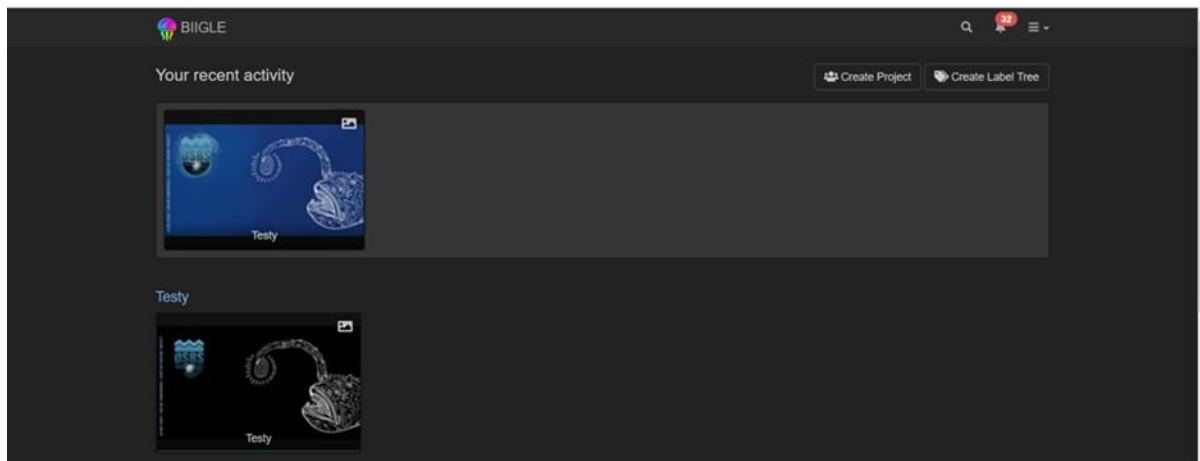
Click on create project

Click on create volume

Here you can name your volume, for example it might be a cruise name, a project name or other useful name.

In volume URL paste the URL for your Amazon S3 bucket and folder. In the volume files you will need a list of the names of all the images that are in the file. Below is an R Script to read a directory containing images, and place a comma separated list on the clipboard for you to paste into BIIGLE or a text file for later use.

```
setwd("F:/DSCRU-frames-main/renamed frames/DIVE295_SCORPIO_4")
list.files() -> imgs
    writeClipboard(
      paste0(
        unlist(paste(imgs,collapse = " , ")
              ) ,
              collapse = " , "))
```

2.1.1 Enabling others to access your S3 AWS bucket when working in a team

To set permissions for users of your S3 AWS you will need to use IAM. Click on services in the top left of the AWS management console and type IAM in the all services search facility.
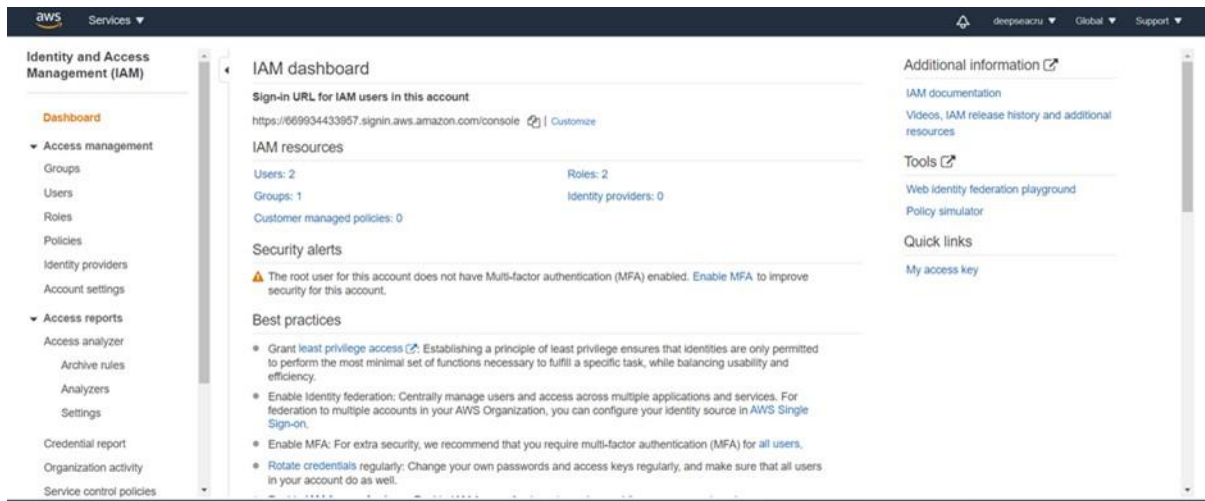
Click on groups on the left menu

Create new group

Name the group something sensible.

Now you need to state what permissions the group gets. So on the attached policy screen type "S3" in the policy type search. Then click on AmazonS3FullAccess.
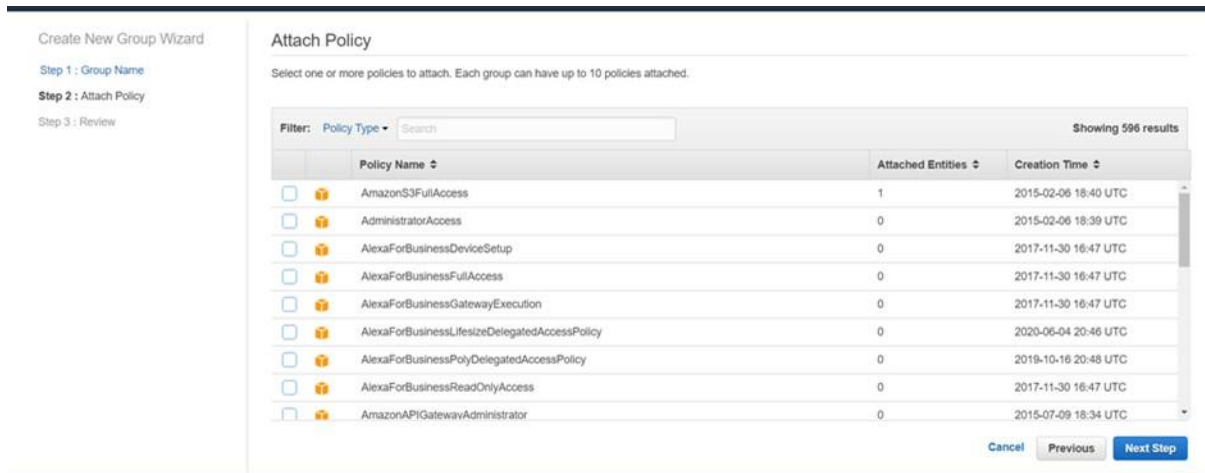
Then click next, then create group.

Now you need to add users to your group, so click on users on the left hand menu

Click on add user

Add user name and click both "access type" check boxes.

Select password options, then click next permissions

Check the group you want the user to belong to

Click "next: tags"

Click "next: review"

Click "create user"

Now it is very important to download the CSV file that gives you the key info for this user, you cannot get this info again if you change screen before downloading so do it now!

Ok you are all set!

2.2 Formatting your data for use in BIIGLE

BIIGLE only accepts still images as either PNG or JPEG format.

Video files: need to be converted to MP4 format (codec 254). This conversion can be done using the free software VLC (https://www.videolan.org/).

2.3 Converting video file to upload to BIIGLE

Take the following steps:

Media - convert/save - + add (find the file) – convert/save – profile (choose H.264 + MP3 (MP4) option from dropdown menu).

VLC media player

Media Playback Audio Video Subtitle Tools View Help

Open File...   Ctrl+O
Open Multiple Files...   Ctrl+Shift+O
Open Folder...   Ctrl+F
Open Disc...   Ctrl+D
Open Network Stream...   Ctrl+N
Open Capture Device...   Ctrl+C
Open Location from clipboard   Ctrl+V
Open Recent Media   ▶

Save Playlist to File...   Ctrl+Y
Convert / Save...   Ctrl+R
Stream...   Ctrl+S

Quit at the end of playlist
Quit   Ctrl+Q

17%

VLC media player

Media Playback Audio Video Subtitle Tools View Help

Open Media   — ☐ ✕

📄 File   💿 Disc   🖧 Network   💻 Capture Device

File Selection
You can select local files with the following list and buttons.

                                          ➕ Add...
                                          ➖ Remove

☐ Use a subtitle file
                                          Browse...

☐ Show more options

                        Convert / Save ▾   Cancel

17%

Choose where to save it and give it a sensible filename, see below

FAQs/Troubleshooting:

- VLC may corrupt when converting the video therefore check the length of the video converted matches the original.
- VLC typically take the length of the video to convert.
- You can set up batch conversion.

2.4 Recommendations for the file structure of your bucket and file-naming conventions

It is worth giving some careful thought as to how you set up your file structure under your bucket. Different people use different systems and have different ways of storing data but there are some common features such as CAMPAIN_NAME and DEPLOYMENT_NAME. We set ours up as follows, but you need to work out what works for you.

- `CLOUD_STORAGE_REPOSITORY_FOR_PLATFORM/`
  - `CAMPAIGN_NAME/` (for example research cruise number)
    - `DEPLOYMENT_NAME/` (for example ROV dive number)
      - `VIDEO_NAME/` (Video segment of ROV dive)
        - `ROV_20100625T130726_8888.jpg`
        - `ROV_20100625T130727_7234.jpg`
        - `...`

In practice for us that looks like this for one research cruise and considering image data:

- Amazon S3 > Buckets > 'bucket name'
  - JC136/
    - DIVE269/
      - DIVE269_SCORPIO_1/
        - 1000__Isis-DIVE269_SCORPIO_1_20160517-181525.jpg
        - 10000__Isis-DIVE269_SCORPIO_1_20160517-182125.jpg
        - 100000__Isis-DIVE269_SCORPIO_1_20160517-192125.jpg
        - ....

At this point it is also worth mentioning that it is good to follow a standard convention in how you name your images as it will make your life easier in the long run (python and R code you write for onward data use will run on all your files!). A possible standard is as follows, but again you must find what works for you and your data. Our ROV dive numbers are research cruise specific so we do not need to use research cruise in the image name, you might want to:

Example naming convention for image frames extracted from video ([CruiseNumber] is inserted as an example of where you might include it)

VideoFrameNumber_[*CruiseNumber*]_PlatformName_DiveNumber_VideoSegment_YYYYMMDD_HHMMSS

Example naming convention for video files (again with cruise number inserted)

[*CruiseNumber*]_PlatformName_DiveNumber_VideoSegment

There are no agreed standards on this issue and it would be useful if there were 🙂.

IMPORTANT NOTES: BIIGLE does not like spaces in file names, so use an underscore to separate out key parts of your file name. This also makes your file names machine readable. If you want your files to display in order when in BIIGLE you need to have numeric values at the start of the file name (e.g. frame number in the above image file example) or (in our video file example) start your video segment numbering with triple digits. i.e. 001 not single digit i.e. 1.

3. Setting up projects and volumes in BIIGLE

Once you have your files hosted, e.g. cloud storage set up, you can then go into BIIGLE to set up your projects and volumes. This process is relatively simple and you may already have done this when testing that your cloud storage works (see previous section).

First log into your BIIGLE account.

Go to your dashboard.
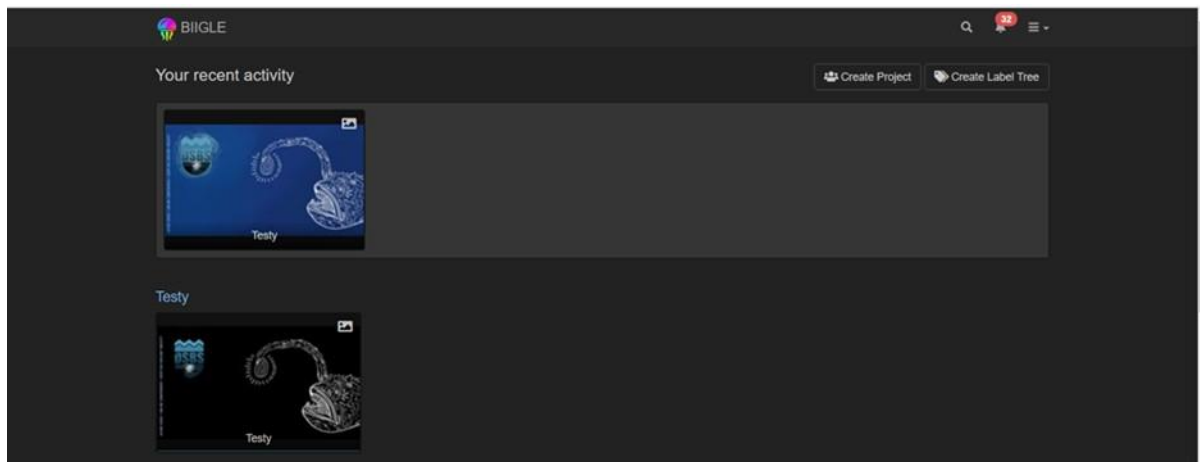
Click on create project

Adding imagery to projects can be done in two ways:

1. Create volume. This is done by attaching files from your remote server, e.g. from the Amazon server.
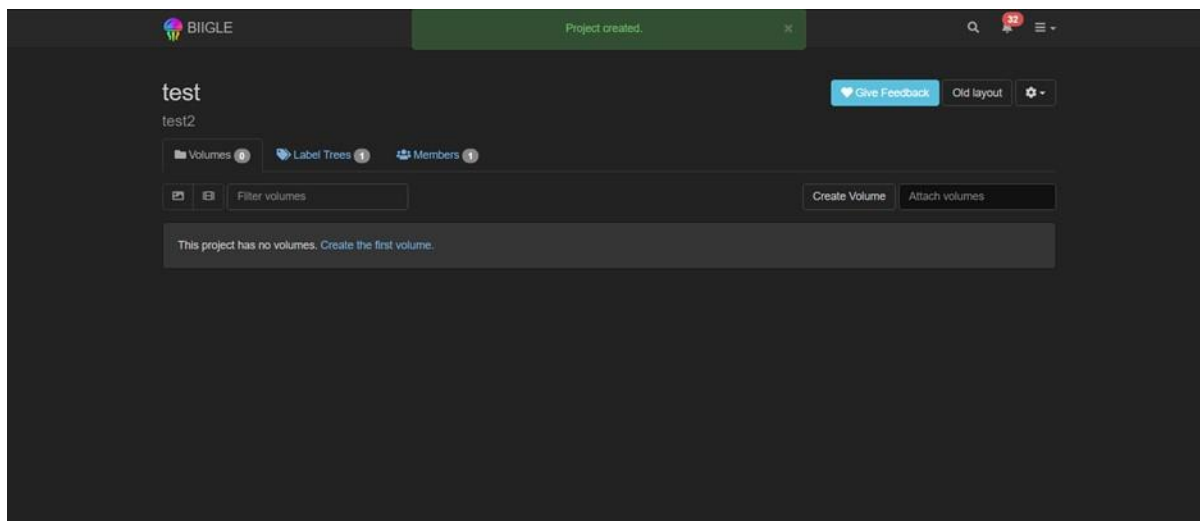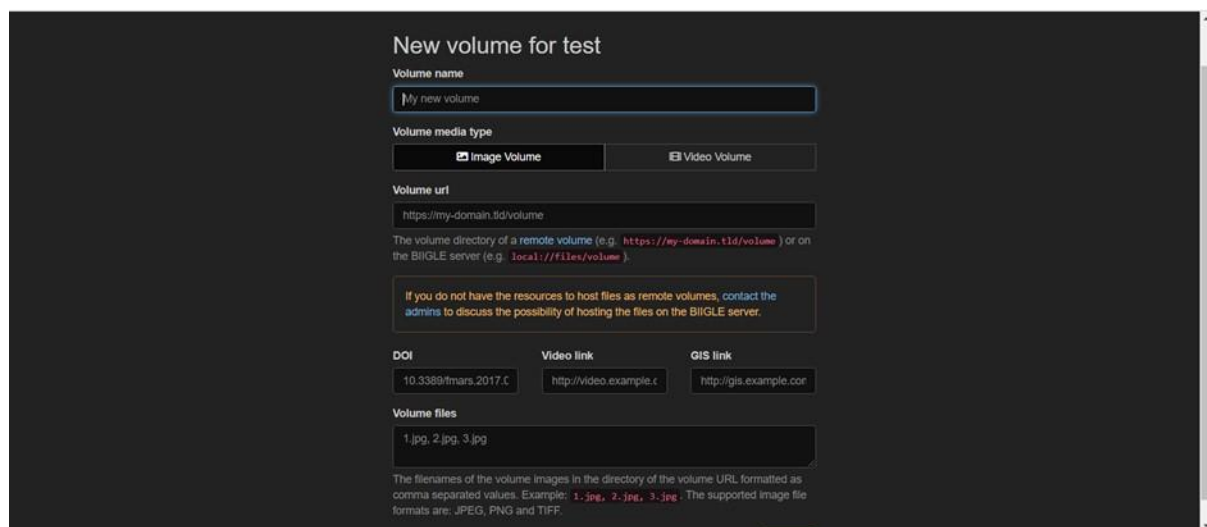2. Attach volumes. This is for volumes which are already in BIIGLE and you may need to attach the same volume to a different project. Start to type in the volume name, and it will display and you can click and it will add. Note, adding volumes this way attaches any annotations that have been made on the volume; if the original volume is deleted, the attached volume remains in the second project.

Creating a volume

Click on create volume

Here you can name your volume, for example it might be a cruise name, a project name or other useful name. You need to then specify whether it is an image or video volume.

In 'volume URL' you need to paste the URL for your Amazon S3 bucket and folder that relates to your volume. You will find this by doing the following: log into your AWS account (or equivalent remote server facility), go into your bucket and then file of interest, click on an image in your file of interest, in the overview tab you will see the object URL. Copy-paste 🙂

In the 'volume files' you will need a list of the names of all the images (or videos) that are in the file. Below is an R script that can produce this list for image data. This script produces a list in the correct format and places it on your clipboard. You can then paste this into the 'volume files' box.

```
# R Script to read a directory containing images, and place a comma separated
list on the clipboard for you to paste into BIIGLE or a text file for later
use

setwd("F:/DSCRU-frames-main/renamed frames/DIVE295_SCORPIO_4")
```

```
list.files() -> imgs
   writeClipboard(
     paste0(
       unlist(paste(imgs,collapse = " , ")
              ) ,
              collapse = " , "))
```

Now hit 'create', and this will create your volume. Your images (videos) should appear in that volume after a few seconds.

3.1 Recommendations for project / volume structuring in BIIGLE

It is worth considering how you set up the volumes. BIIGLE has evolved in recent years so that image and video volumes can now sit in the same project. This means that if you undertake Largo at the project level (see section 5) both image and video annotations will be displayed. Largo can also be undertaken at the volume level, but not on individual files (images or video) within a volume. When creating projects / volumes in BIIGLE, you can either:

- Have each video file as an individual volume
- Have all video files in a single volume.

There are pros and cons to both.

Separate volumes:

- Allows Largo to be undertaken for individual videos
- Can use the filter volume function to find a video
- Have more thumbnails for videos filling up volume level

Combined volumes:

- Better storage of all videos in one folder
- **Cannot** Largo individual videos, can only be done at the volume level - so all videos within that folder
- The filter volume function does not work to find videos, you would either have to scroll to find the video, or use the filter functions

3.2 Make label trees

Before you can begin labelling you need to create a label tree. We strongly advise you to use the SMarTaR-ID trees as this provides a standard ensuring that your annotations will be comparable with others annotations, e.g. your branching sponge morphospecies 5 is the same as someone else's. However, BIIGLE is very flexible and allows you to make your own trees if you prefer.

3.2.1 To use a publicly accessible (standard) tree

To use these or other publicly accessible trees we suggest you first create a 'fork' of that tree (e.g. your own copy) so that you can edit it and add your own new taxa that you will come across. To do this go to the top right drop down menu on the BIIGLE dashboard and click on 'label trees'.
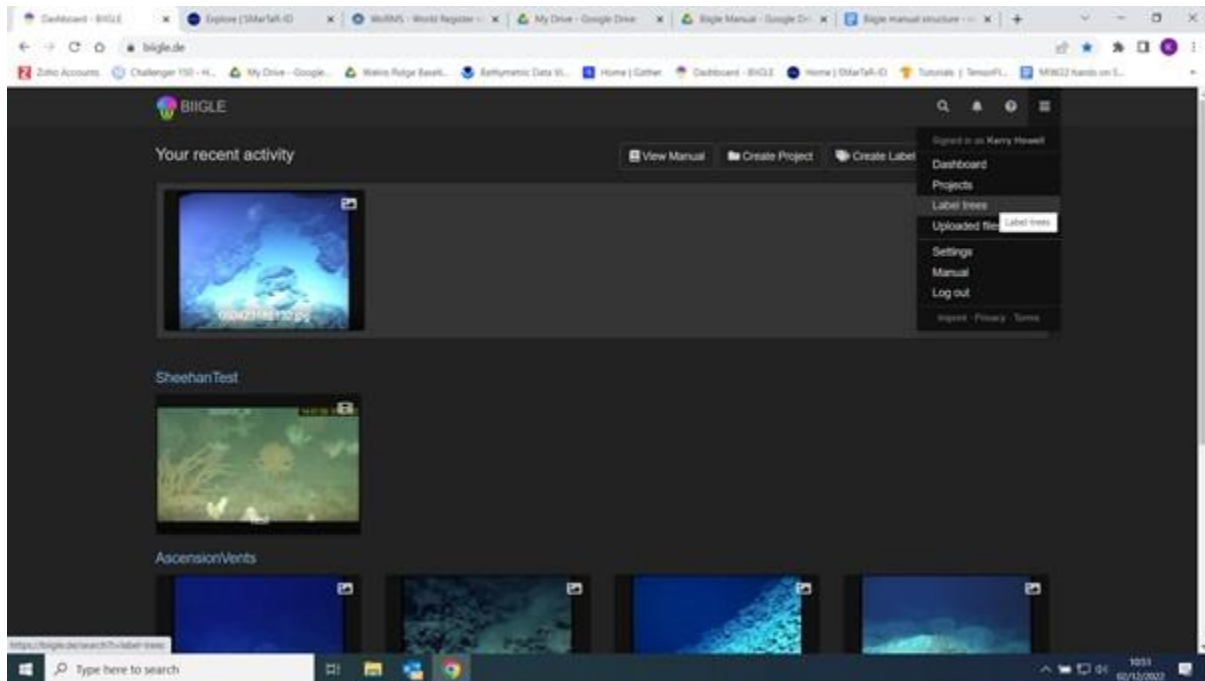
This will take you to the label trees screen where you can see publicly available label trees.
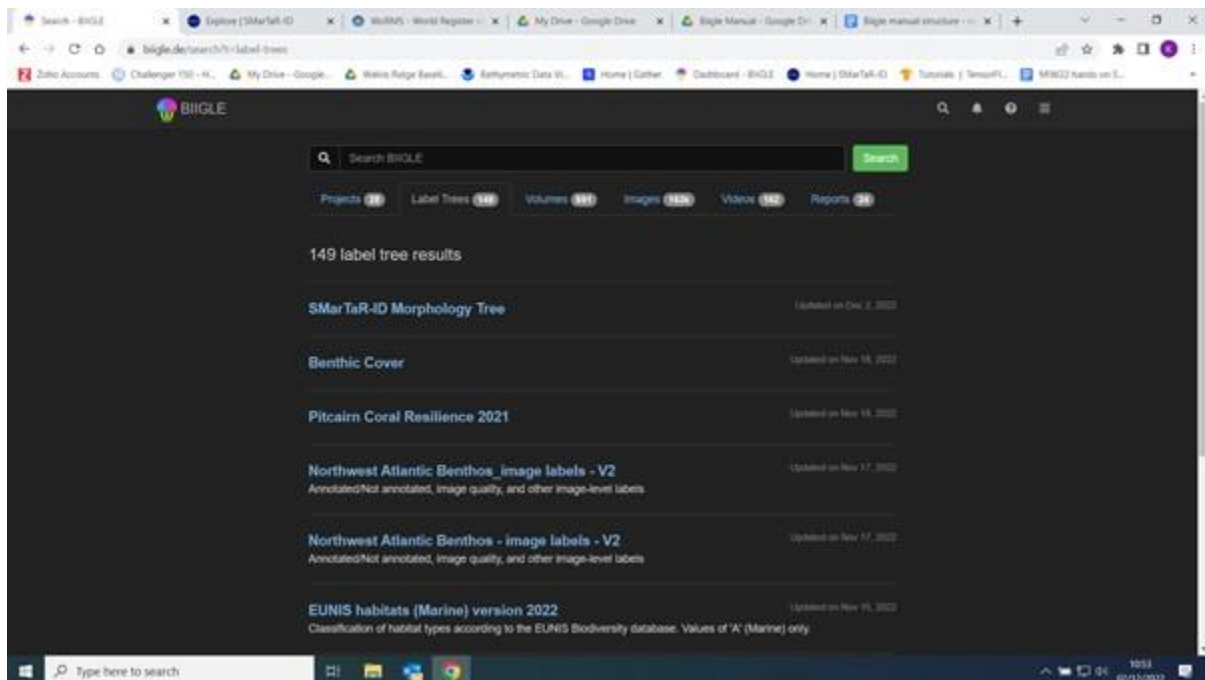
Search for a tree you want to use and click on it. Then in the settings drop down menu, choose 'fork'. Give it your own name, choose to keep it private or make public, and create.
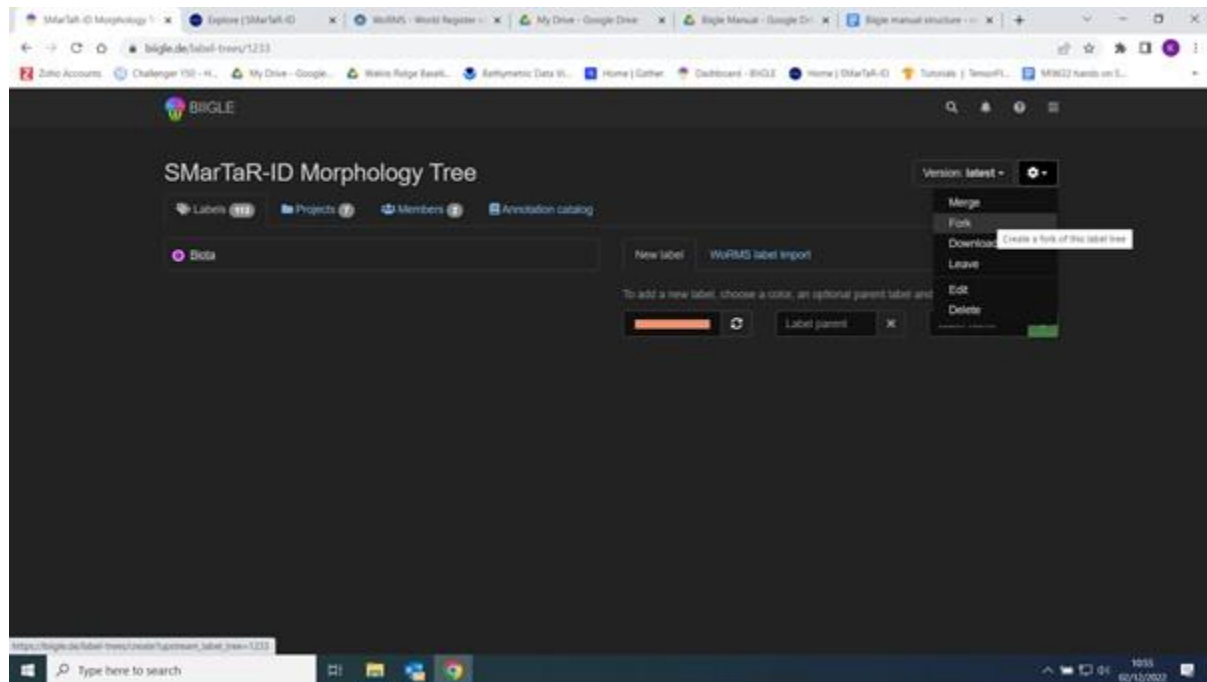
3.2.2 Creating your own tree from scratch

If you need to, you can create your own tree from scratch. From the dashboard, click on 'create label tree'. Name your tree, decide on its visibility and create. You can then add labels to your tree using the 'Label parent' / Label name' boxes on the left of the screen. However, if you are adding taxonomic names, we strongly advise you to use the 'WoRMS label import' tab, which will attach the WoRMS Aphia ID to the label such that you can retrieve taxonomic information from WoRMS on your BIIGLE outputs (World Register of Marine Species - https://www.marinespecies.org/index.php).
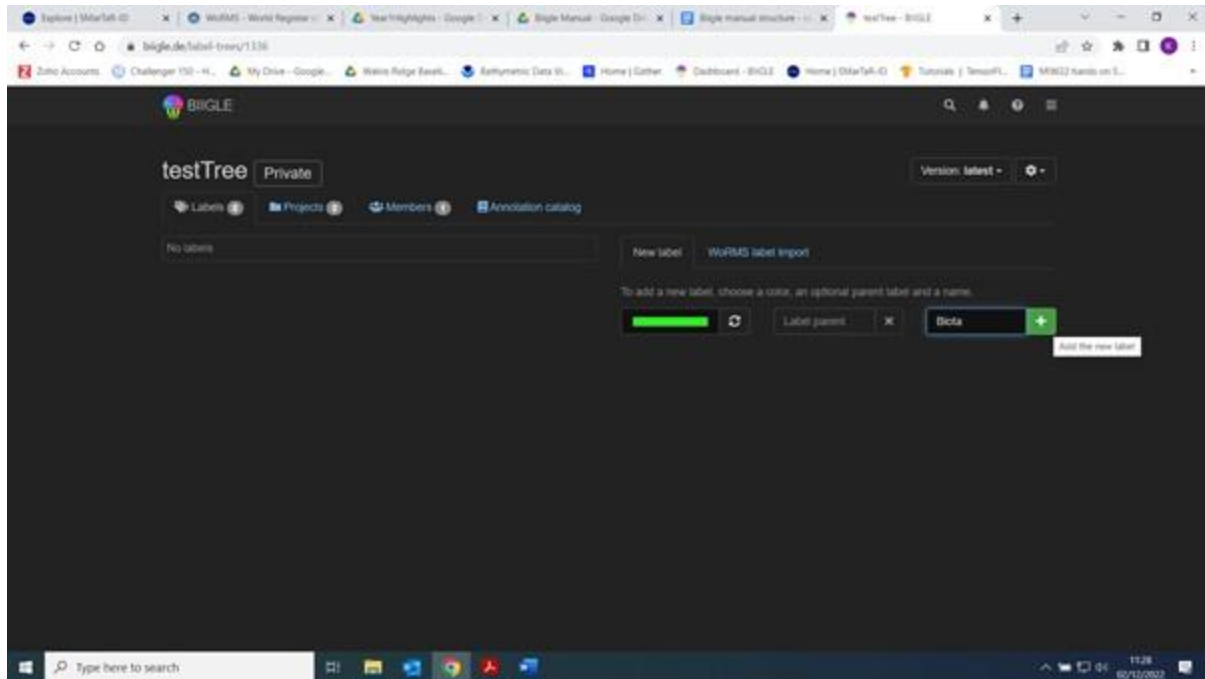
3.3 Attaching a label tree to your project

Once your tree exists you need to authorise projects to use it. To do this simply go back to the label tree screen, click on your new label tree and then click the projects tab. You then need to type your project name into the 'add authorised project' box, this will allow your project to use this tree. Next you need to add the tree to your project. To attach a tree to your project, simply go into your project, click on the 'label trees' tab, and then start typing the name of the label tree into the search box, it will offer you possible trees you can then click on to attach to your project.
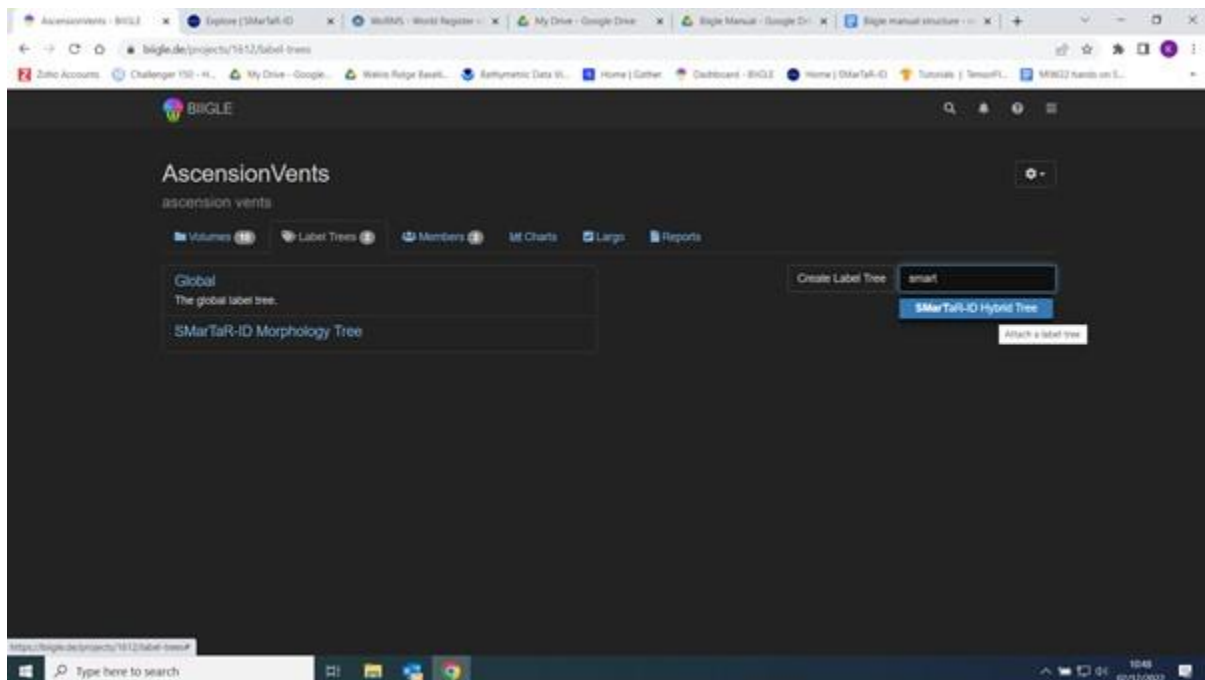
Now you are all set to start annotating 🙂

If you need to edit your label tree mid annotating, for example if you come across something not in your label tree, then you can do this by going back to the label tree screen, click on your new label tree and add additional labels using the 'add label' or 'WoRMS label import' tabs.

4. Make annotations

Before we go any further it is useful to just familiarise yourselves with the different tools / tabs available at the project and volume level.

In the project window of BIIGLE, you can:

- View image and video volumes uploaded to the project,
- View/edit label trees attached to the project,
- Add members who can view or edit annotations,
- Largo the annotation. The largo function allows you to view annotations of a specific label and re-assign if incorrect, or retrospective identification,
- Export report for the entire project (see manual for details for each report type).
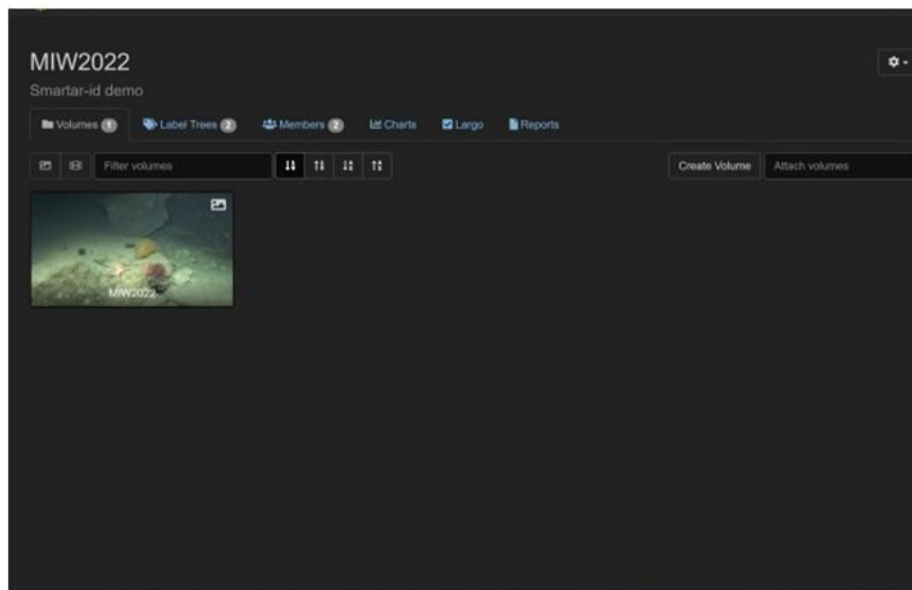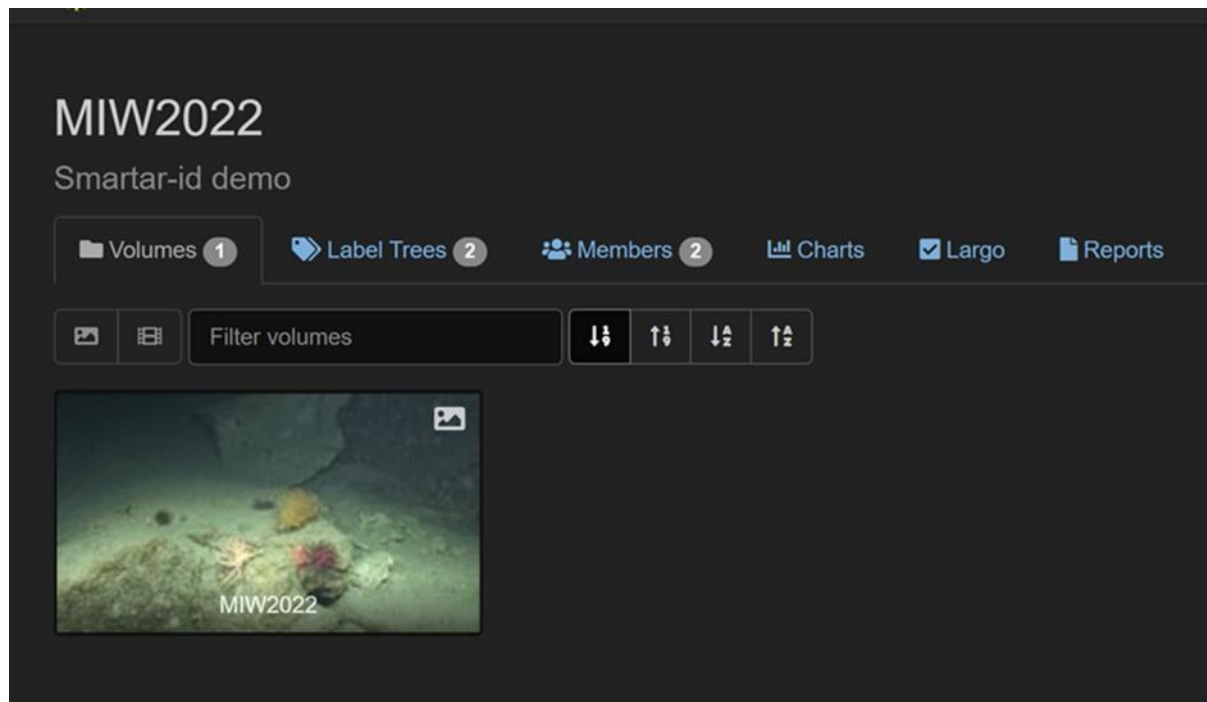
In the volume window of BIIGLE, you can:

Edit the volume

Add image/video label. You may use this to assign quality for example.

Filter, this could be file name etc.

Sort files

Largo the annotation. The largo function allows you to view annotations of a specific label and re-assign if incorrect, or retrospective identification.

Reports: Request report from the volume. Export report for the entire project (see manual for details for each report type).

Area: Using lasers, the area of the images can be calculated if you know the distance of lasers

MAIA: Create a MAIA job

4.1 Image annotation

To begin annotating you must select a project and volume to annotate. To view the projects, click on the three lines on the top-right corner of the webpage, and select "Projects" in the drop-down list, or just click on the projects tab. Projects will be listed on the page, and will consist of

the name, preview, description, and date of last update. Double-click on the project of interest to open the volumes for each project. Alternatively, click on the "Volumes" button to open a window containing all the volumes available for access.
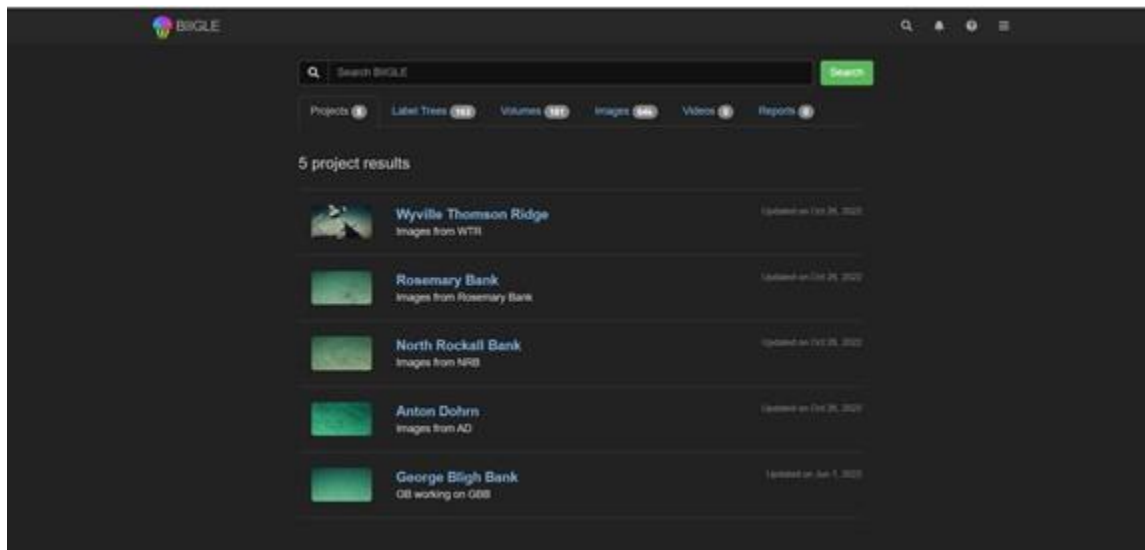
Click on the volume of interest to open it and reveal all the images contained in it, then click on an image to start annotating. The image view in BIIGLE looks something like this:
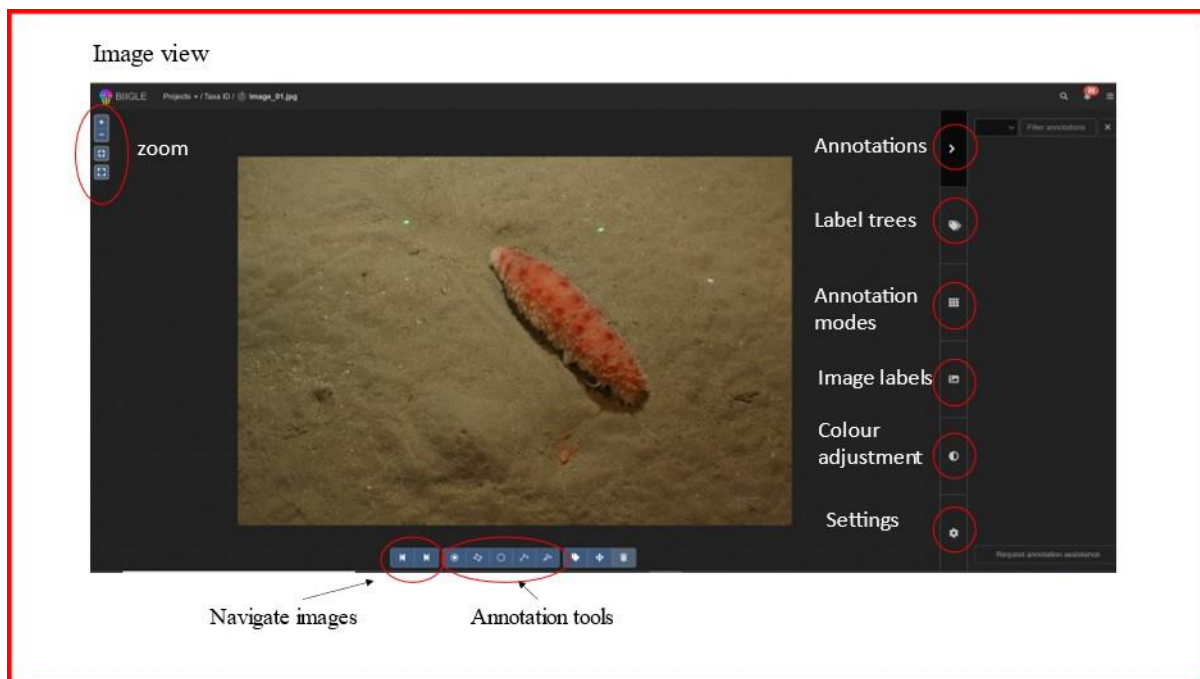
First select the appropriate tree from the "label tree" section of the panel on the right. To choose a label, you can either:

- Find the label in the label tree

- Use the search function, you can search for any part of a label, including text or numbers
- Use shortcut keys for your favourite list (max 9)

To find the label in the label tree click on the coloured dot on the side of the label to show lower taxonomic or morphological levels. Select the label of interest and the name will appear in a box on the bottom-right corner of the screen.
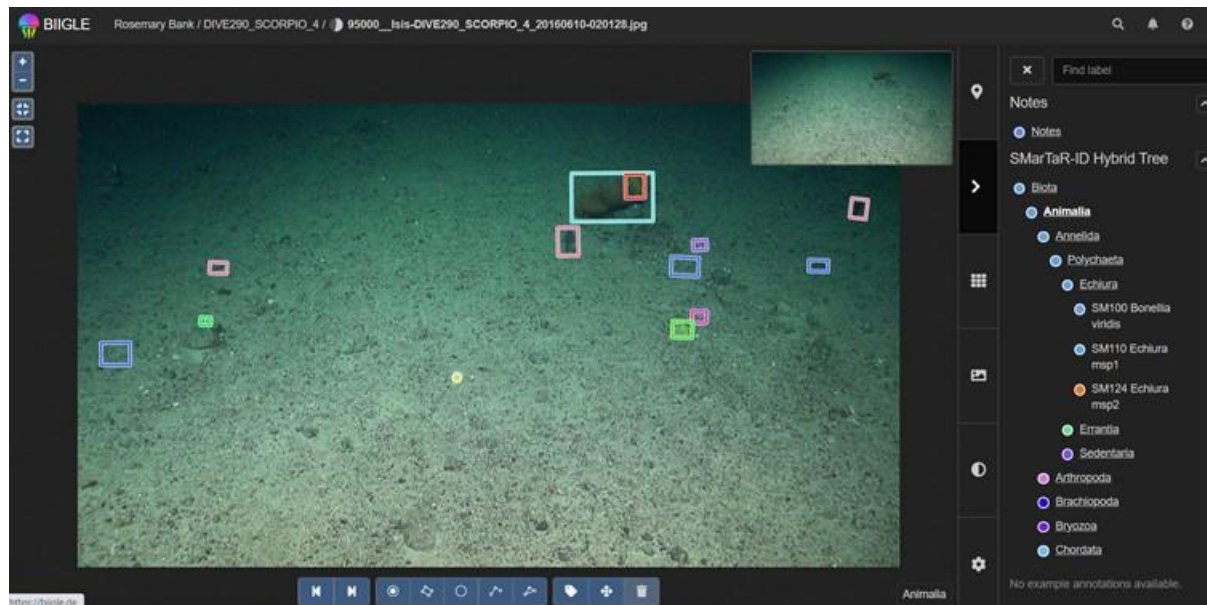
Now that the label is selected, click on the "draw a square" button on the bottom, and trace a square around the object of interest by clicking on its upper- and lower-middle points and extend the box as needed. The box will be of the same colour of the chosen label. The shape of the box can be modified by unselecting the "draw a square" button, clicking on a corner of the box (which will turn red) and dragging it to reshape it. After selecting the box, it is also possible to delete the annotation by clicking on the bin icon in the bottom panel.

Tips: * For more specific details on annotations, refer to annotation section on BIIGLE website (https://biigle.de/manual/tutorials/annotations/creating-annotations) * Use shortcut keys for annotation tools, this will speed up annotation sessions, these used in combination with the favourite keys (1-9) can significantly speed up annotation
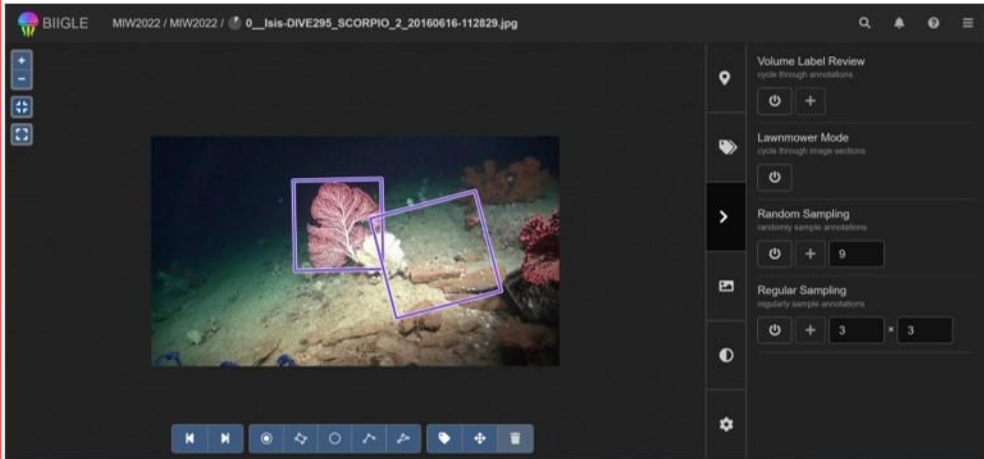
Annotation type keyboard shortcuts

When using annotations tools, things to consider:

- If you want to use your annotation for training AI models try to use the rectangle tool only and draw your rectangles in the same orientation as the image as in the example above. Try to keep your rectangle tight around your animal (e.g. do not include lots of background in your annotation). If you need to use another shape (e.g. point) that's OK but a second conversion step will be needed outside of BIIGLE to use it for onward AI model   training.

- size of annotation tool, e.g. if using the point, try and annotate the same part of taxa each time as this will aid viewing in Largo mode.
- If annotating for area use the polygon tool. You can either draw this manually with the mouse/stylus (with compatible computer/IPad) or using the magic wand tool.

The annotation modes section has some useful options that may also help you.

4.2 Video annotation

To begin annotating you must select a project and volume to annotate, then select a video file (if you have organised your videos in multiple videos per volume). The video view in BIIGLE looks something like this:
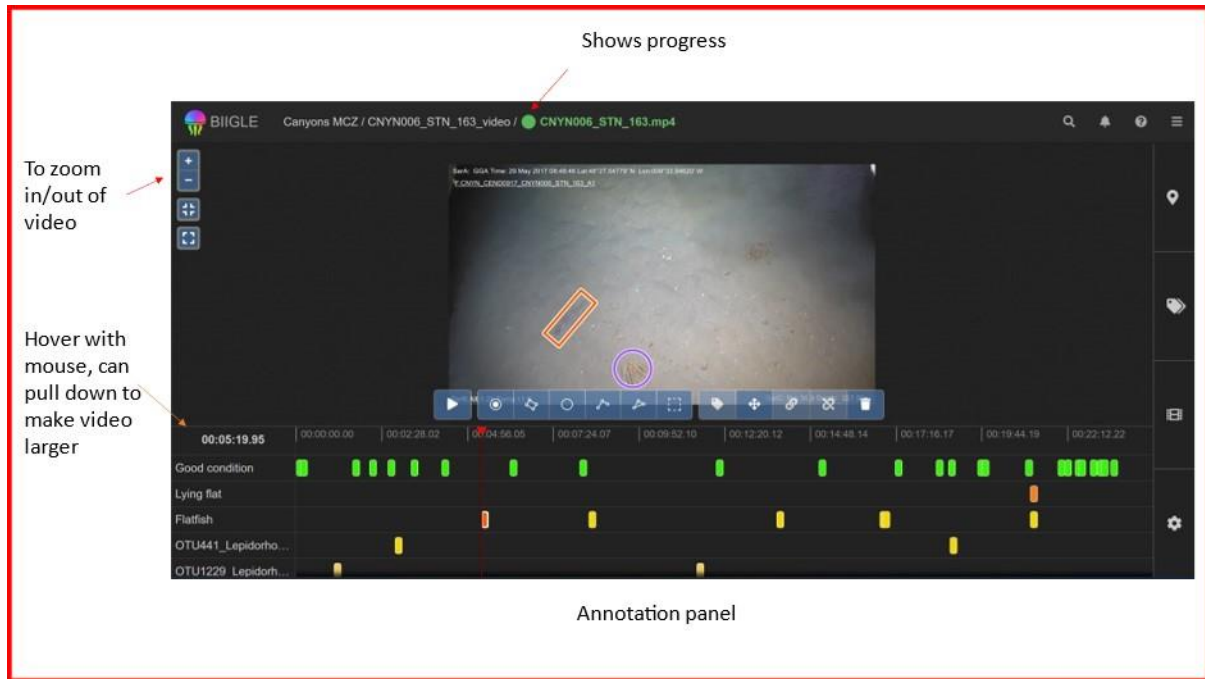
Annotation tools for video are mostly the same as that used for image annotation (see above), the exception is that you need to click the tick to accept the annotation before moving on to the next annotation. If this is not done, the annotation will not be complete.

An additional annotation tool, whole frame annotations are also available. This allows you to in effect, segment the video. This could be to categorise by quality, habitat or even to record when an animal/object is in frame.
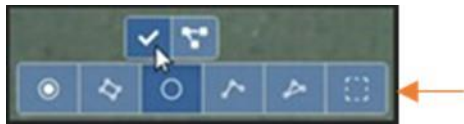
The settings tab has some useful options that may also help you.

Setting tab

Can adjust the visibility of annotations

Can adjust playback speed of video

Allows you to see the annotation label when hover over
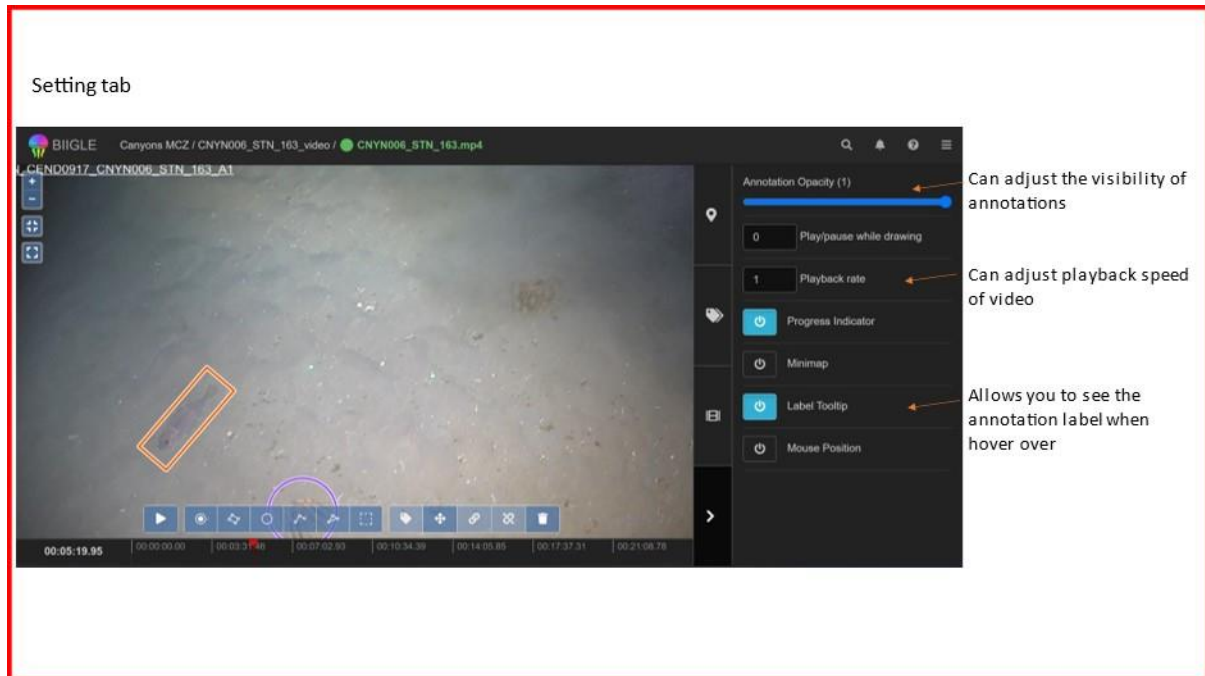
4.3 Setting an annotation session

Annotation sessions can be set up for each volume allowing you to set a session where annotators cannot see each other's annotations. These are ideal for training sessions and quality assurance to check for inter-analyst variation.

At the project level go to the Member tab and add all participants who will undertake the annotations.
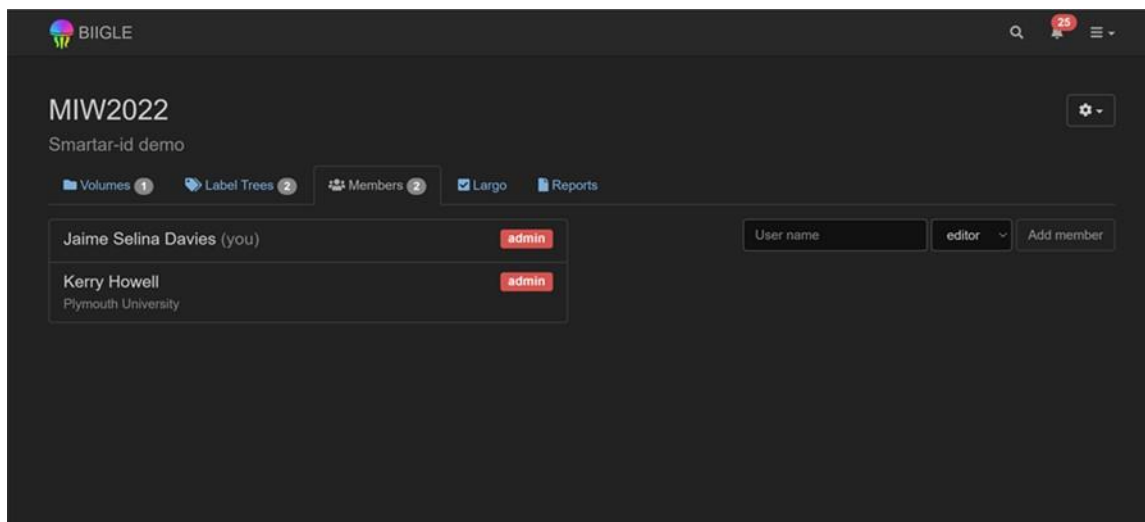
Then set up the annotation session. Go to volumes tab – and click on the volume thumbnail to take you to the volume view:
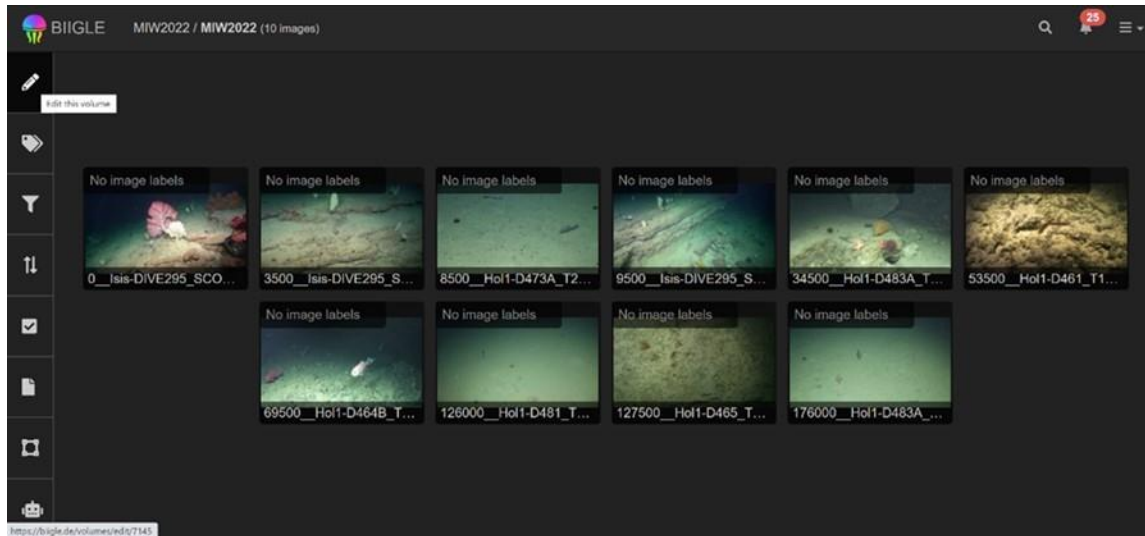
Click on the edit button (pencil) – then edit annotation session

Fill in the annotation session box, all must be filled in (example below). Add the users names (new users box) and when you click on them they will add to the right hand box 'users'. When you have added everyone, press create.

It will display like this:

If you want to edit the session, i.e. remove the annotation session, click on it. Then you can delete it or just change the date to earlier than the date and that will mean it has expired and all users will be able to see each other's annotations. Sadly deleting the annotation session does not delete the annotations, which is a bit annoying when using for training.

5. Quality control and the largo tool

The largo tool is just about the most useful tool you can imagine. It can be used to:

- Check annotations for consistency
- Separate coarsely classified groups in lower taxonomic / morphological groups

You can use the largo tool at the project or volume level. In essence it allows you to view all images with the same label at once and make changes to the labels.

At the project level simply click on the Largo tab:
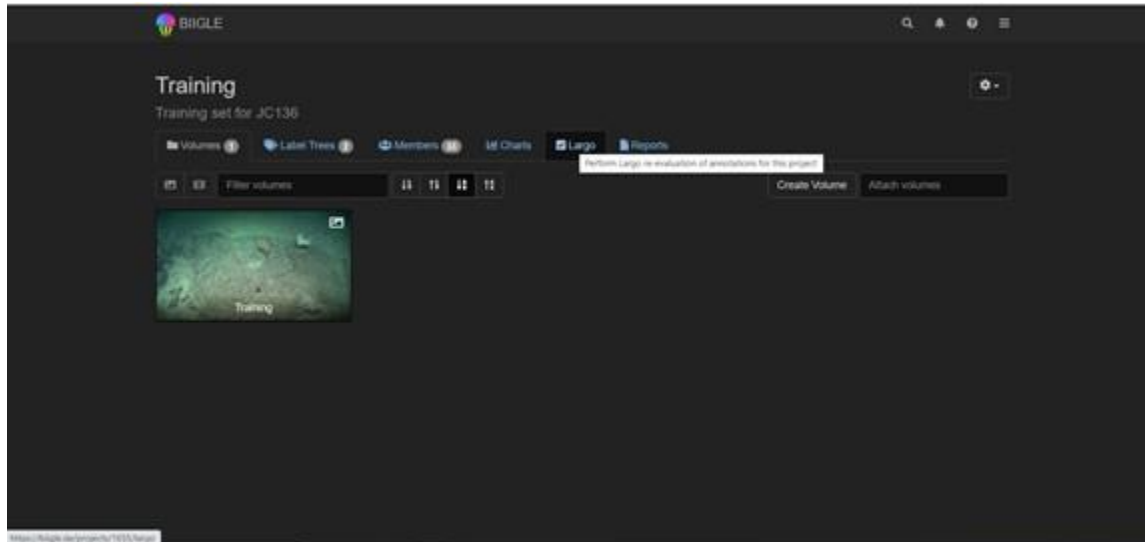
At the volume level the tab is here:
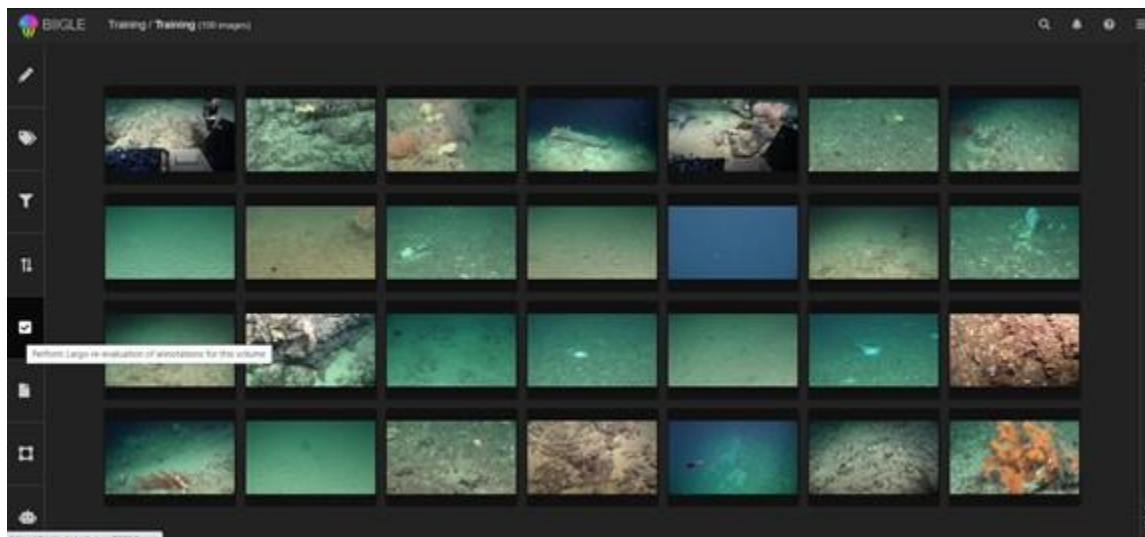
In both cases you will then need to select a label from your tree. In the example here (at the project level) we have selected the 'Echinoderms' label, and you can see that anything that has been labelled 'Echinoderms' in the project is now shown on the screen. You can scroll through them using the scroll bar on the left.
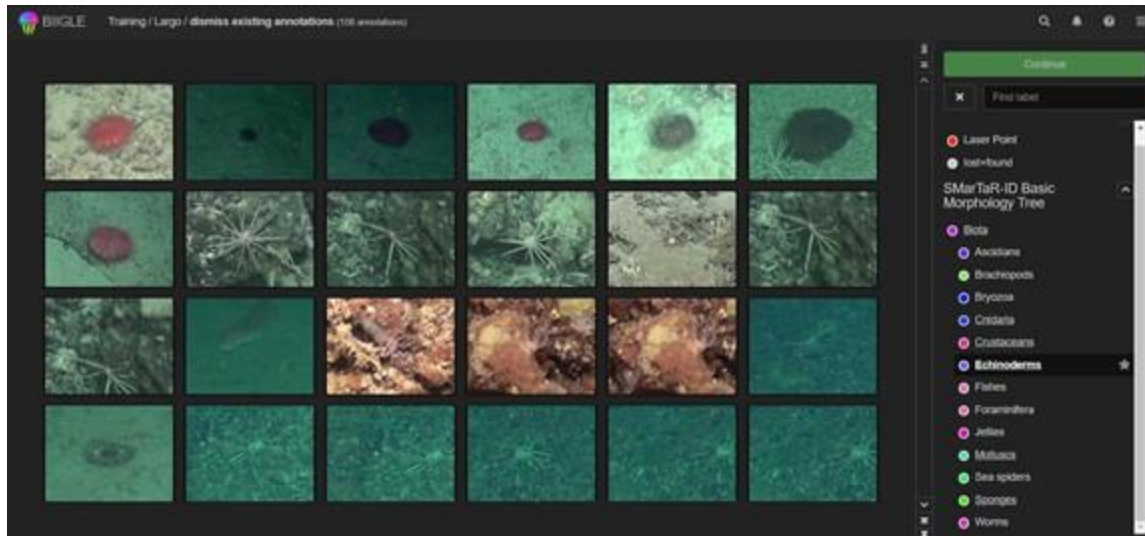
If you look carefully you can see that someone has labelled a cup coral as an echinoderm (image in the bottom right corner). This is clearly wrong and so this image needs to be corrected. To do this click on the image, this will make a cross appear on it. Then click continue, which will take you to the relabelling step.
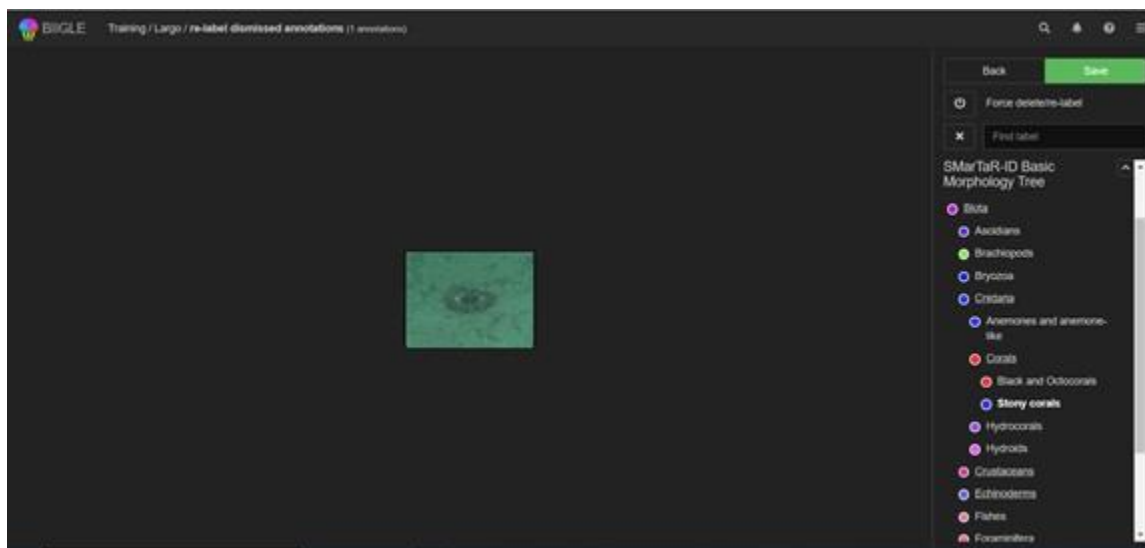
Now you need to select a new label for your animal. We will relabel this as a stony coral. First select the correct label from your label tree, then click on the image to assign the new label. Then you need to decide whether to delete the old label and replace it with the new label, or just give the animal two labels. Most of the time you will want to delete the old label and replace it with the new label. To do this you need to select the 'force delete/re-label option on the top left of the screen.
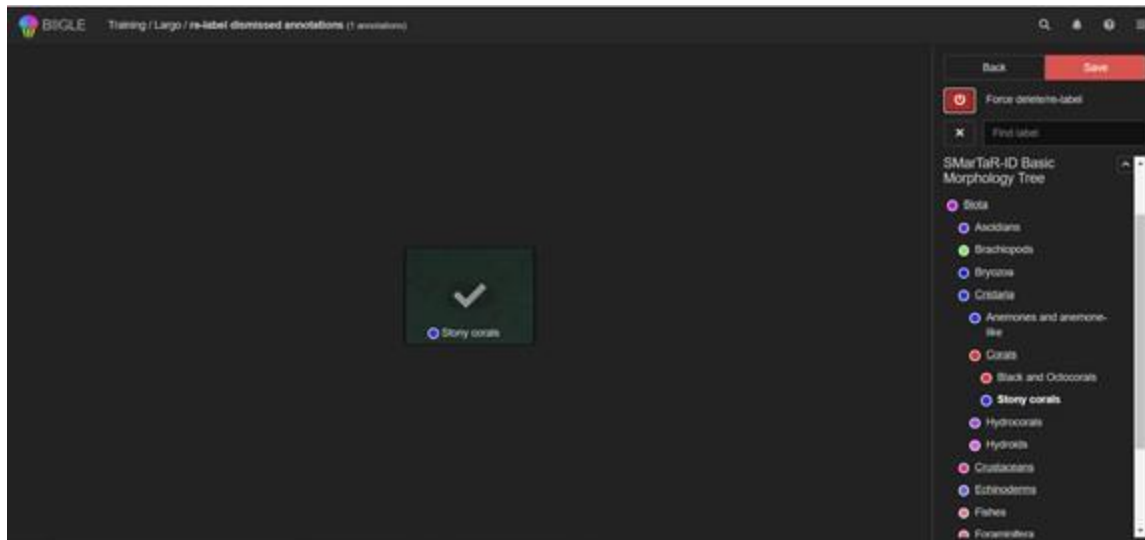
You then need to hit the red save button to make the change.

You can use largo for multiple changes at once.
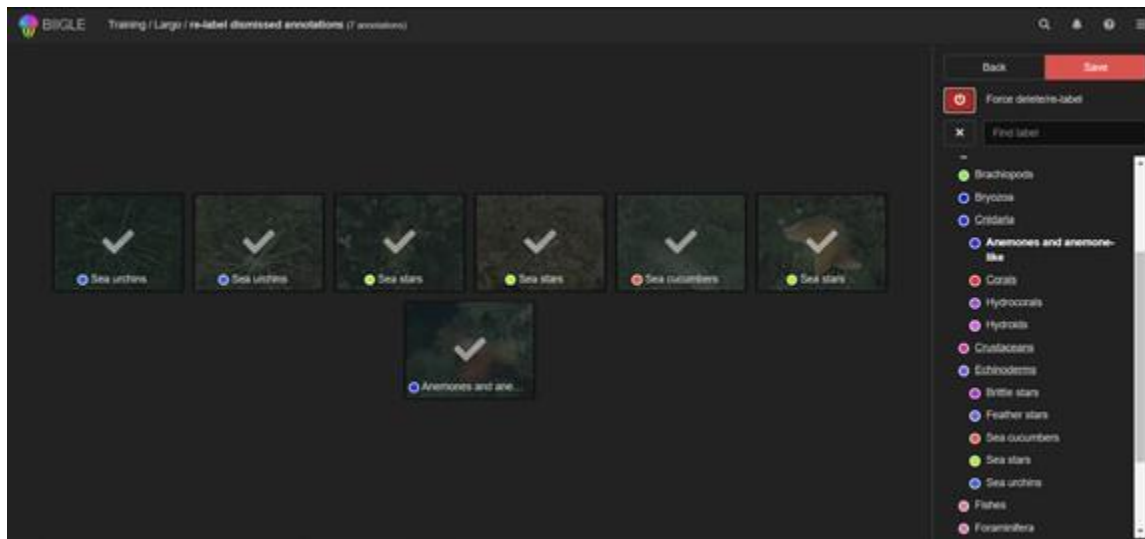
You can also call labels on one label tree and relabel them to a different label tree.
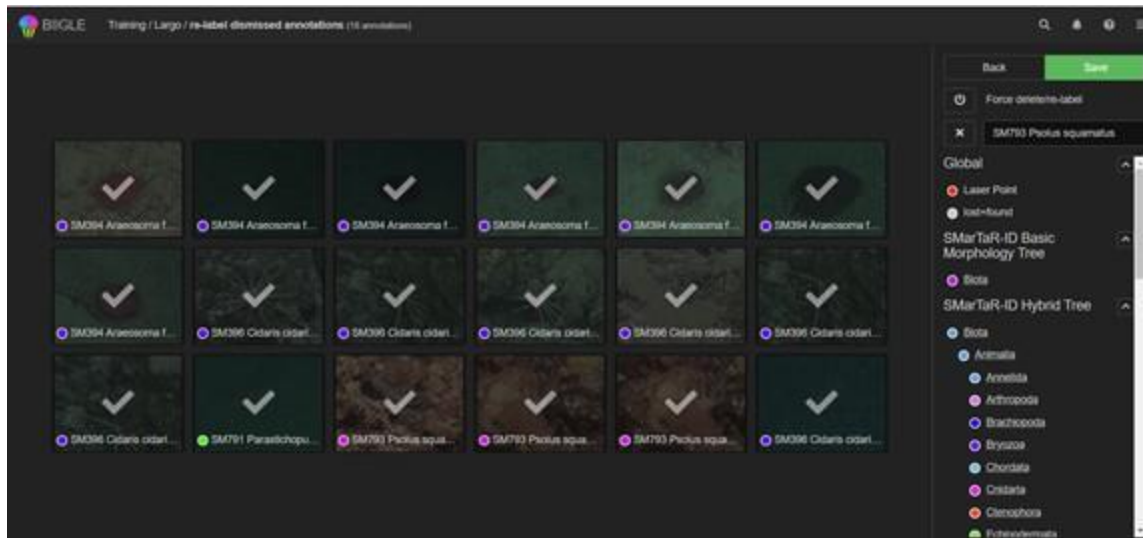
Sometimes you may want to review the animal in its original image to help you to decide if / what to re-label it as. To do this simply click on the top right hand corner of the thumbnail in the original largo screen, and this will open the original annotated image in a new window.
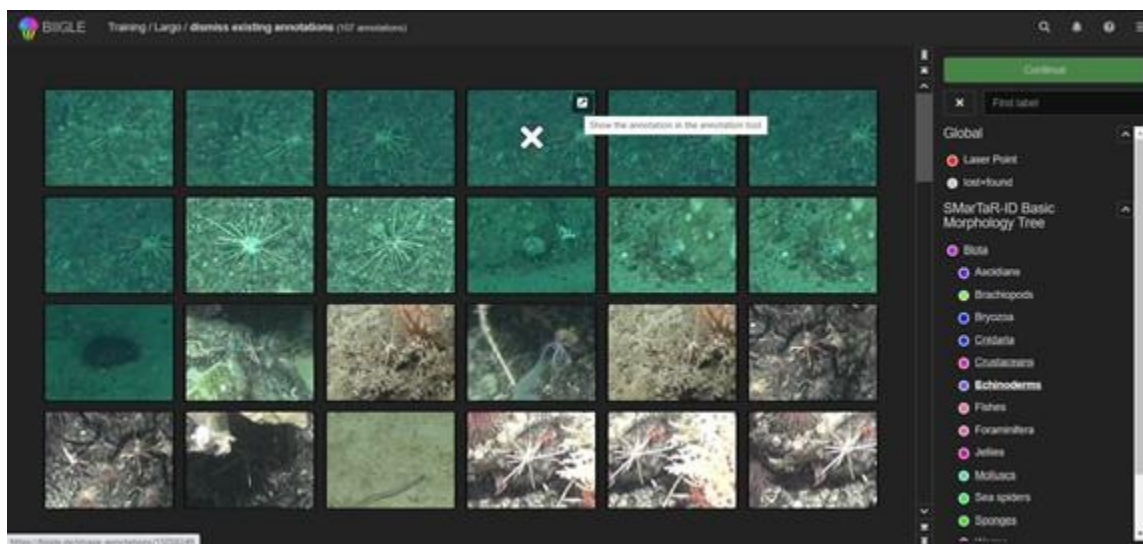
5.1 Suggested best practice in annotation

We have found through much experience that the way to get the most consistent annotations in the shortest time is to use a two-pass approach to annotation. Anyone familiar with sorting grab samples will understand where this comes from. We recommend you make a first pass through the image / video using high level taxonomy / morphology. We use the public SMarTaR-ID Basic Morphology Tree to first classify things to a high level standardised morphology. You may wish to make your own tree or just discipline yourself to only using high levels of your taxonomy tree.

Once we have annotated all our images / video to this level we use the Largo tool to then classify all images within each high level group to a lower level of taxonomy / morphology. Largo

allows you to see all images classified under the same high level group at once. It is then a short and more consistent job to separate the images into lower level groups.

SMarTaR-ID basic morphology tree			
Biota			
	Ascidians		
	Brachiopods		
	Bryozoa		
	Cnidaria		
		Anemones and anemone-like	
		Corals	
			Black and Octocorals
			Stony corals
		Hydroids	
		Hydrocorals	
	Crustaceans		
		Amphipods	
		Barnacles	
		Chirostylids	
		Crabs	
		Isopods	
		Lobsters	Squat lobsters
			True lobsters
		Prawns / Shrimps / Mysids	
	Echinoderms	Sea stars	
		Brittle stars	
		Feather stars	
		Sea urchins	
		Sea cucumbers	
	Fishes		
	Foraminifera		
	Jellies		
	Molluscs	Bivalves	
		Cephalopods	
		Gastropods	
		Chitons	
		Scaphopods	
	Sea spiders		
	Sponges	Crust-like	
		Cup-like	
		Erect	
		Massive	
	Worms		

6. Downloading data and reformatting

As with any image analysis software, the results of the image analysis conducted on BIIGLE is stored there until it is retrieved by the user in order to analyse it further.

The data is exported via BIIGLE Reports for a single volume or the entire project.

To export a report, go find it in the aptly named 'Reports' tab at the root of a volume on the right handside:
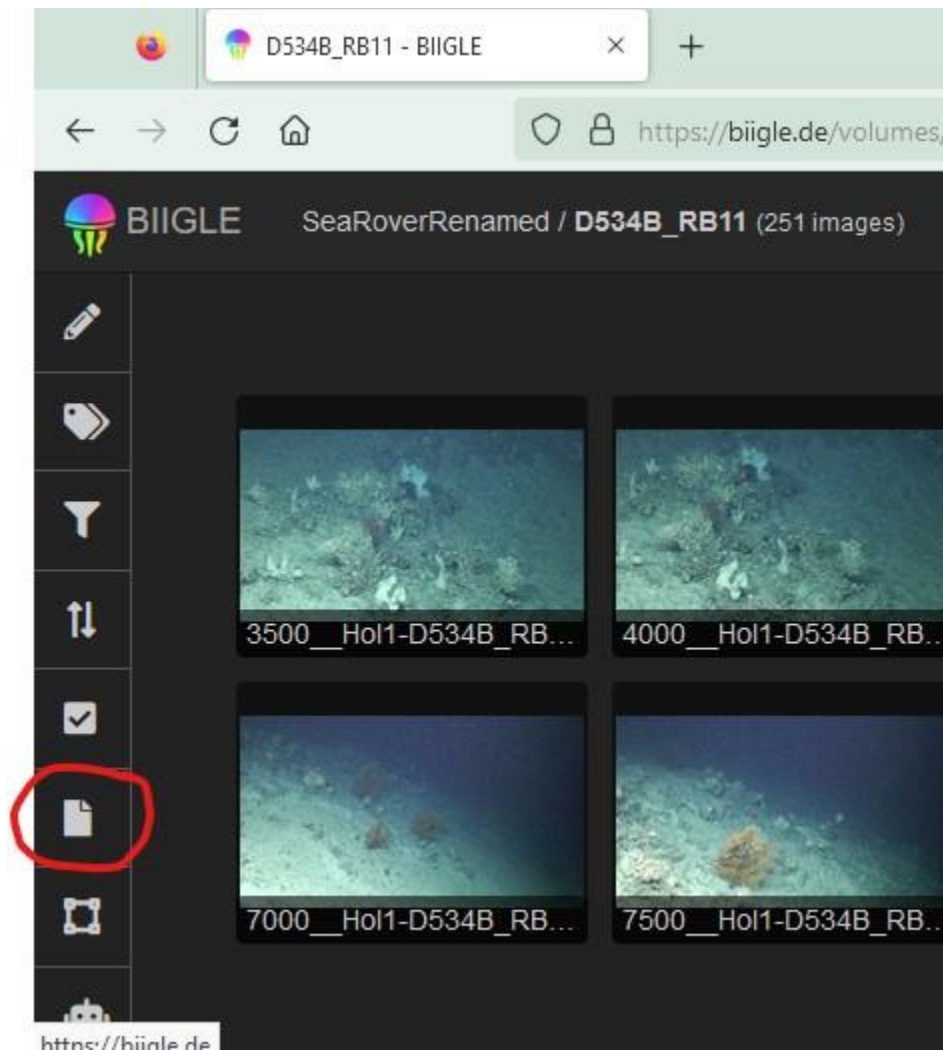
The same is possible for the entire project
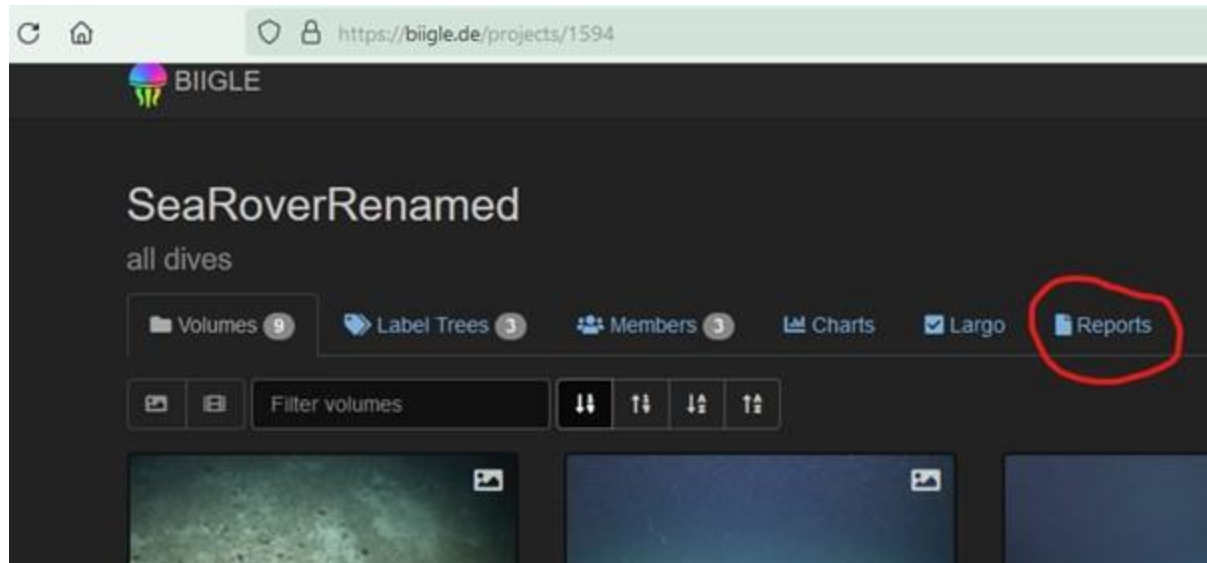
This takes you to a menu where you can choose the report type, and several other options to organise the data you are downloading.

6.1 Get and use a report

BIIGLE has many report formats displaying different types of information. The list of reports and specificities of each types are listed here: https://biigle.de/manual/tutorials/reports/reports-schema#annotation-abundance-report

The reports can contain the image annotation, the image labels, the metadata (be it geographical or concerning the images' characteristics). They can be nested by user, annotation tree or other.

6.1.1 Image annotation report - CSV variant

For ecological analysis where you need to know the exact list of species present within each image in a volume or a project, the CSV format is the most practical. To get a CSV report, just select that option in the 'Report variant' drop down menu and leave the other boxes unticked.

R code to combine reports and format their content into a species x image matrix can be found along with instructions here: https://github.com/DeepSeaCRU/BIIGLE-resources.git (in the species-matrices folder)

The results of the SMarTaR-ID reports.Rmd are excel spreadsheets that can be used in R, python or any other software that takes tables as input.

It also gives the possibility to re-arrange the labels hierarchy (the upper taxonomic levels) to make it easily available for later analysis.

	A	C	D	G	H	O
1	filename	OTU339	OTU23	OTU142	OTU11	OTU205
2	level1	Biota	Biota	Biota	Biota	Biota
3	level2	Animalia	Animalia	Animalia	Animalia	Animalia
4	level3	Arthropoc	Cnidaria	Arthropoda	Arthropoda	Arthropoc
5	level4	Crustacea	Anthozoa	Crustacea	Crustacea	Crustacea
6	level5	Multicrust	Hexacoral	Multicrustacea	Multicrustacea	Multicrust
7	level6	Malacostr	Actiniaria	Malacostraca	Malacostraca	Malacostr
8	level7	Eumalaco:	Halcampo	Eumalacostraca	Eumalacostraca	Eumalaco:
9	level8	Eucarida	OTU	Peracarida	Eucarida	Eucarida
10	level9	Decapoda	OTU	Mysida	Decapoda	Decapoda
11	M116_13331848_13109013315679.jpg	1	1	0	0	0
12	M116_13331848_13109013316679.jpg	1	1	0	0	0
13	M116_13331848_13109013317679.jpg	1	5	1	1	0
14	M116_13331848_13109013318679.jpg	1	11	1	0	0
15	M116_13331848_13109013319679.jpg	1	8	0	0	0

7. Using the BIIGLE API

7.1 Intro to BIIGLE's Application Programming Interface

API stands for application programming interface, which is a set of definitions and protocols for building and integrating application software. It provides a standardised way for users to interact with databases and websites through command lines and lets them efficiently manipulate data.

The API works by sending requests to a server in the form of a command line rather than a point and click interface you normally use on a website. In practice it gives you a dialogue box in which you can input commands to interact with BIIGLE and script your action so it is faster, reproducible and can be automated. It is less intuitive because you have to write your request correctly in order to be correctly understood by the API and do what you intended. It is also faster as each command can perform a complex task that would require many clicks. This leaves less room for error than the manual interface where you may lose track of what you have and have not done, or click the wrong button without even realising it. This means that instructions can be written down and applied to the letter with no other input from you. You can, for example, request the information attached to a specific object, modify them or add new ones. For the habitat mappers out there, this is akin to the console you open in ArcGIS to enter scripted instructions rather than use the drop-down menu and the GUI of each tool. In effect, you are pointing a command that corresponds to an action towards a specific object, the action gets performed and the expected output is sent back to you.

An important note here is that to give the instructions to BIIGLE you need to identify the objects you want to modify. BIIGLE does not use the names of volumes and images that you are using. Probably because these can be quite long, have special characters and can be duplicated across users and projects. Usually, new objects (project, volume, image, label tree, label, annotations, etc) receive a numerical ID that is unique in the whole BIIGLE ecosystem. You need to know these IDs before you can form the right API requests that BIIGLE understands. All the available commands in the BIIGLE API are listed here: https://biigle.de/doc/api/index.html

Note that 'listed' does not mean 'explained'. There is barely a description of what each command does, what input is expected and, succinctly, what the output will be; and no explanation of individual arguments needed in the body of each function. Some changes are regularly implemented in the commands, so it is worth taking a look there regularly in case your commands become obsolete.

7.2 Basic requests
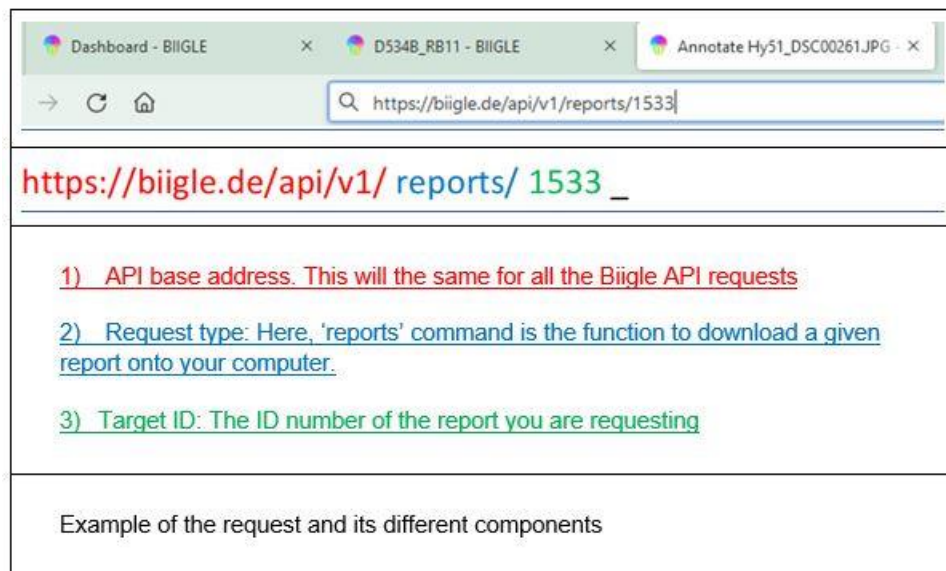
To make a request, the task and targets must be combined to form the command. The API requests structure works a bit like a windows explorer pathway or a website address with hierarchical elements pasted together from left to right to form a sort of sentence.

Here are simple examples:

7.2.1 Requesting a BIIGLE report

When requesting BIIGLE reports you receive an email with a link to click, that is in effect an API request to download the report created.

Specifically, if it requests the report of number 1533, the request code will be: https://biigle.de/api/v1/reports/1533

Example of the request and its different components
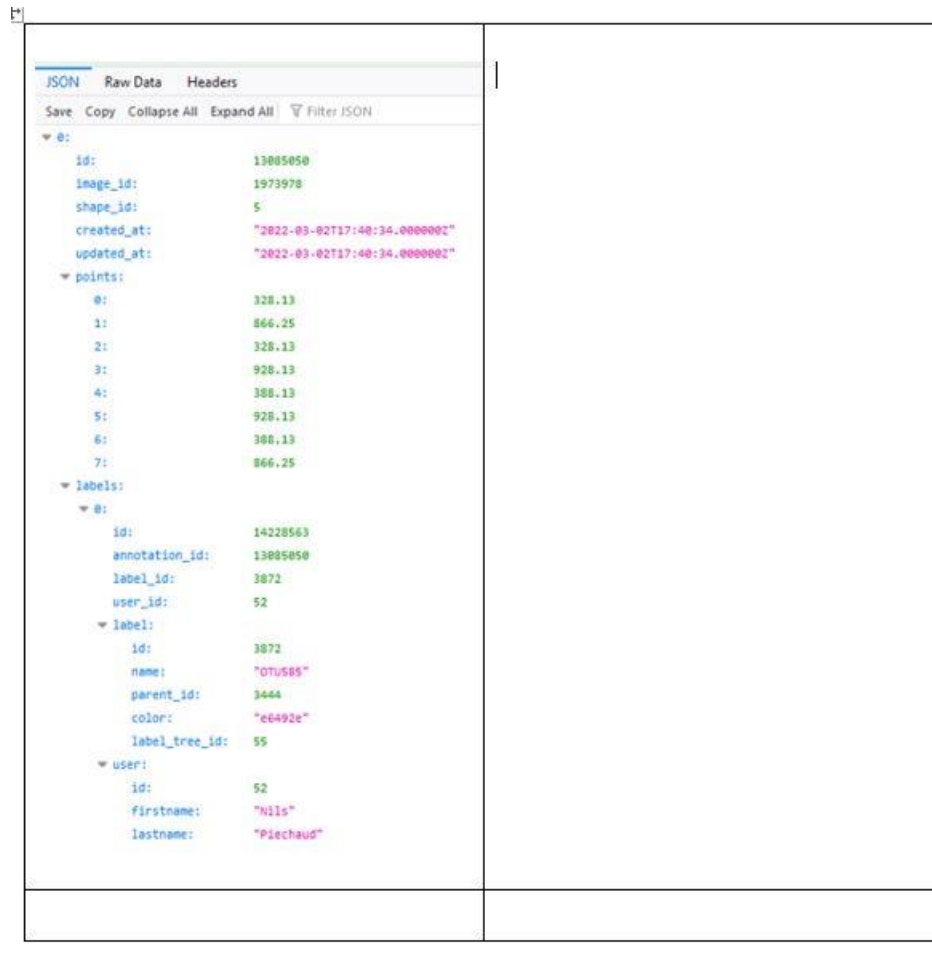
The easiest way to use that command is to paste it into your web browser's (Chrome, Firefox, Opera, etc) address bar and it will send the request to the BIIGLE server, which will in turn do as instructed, in this case, prompt the download of report number 1533. There is more going on behind the scenes, for example, this will only work if you have been granted access to the report you are requesting (granted to you by the creator of the project). So, in your web browser, you need to be logged onto BIIGLE for the download to start, and your account needs to have been granted access to this volume.

7.2.2 Requesting the ID numbers of BIIGLE objects

Another example is for requesting some open label tree (gets a page displaying the tree in JSON format): https://biigle.de/api/v1/label-trees/55

So you can explore all the BIIGLE objects that way and get all the pretty stuff displayed in green and pink numbers so you can make sense of it, maybe. So if you pick one project that has a specific ID: 1129

First you might want to know what volumes there are in there: https://biigle.de/api/v1/projects/1129/volumes

If you want to display the BIIGLE ID of the images you see in a specific volume: https://biigle.de/api/v1/volumes/6063/files

Every image added to BIIGLE gets unique Image ID. Not the ones on your remote storage server (not the ones that you have linked to a volume on BIIGLE).

To get some info on a specific image: https://biigle.de/api/v1/images/1973978

and what annotations it has: https://biigle.de/api/v1/images/1973978/annotations

The results are displayed by the browser and give you the information you need to interact with the browser.

What you can do with the API:

- Automatically annotate
- Upload annotation made by Machine Learning algorithm outside of BIIGLE
- If you have some data that was annotated in another software, you can add it to the corresponding images on BIIGLE
- You duplicate your volume. Take all or part of the annotation from one volume and add them to another without having to re-annotate the images
- You can also selectively transfer part of your annotations or part of your images to another volume
- This means you can back-up your volumes at a given time and bring it back to BIIGLE at a specific stage if you want to
- You can also easily gather data about your project. How many images, how many volume, the size of each volume, the amount of annotation by each user… etc
- You can also gather info about label trees and modify/update it

These examples are just pulling information out of BIIGLE, not actually inputting anything on the server. Other types of request exist with a "body" of data that is uploaded when the request is sent. These are better done through a terminal through R or Python.

7.3 Using the API with R and Python

You guessed it, if there are a lot of instructions to send, manually typing them one by one is not going to be feasible. Thanks to interfaces with your favourite coding languages, you can script your interaction with the API the same way you script other repetitive tasks. This comes down to manually "crafting" your command by gluing arguments together and then, sending them to the API directly through the console of your integrated development environment (IDE). This makes the interactions much more efficient as you program what commands are sent, which is considerably faster. This is relevant when many commands must be sent. Typically, when inputting data into individual volumes or images.

To add an annotation to each image in volume, a single command must be sent for each image. A volume often hosts hundreds if not thousands of images. Hence the need to automate the process to speed it up and avoid errors.

So how do you use BIIGLE in R or Python?

R and Python are the most commonly used programming languages in the community of BIIGLE Users. They are not the only languages that allow users to interact with an API but other options will not be covered here.

What you need to get started:

- A viable installation of R and Python
- An integrated development environment (IDE) is optional but recommended
- Libraries that let you send requests to an API in R (httr) and python (requests): httr and requests in R and Python are examples
- Your credentials

It is good practice to have a recent if not the latest version of R installed on your computer to ensure compatibility between all code and libraries. This is less straightforward with Python where recent and old versions can co-exist on parallel update paths but try to install version 3.7.10 or later.

An IDE is strongly recommended. Rstudio and Jupyter in Anaconda are good options for R and Python respectively, but if there is an environment you are familiar with, then that is really all you need.

Base R and Python are not able to communicate with an API so you will need to import new packages or libraries to get the functions that can. In R, the code presented here uses "httr" (https://cran.r-project.org/web/packages/httr/index.html) and Python Code uses "requests" https://pypi.org/project/requests/. Make sure they are installed and up to date before starting.

You will need your credentials to tell BIIGLE who is sending these requests so it can check if you have access to the object you want to interact with. Your email address acts as your login. You will need your token that you can get here: https://biigle.de/settings/tokens. Once you have your token, store it in your IDE so it can be attached to your request each time your send one:

Also, try some easy tasks first to make sure you understand the whole workflow correctly as you can easily make large scale errors when using the API.

Below are basic examples of API use:

```python
# Python code to get basic information about an image on Biigle

 import requests
from requests.auth import HTTPBasicAuth
# credentials
login = "name@place.ac.uk"
token = "token"
# your request url:
url = "https://biigle.de/api/v1/images/934788"
# send the request to the API
request = requests.get('https://biigle.de/api/v1/images/934788',
auth=(login,token))
print(request.text)
```

The same example in R keeps the same structure but, of course the commands are different.

```r
# R code to get basic information about an image on BIIGLE
```

```
library(httr)
# put your credentials here:
login <- "name@place.ac.uk"
token <- "token"
# enter the URL
url <- "https://biigle.de/api/v1/images/934788"
GET(url, authenticate(login, token) ) -> request
content(request, as = "text", encoding = "UTF-8")
```

7.3.1 Accessing label tree information

BIIGLE takes its annotation label from a label tree that has to be manually created (see section 3.2 Make label trees).

In this label tree, each label is given a specific ID generated by BIIGLE. So with this ID number, BIIGLE understands how to link an annotation to a given tree and label in the tree.

The tree can be downloaded by following a link with the relevant number. This number will appear in your browser's URL bar when you access the Label tree.

For example, to get tree number 1, follow: *https://biigle.de/label-trees/1*

Replace the number at the end of the URL to get another tree. This will provide a JSON table or csv spreadsheet.

Note that you must have access to the tree to see its ID number and download it.

Here is what it should look like:

	id	name	parent_id	color	label_tree_id	source_id	label_source_id
1	3365	Biota	NA	5e0d5b	55	1	1
2	3366	Animalia	3365	5e0d5b	55	2	1
3	3372	Nemertea	3366	9d49f2	55	152391	1
4	3374	OTU384	3372	70e667	55	NA	NA
5	3376	Phoronida	3366	5ef23d	55	1789	1

The ID here will be needed later to upload annotations with a given name to BIIGLE

7.4 Uploading annotations to BIIGLE

A more significant sort of task you will perform with the API are those that create information. Most importantly of all, creating annotations.

This task is usually the most time consuming and its automation can save a lot of time if the necessary information about what animals are present in an image and where they are within it already exists in a list of annotations stored in a table on your computer.

This can be from one or more other BIIGLE volumes, from another annotation software, or may have been generated using Artificial Intelligence approaches. The API allows you to create BIIGLE annotations of corresponding parameters with the information stored in the table, which is much faster than the manual point and click input.

However, it will need to contain the following information required by the API

What you need to make annotations:

- **Volume ID**
    – The Volume ID is specific to the volume you want to modify. You can use the API request shown above or navigate to the volume itself and see the ID in the address bar of your browser.
- **Image ID**
    – The Image ID is attributed to an image in a volume. You will need to get them from that volume and the same image in different volumes will have a different ID. To get the image ID, it is best to use the API
- **Label ID**
    – The Label ID is specific to the species name / OTU name / label name that you want to give to your annotation. This ID can be found in the label tree as indicated above (section 7.3.1). Every OTU / species / morphotype in your list of annotations must be matched to an ID in an existing label tree before it can be uploaded to BIIGLE. Note that the label tree must be linked to the volume you want to modify.
- **Annotation shape ID**
    – The Shape ID: specifies what shape the annotation will be. BIIGLE can take annotations as point, circle, rectangle, polygons and other shapes. This needs to be fed into the API so it knows how to interpret the set of coordinates you are providing. This is a numerical ID where the shape of the annotation 1 is point, 4 is circle, 5 is rectangle. You can see all the shapes with this API request: https://biigle.de/api/v1/shapes
- **Coordinates of the annotation**
    – The Coordinates: As you have specified what you see and the shape the annotation will have, you need to say where in the image it is. The coordinates are the X and Y of the annotation in pixels within the image. The origin is the top left corner. It is a string of numbers specific to each shape: points take [X,Y], circles take [X,Y, Radius], rectangles take [X top-left, Y top-left, X bottom-left, Y bottom-left, X bottom-right, Y bottom-right, X top-right, Y top-right].
- **Confidence**
    – The Confidence is a number between 0 and 1 (1 meaning maximum confidence). It can be arbitrarily set to a constant by the user for all annotations.

In summary, the list of annotations must look like this:

	biigleid	filename	volume_id	label_id	label_name	shape_id	shape_name	points	confidence
1	2353607	M116_13331848_13109013355679.jpg	8188	3720	OTU261	4	Circle	[529.52,1641.59,43.67]	1
2	2353607	M116_13331848_13109013355679.jpg	8188	3940	OTU603	1	Point	[977.77,1755.15]	1
3	2353607	M116_13331848_13109013355679.jpg	8188	3940	OTU603	1	Point	[811.62,1626.05]	1
4	2353607	M116_13331848_13109013355679.jpg	8188	3940	OTU603	1	Point	[699.26,1875.28]	1
5	2353607	M116_13331848_13109013355679.jpg	8188	3940	OTU603	1	Point	[515.78,1534.02]	1

This information can easily be retrieved from an existing BIIGLE 'CSV' report. Note that the original 'image_id' and 'volume_id' in the report need to be replaced by the new image and volume ID corresponding to the new destination volume.

If your annotations are stored in a different format (from another software or other custom archiving scheme), the equivalent table must be recreated.

This R code example shows the basic code needed to create a label

```
# Example API code to create annotations

# put your credentials here:
login <- "NAME@plymouth.ac.uk"
token <- "yourtoken"

# make the base Url

Image_id <- "10000" # image in: https://biigle.de/images/10000/annotations

 url <- paste0("https://biigle.de/api/v1/images/",Image_id,"/annotations"  )

# make a point annotation - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - - - - - - - - - - -
# make a list with the body of the request
list( shape_id = 1, # for a circle
      label_id = 3720, # label ID for each OTU should be found in the
catalogue's table
      confidence = 0.75,
      points = c(150,150)) -> body_list
# send to the API
POST(url,
      config = authenticate(login, token),
      body = body_list,
      encode = "json" ) -> rp
# server response. If Status: 201, it is a success
  print(rp)

# make a circle  - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
- - - - - - - - - - - - - - -
   # make a list with the body of the request
  list( shape_id = 4, # for a circle
      label_id = 3720, # label ID for each OTU should be found in the
catalogue's table
```

```
        confidence = 0.75,
        points = c(150,150, 50)) -> body_list
  # say that to the API
   POST(url,
    config = authenticate(login, token),
    body = body_list,
    encode = "json" ) -> rp
   # server response. If Status: 201, it is a success
  print(rp)
```

R code and instructions to add annotations generated by a YoloV5 computer vision algorithm are available here: https://github.com/DeepSeaCRU/BIIGLE-resources.git (in the YOLO folder)

The YoloV5 to Biigle.Rmd is a follow up to the use of Yolov5 CNN to make predictions on images. You may want those predictions in BIIGLE along with your other annotations where it can be shared and manually curated. This R script will reshape the annotations produced by Yolo, particularly with regards to the names and IDs it needs to link the annotations to a label tree. It will then upload them to BIIGLE in a specific volume.

Download the repository on your computer and use the API - Biigle to Biigle.rmd script to add annotations from a BIIGLE CSV report into a volume of your choice.

8. Exporting BIIGLE files for use in YOLO

All machine learning frameworks require input data in a format different from what BIIGLE provides. We provide code to transform the BIIGLE output into a set of files needed to train a YOLOV5 CNN: https://github.com/DeepSeaCRU/BIIGLE-resources.git (in the YOLO folder)

9. Future updates to this manual

We have taken the decision to publish this manual as is but there are many aspects we would like to add in the near future, for example extracting images from video for use in BIIGLE and use of IFDOs. If you would like to help make this manual better you can either go to the living version on GitHub and raise an issue there that we can correct / add to in the next version, or contact us to contribute a section to the next version.