An Ecologist’s Guide to BIIGLE

This is the living version of the document where users can suggest edits to improve the Ecologist’s Guide to BIIGLE. You can find the finalized, original version on Zenodo here: LINK

# 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. We intend this to be a living document that others can contribute to here https://github.com/DeepSeaCRU/CRU-resources. 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.

|  |
| --- |
| image 1 |

Happy annotating!

# Setting up on BIIGLE

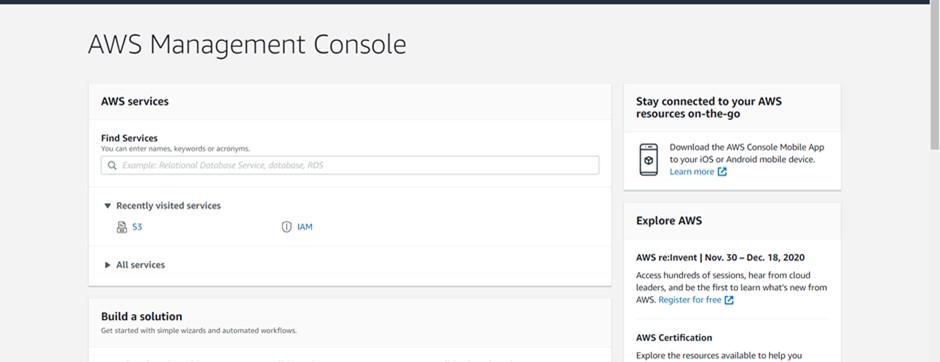
Getting set up on BIIGLE (and other annotation softwares) first requires that you host your imagery data, there are three options for this: via BIIGLE (upload files on homepage); local instances and remote volumes. See the BIIGLE manual for more details. Many institutes use a cloud-based storage solution 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.

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

Create an Amazon Web Services account.

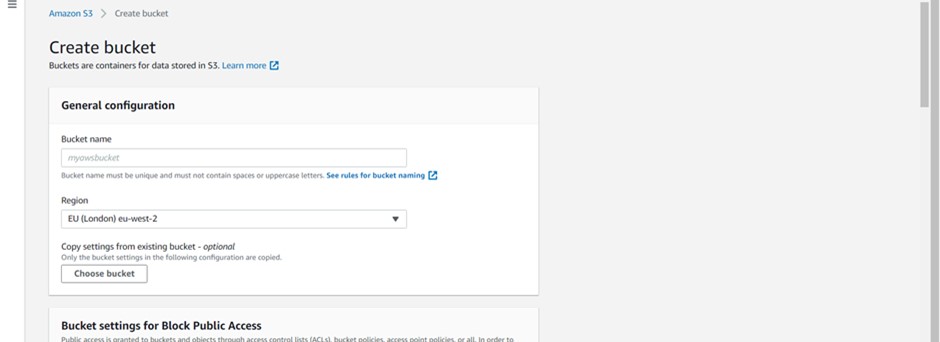
On the AWS management console type “S3” into the find services search box.



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”



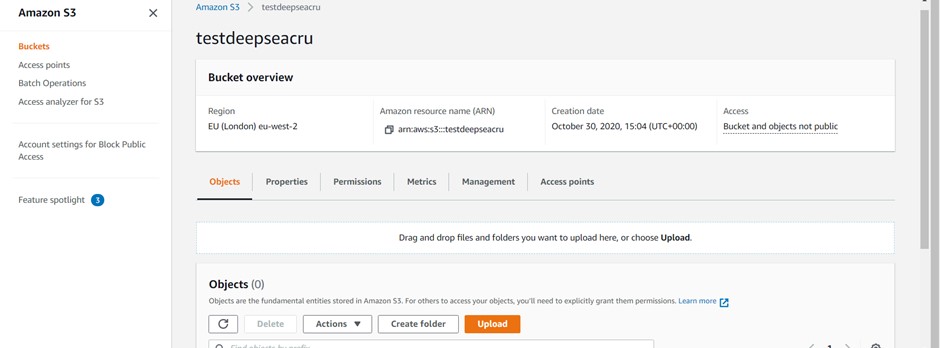
Name your bucket in a sensible way. BIIGLE recommends you use their personal random string generator [available here https://biigle.de/manual/tutorials/volumes/remote-volumes] to do this as this is hard to guess and keeps your data safer. But you can use whatever name you like.

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.

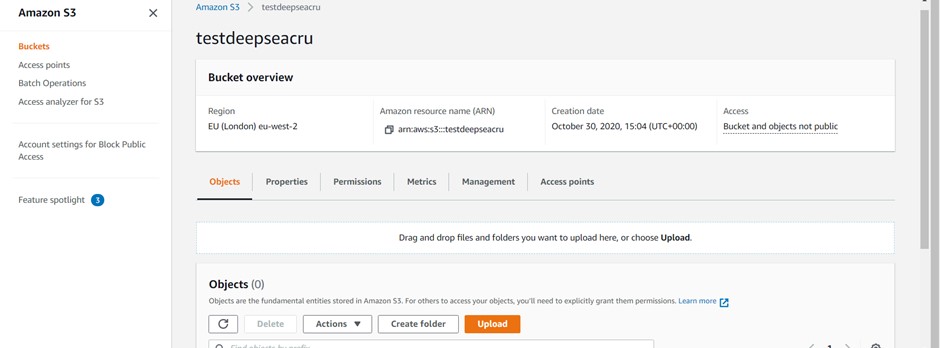
Copy and paste the script below into the text box. Then save changes.

[  
 {  
 "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.



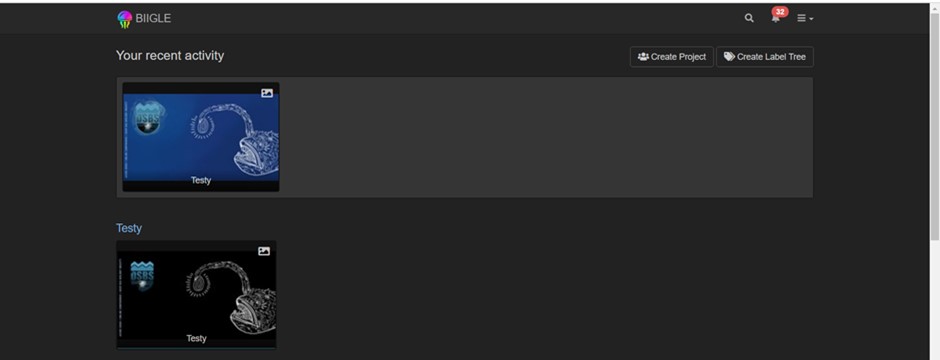
This is where you can upload files, so create a new folder, for example called test.

Click on your folder and upload a couple of images.

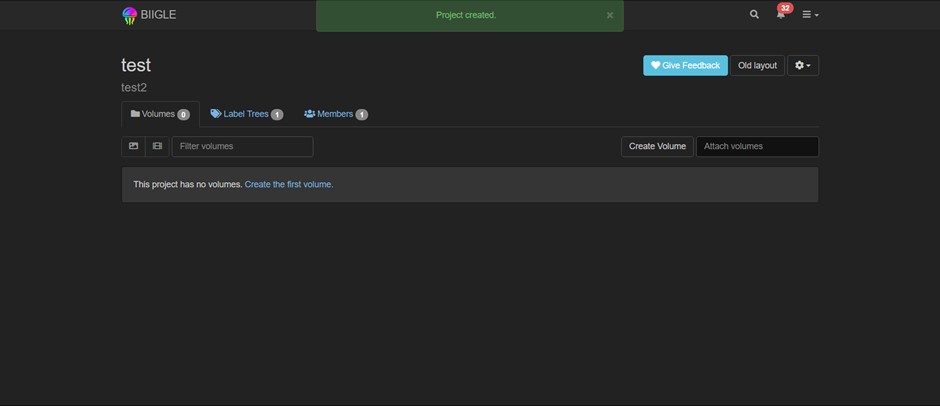
Click “upload”, click “add files”, add your files, scroll down to the access control list and check the “everyone (public access) read check box. A new check box will appear to confirm that you understand the implications of making your files public access, just check it, then scroll down to the bottom 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.

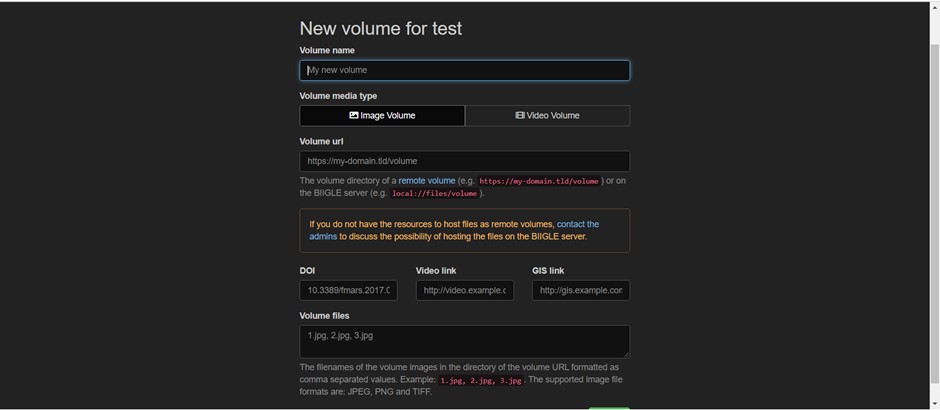
Now log into BIIGLE and go to your dashboard.



Click on create project



Click on create volume



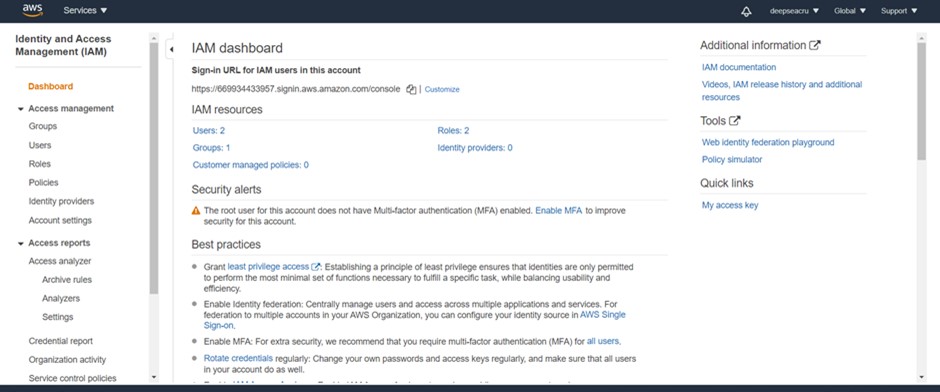
Here you can name your volume, for example it might be a cruise name, a project name or other useful name.

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 = " , "))

### 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.

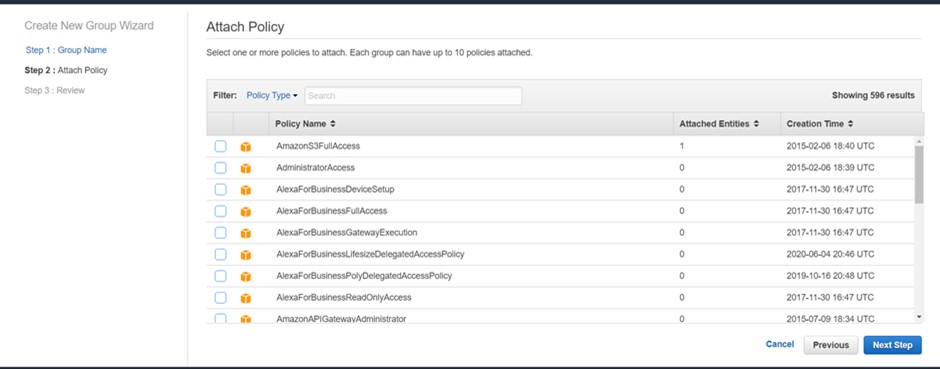


Click on groups on the left menu

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.

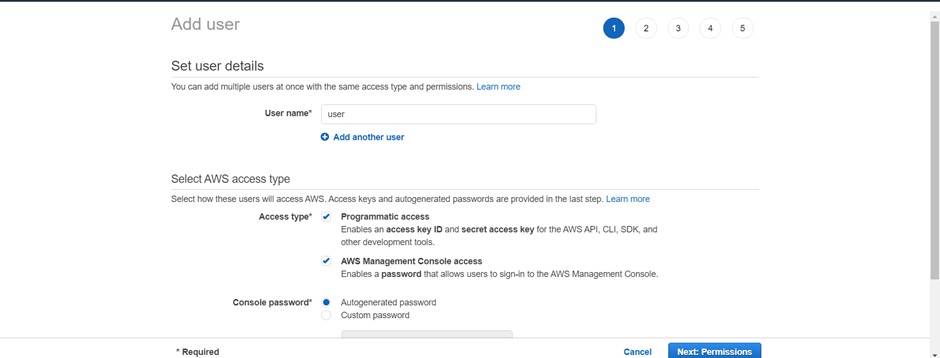


Then click next, then create group.

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!

## Formatting your data for use in BIIGLE

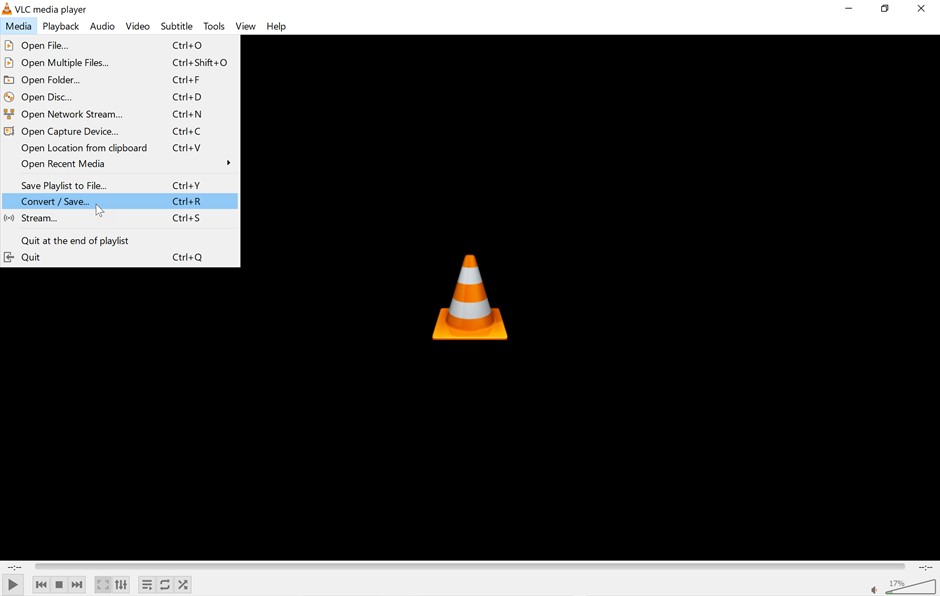
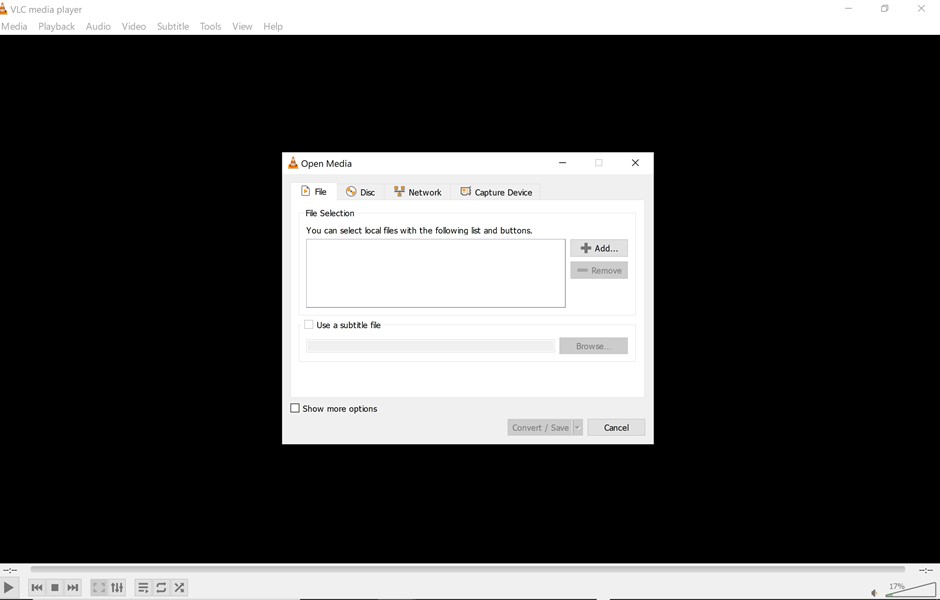
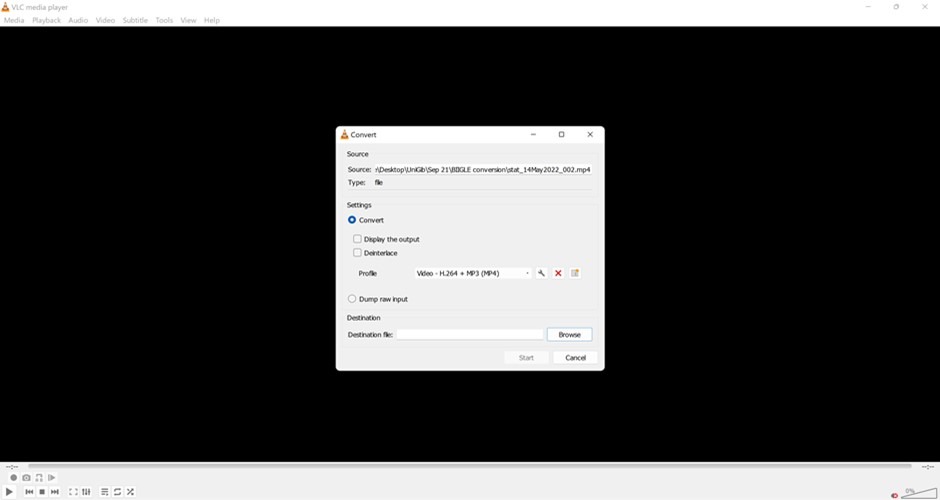
BIIGLE only accepts still images as either PNG or JPEG format.

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/).

## 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).

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.

## 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 🙂.

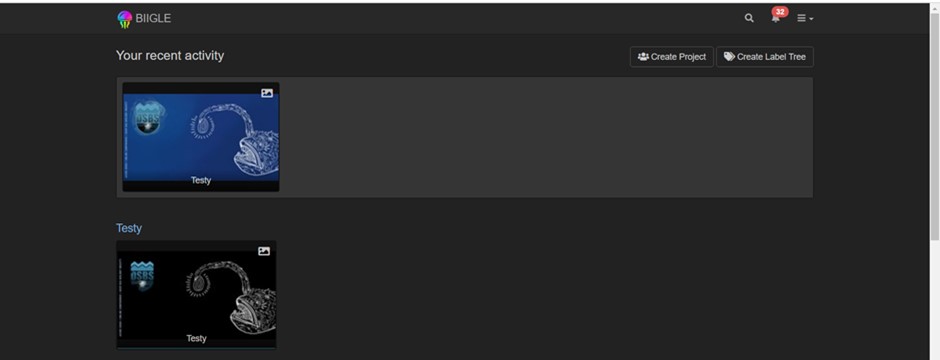
**IMPORTANT NOTES**: BIIGLE does not like spaces in file names, so use an underscore to separate out key parts of your file name. This also makes your file names machine readable. If you want your files to display in order when in BIIGLE you need to have numeric values at the start of the file name (e.g. 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.

# Setting up projects and volumes in BIIGLE

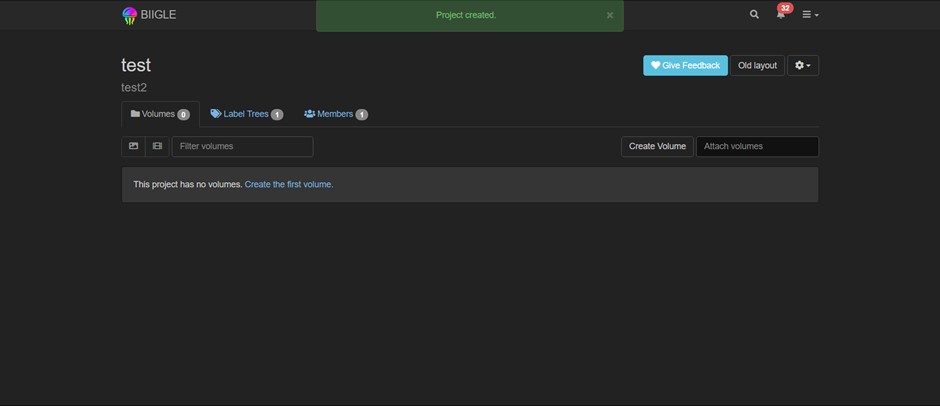
Once you have your files hosted, e.g. cloud storage set up, you can then go into BIIGLE to set up your projects and volumes. This process is relatively simple and you may already have done this when testing that your cloud storage works (see previous section).

First log into your BIIGLE account.

Go to your dashboard.



Click on create project

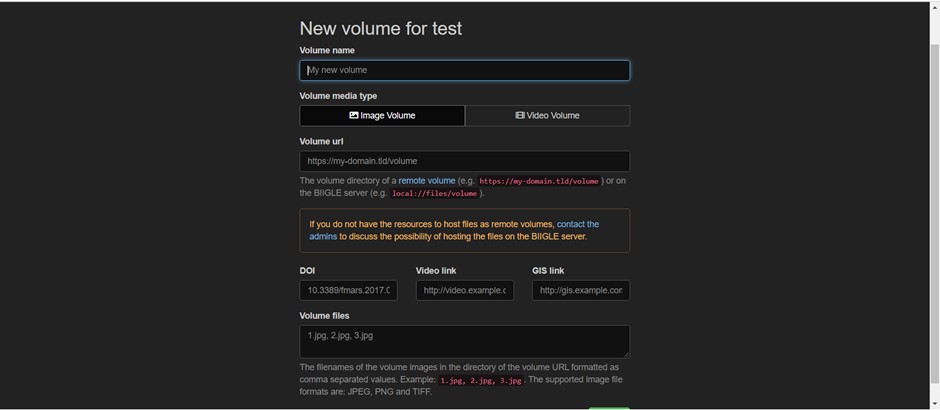


Adding imagery to projects can be done in two ways:

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.

## 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

## Make label trees

### To use a publicly accessible (standard) tree

### Creating your own tree from scratch

## Attaching a label tree to your project

# Make annotations

## Image annotation

## Video annotation

## Setting an annotation session

# Quality control and the largo tool

## Suggested best practice in annotation

# Downloading data and reformatting

## Get and use a report

### Image annotation report - CSV variant

# Using the BIIGLE API

## Intro to BIIGLE’s Application Programming Interface

## Basic requests

### Requesting a BIIGLE report

### Requesting the ID numbers of BIIGLE objects

## Using the API with R and Python

### Accessing label tree information

## Uploading annotations to BIIGLE

# Exporting BIIGLE files for use in YOLO

# Future updates to this manual