ANERIS

Operational Sensing Life Technologies for Marine Ecosystems

`

Deliverable D3.4 – AIES-ZOO Code and Documentation

Subject to change

Lead Beneficiary: EMBRC-SU

Author/s: Madeleine Walker, Marc Picheral, Camille Catalano, Julie Coustenoble, Jean-Olivier Irisson, Lionel Guidi

21/11/2024

Funded by the European Union

Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the EU nor the EC can be held responsible for them.

Prepared under contract from the European Commission

Grant agreement No. 101094924

EU Horizon Europe Research and Innovation action

Project acronym:	**ANERIS**
Project full title:	**operAtional seNsing lifE technologies for maRIne ecosystemS**
Start of the project:	January 2023
Duration:	48 months
Project coordinator:	Jaume Piera
Deliverable title:	AIES-ZOO Code and Documentation
Deliverable n°:	D3.4
Nature of the deliverable:	Report
Dissemination level:	Public
WP responsible:	WP3
Lead beneficiary:	EMBRC-SU
Citation:	Walker, M., Picheral, M., Catalano, C., Coustenoble, J., Irisson, J-O. & Guidi, L. (2024). AIES-ZOO Code and Documentation. Deliverable D3.4 EU Horizon Europe ANERIS Project, Grant agreement No. 101094924
Due date of deliverable:	Month n°24
Actual submission date:	Month n°24

Subject to change

Deliverable status:

Version	Status	Date	Author(s)
1.0	Draft	22 November 2024	Madeleine Walker (SU-EMBRC)
1.1	Draft	06 December 2024	Marc Picheral (SU-EMBRC), Camille Catalano (SU-EMBRC)
1.2	Draft	09 December 2024	Marc Picheral (SU-EMBRC), Camille Catalano (SU-EMBRC), Madeleine Walker (SU-EMBRC)
1.3	Review	17 December 2024	Xavier Salvador (CSIC)

1.4	Review	19 December 2024	Christos Arvanitidis, Cristina Huertas (LW ERIC)
1.5	Review	19 December 2024	Simon Korman (UoH)
1.6	Final review	20 December 2024	Xavier Salvador (CSIC)

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union

Table of Contents

Subject to change

Preface

To handle the large amounts of quantitative marine imagery data generated by the Underwater Vision Profiler 6 (UVP6) deployed operationally on EMSO moorings at OBSEA (Spain) and SMARTBAY (Ireland) at the start of the ANERIS project, the existing UVP6 software and applications are being overhauled to more fully automate the data pipeline and analyses. This pipeline is part of the Automatic Information Extraction System for Zooplankton Images, or AIES-ZOO, and involves data extraction, processing, and analysis for UVP6 images. In the deployed version of the AIES-ZOO pipeline, the UVP6 data collected (which includes size measurements of particles between 100 µm and 2 mm, and images of plankton and marine snow between 0.8 mm and 1 cm) is downloaded using the OctOs terminal application, transferred by Ethernet, and then processed with UVPapp, a user-friendly application developed for the UVP. The particle information is then uploaded to the web application EcoPart, while the images are sent to the web application EcoTaxa for taxonomic classification. To handle the millions of images and size spectra collected at the two cabled observatories in ANERIS, AIES-ZOO is being updated by i) fully re-coding EcoPart, including developing APIs to bridge the different applications automatically; ii) training deep learning classifiers to be deployed as a dedicated service for UVP6 image classification; iii) defining periodic validation procedures to control the performance of those classifiers.

Subject to change

Currently, five million image frames have already been captured by the two ANERIS UVP6s, with about 500,000 images of objects large enough for sorting and taxonomic classification. This deliverable provides the first version of the code and documentation for the updated AIES-ZOO pipeline, primarily concerning the specifications and code (in various stages of development) for the recovery of UVP data, the new version of EcoPart, and the EcoPart API. This code will also be updated to accommodate the UVP6-micro (UVP6m), a new version of the instrument which extends the objects imaged to microzooplankton and large phytoplankton, counting particles between 10 µm and 1000 µm and imaging objects from approximately 100 µm to 1000 µm. While not initially planned for deployment in the ANERIS project, the UVP6m is planned to be installed at both SmartBay and OBSEA by the end of the project. The code and documentation detailed in this document will continue to be developed and updated over the course of the project.

Summary

This deliverable D3.4 contains the code and documentation for

1) The manual recovery of UVP data at regular intervals for import in EcoPart and EcoTaxa, and the specifications for the automated data recovery
2) The new EcoPart specifications and backend
3) A description of the UVP6 image classifier

Links to the existing code and documentation are regrouped in Section 1.5, Table 2.

List of Abbreviations

AIES-ZOO - Automatic Information Extraction System for Zooplankton

ANERIS - operAtional seNsing lifE technologies for maRIne ecosystemS

API – Application Programming Interface

JB - Junction Box

NetCDF - Network Common Data Form

OctOs – OctOS.exe command line software

REST - Representational State Transfer

ROI – Region of Interest

SFTP – Secure Shell (SSH) File Transfer Protocol

UV-C – Ultraviolet C

UVP6 – Underwater Vision Profiler 6

UVP6m – Underwater Vision Profiler 6 micro

Subject to change

1.1 Instrument deployment

The UVP6 is an underwater, in-situ sensor that provides particle size spectra and plankton images, and is deployable on an array of autonomous and cabled platforms (Picheral et al. 2021). The UVP6s deployed at SmartBay and OBSEA (Figure 1) in ANERIS are connected to the observatory junction boxes for Serial and Ethernet connection, and programmed using the OctOs terminal application to set the recording and UV-C anti-biofouling parameters. The UVP6s are currently programmed to take an image every 10 seconds, with the UV-C antifouling lights switched on for 30 minutes and then off for 30 minutes in order to protect the optical surfaces of the instrument. These parameters will be evaluated and revised as needed following scientific analysis of the data and images captured. Data and images are stored internally by the UVP6 prior to download.

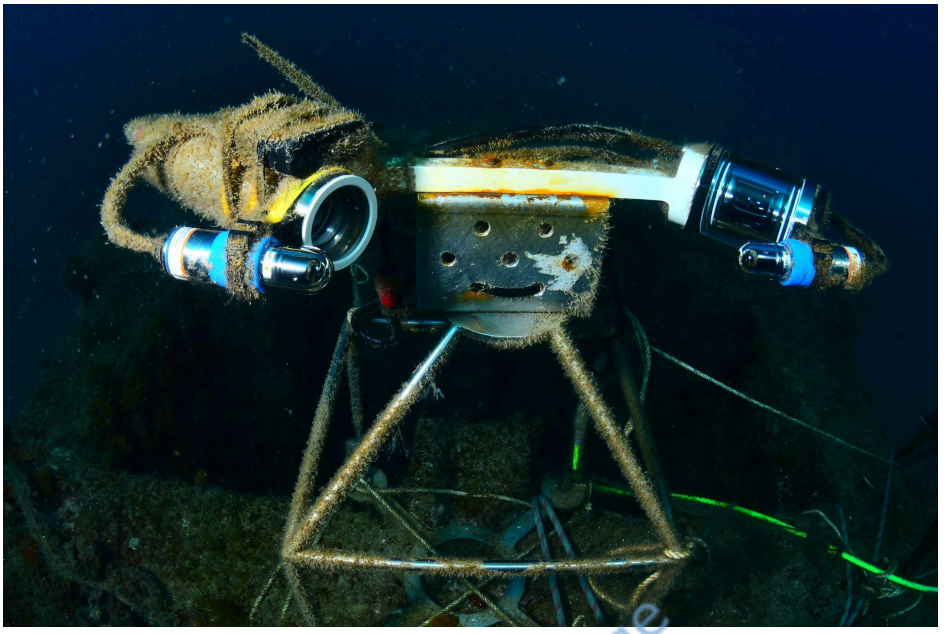


Figure 1. Image of the UVP6 deployed at the OBSEA cabled observatory.

Subject to change

1.2 Data Download & Processing

Data from the UVP6s deployed on the SmartBay and OBSEA cabled observatories is downloaded every month. The current manual protocol is depicted below in Figure 2, with an automated procedure under development described in Section 1.2.2 and depicted in Figure 3.

1.2.1 Manual data download & processing

The current operational procedure is manual, and is based on previously existing software (UVPapp) along with new scripts and protocols developed in ANERIS.

In the manual procedure, the acquisition is stopped temporarily as the data is downloaded to two personal computers hosted at OBSEA and SMARTBAY via an RS232 serial link and an Ethernet connection. The UVP6 is then rebooted following the OctOs protocol. The computer with the downloaded data can be accessed remotely via a remote access software. The dataset is reorganised using specific Matlab and Python tools and then processed using UVPapp, an application which prepares and processes the data and metadata for import in EcoPart, and which can also be used to remotely pilot the UVP6. UVPapp can be downloaded from the PIQv website.

In this manual version, two scripts (currently published on Github: https://github.com/ecotaxa/UVP_toolboxpy) are run to generate project metadata. The first script (uvp6_time_merge) allows for different UVP6 acquisition sequences to be merged into 24-hour sequences, while the other creates the corresponding metadata (uvp6_create_meta). Between

the two scripts, the data is manually split into two data projects, based on whether the UV anti-biofouling lights deployed for testing on the UVP6s were on or off.

Data is downloaded in the RAW folder of the data project of the instrument, with a data.txt file with all raw particle counting and sensor metadata plus a ZIP archive of all raw image/vignette files. The header of the data file also contains two metadata lines, one for the instrument hardware configuration and one for the acquisition settings utilized for the sequence. This data can then be uploaded as a project in EcoPart. More information regarding UVPapp can be found in the UVP6 User Guide, https://doi.org/10.5281/zenodo.14533847.

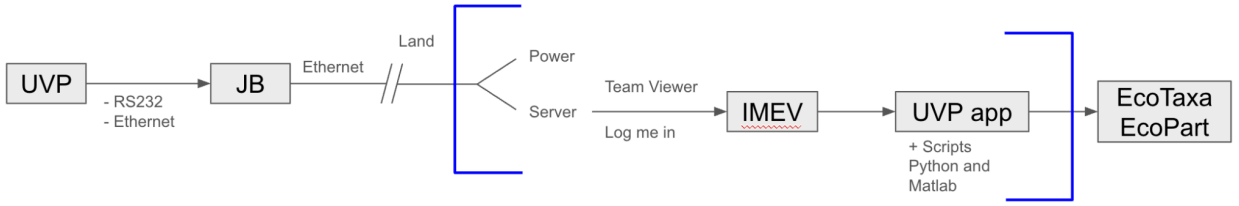


Figure 2. Schema representing the manual AIES-ZOO data pipeline

1.2.2 Automated data download & processing

In the AIES-ZOO pipeline, UVPapp is being replaced by a new application that will automate data download, initial processing, importation in EcoPart and EcoTaxa, and the automatic classification of the images in EcoTaxa. The specifications are described here: https://doi.org/10.5281/zenodo.14532277. In this new version, the data is first split into two projects, based on whether the UV anti-biofouling lights deployed for testing on the UVP6s were on or off. The second part of the pipeline will automate the steps made by the manual pipeline scripts. A code will then be run to segment the images, extracting regions of interest (ROIs) for large particles that can potentially be taxonomically identified (>0.62 mm). The code generates vignettes, descriptors and particle data in an import-ready .tsv file. This file is then imported into an SFTP, and subsequently transferred to the EcoPart application.

Subject to change

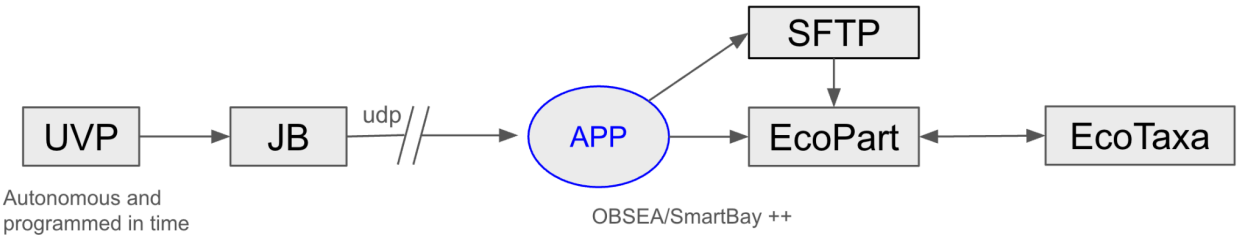


Figure 3. Schema representing the automated AIES-ZOO pipeline

1.3 EcoPart

EcoPart (https://ecopart.obs-vlfr.fr/) is an application specifically developed to manage all UVP data generated globally, but it is also capable of processing particle and image data from other instruments that adhere to the same data formats as the UVP (NetCDF). As part of ANERIS, EcoPart is being entirely re-coded, including an API to make it machine queryable, as well as quality control checks at project and data level (e.g., position, depth, and time metadata checks and basic range checks for particle concentrations). The specifications for the new version of EcoPart are further detailed here: https://doi.org/10.5281/zenodo.14534155. The codebase for the EcoPart backend is being developed using TypeScript, NodeJS, Express, SQLite, and Jest for testing: https://github.com/ecotaxa/ecopart_back. Work on the frontend has not yet started. Communication between the EcoPart backend and the front or a third party app will take place through backend API endpoint calls, with the API designed based on REST principles.

In AIES-ZOO, the pipeline will call the EcoPart API to import the UVP6 data held in the SFTP into EcoPart. Once imported into EcoPart, the ROIs will be automatically sent to EcoTaxa for identification. EcoPart structures its information by size and exports depth-resolved particle size spectra; it also retrieves the taxonomic identifications from EcoTaxa to export concentrations for the taxonomically identified large particles (including plankton). This summarized data will be exportable from EcoPart as NetCDF files.

1.4 Deep learning UVP6 image classification

In EcoTaxa, the ROIs are classified by a deep learning classifier trained on UVP6 images. While this classifier already exists, it will need to be retrained on a learning set of manually classified images from the SmartBay and OBSEA data in order to improve the classifications. These automatic classifications will then be validated by human experts in EcoTaxa, with human validators able to classify about 7000 images per hour. As more images from SmartBay and OBSEA are received and validated, the classifier will improve as its training data increases. The images from OBSEA and SmartBay are available in EcoTaxa, at the following links: OBSEA UV Off, OBSEA UV On, SmartBay UV Off, SmartBay UV On.

At the time of writing, over 4 million UVP frames have been sampled at the OBSEA site, with images extracted for classification on EcoTaxa from over 200,000 of these frames. At the SmartBay site, almost 900,000 frames were sampled, with over 100,000 images extracted for further analysis. The total daily concentration of imaged objects through time is higher at the SmartBay site compared to OBSEA, and there does not seem to be a seasonal signal at either site (Figure 4 & 5).

Subject to change

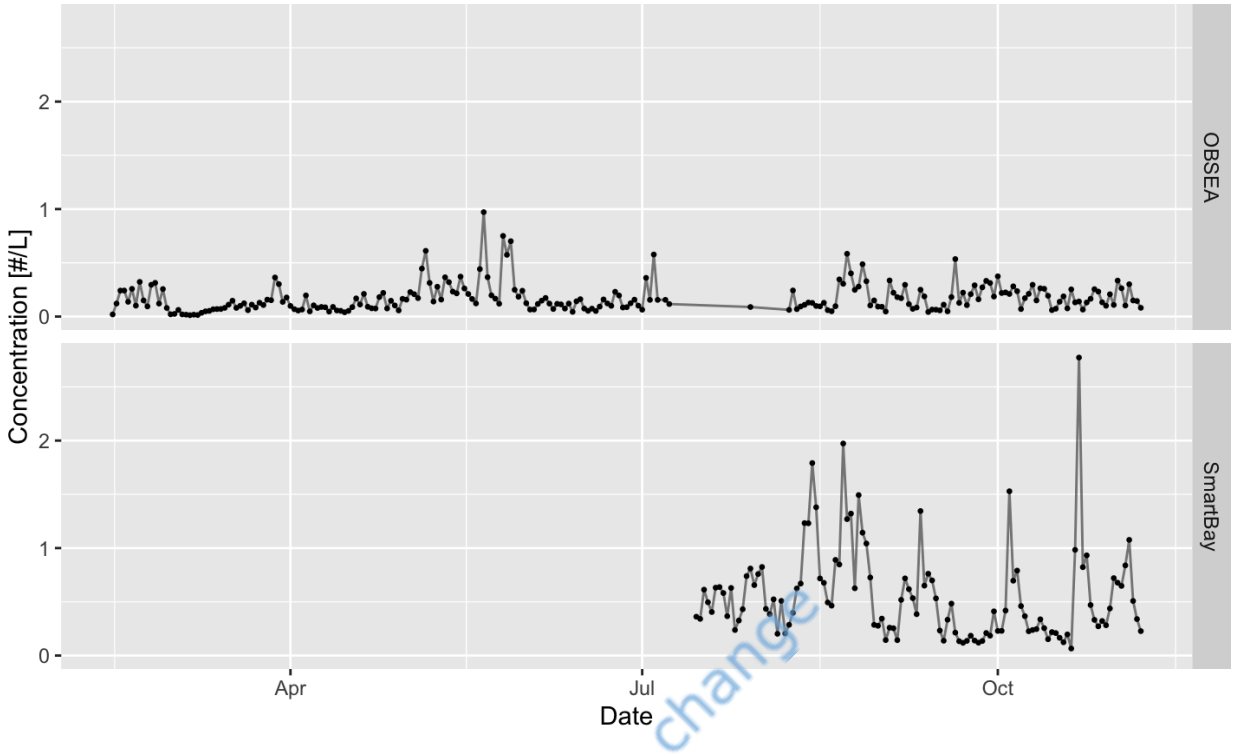


Figure 4. Concentration of imaged objects at OBSEA and SmartBay from February 2024 to present.

Subject to change

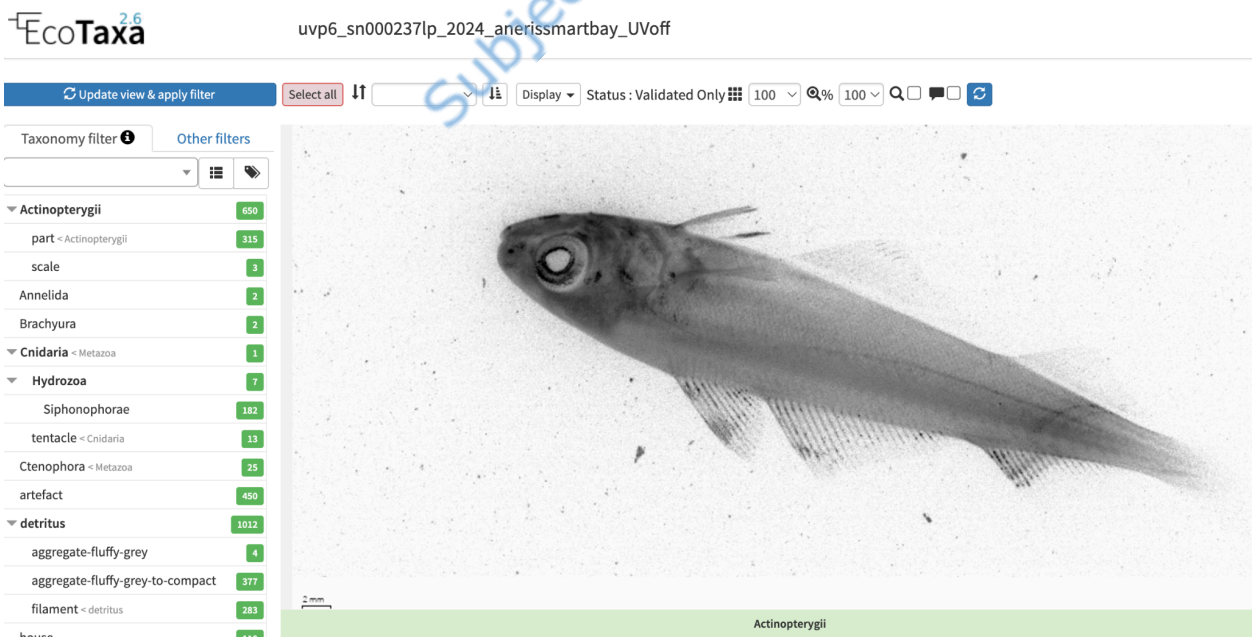
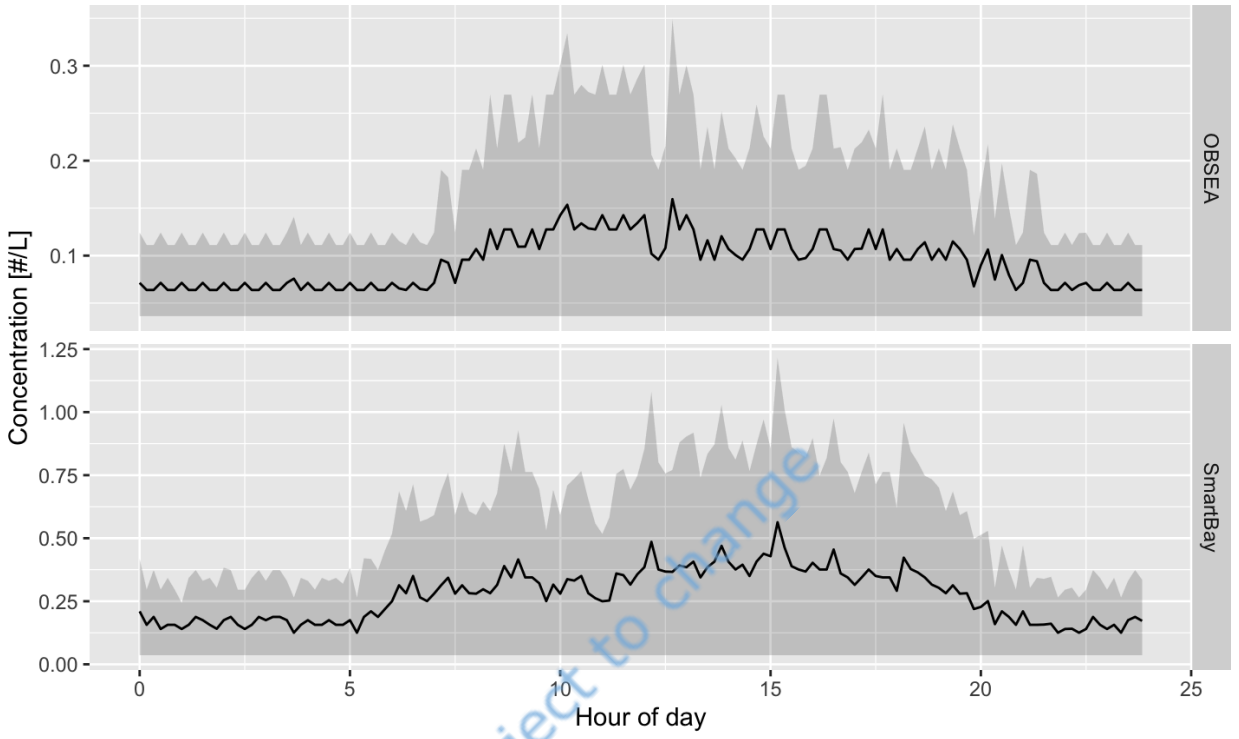


Figure 5. Example of an Actinopterygii image from SmartBay on EcoTaxa.

However, there is a definite daily signal, with very little activity at night at each site, particularly at OBSEA (Figure 6). This will be explored further, to determine whether this pattern is due to biological activity (e.g., no fish active at night), or a light-related artefact.



Subject to change

Figure 6. Daily concentration of imaged objects at SmartBay and OBSEA.

To date, 14,477 of these objects have been presorted on EcoTaxa, with predictions generated by the classifier on the ensemble of the images collected based on this presorting. From these predictions, the preliminary counts by taxon at each site are listed in the table below. Of note are the large numbers of fish (Actinopterygii) imaged, particularly at SmartBay, and the large numbers of appendicularia houses, amphipods, and small crustaceans at the OBSEA site. The SmartBay site also features numerous ctenophores and siphonophores. The prediction quality seems very good for fish and detritus, but needs to be improved for other taxa by increasing the number of human-validated images. In addition, the impact of fish on the plankton imaged will also need to be further explored, since at least some of the fish imaged are planktivorous and may greatly impact the observed plankton signals through predation. For example, a decrease in Appendicularia was observed in the middle of the day, with a concurrent increase in fish that potentially prey upon the Appendicularia.

Table 1. Total predicted object counts by taxon at OBSEA and SmartBay

taxon	OBSEA	SmartBay
detritus	85005	78427
house	67873	2425
aggregate-fluffy-grey-to-compact	33505	21501
Actinopterygii	17856	66306
part<Actinopterygii	17561	41069
aggregate-fluffy-grey	14129	5
scale	10944	5
fiber<detritus	7873	0
Amphipoda	5444	0
filament<detritus	3064	11621
artefact	2639	5521
Eumalacostraca	1637	0
t001	1048	1
seaweed	749	0
bubble	343	0
Cnidaria<Metazoa	148	4
solitaryglobule	103	0
fecal pellets	59	0
Copepoda<Maxillopoda	34	0
t002	32	57
Annelida	3	2
othertocheck	3	2
Collodaria	1	0
Appendicularia	0	1
Brachyura	0	2
Ctenophora<Metazoa	0	1066
Hydrozoa	0	16
Scyphozoa	0	3
Siphonophorae	0	488
tentacle<Cnidaria	0	5717

Subject to change

1.5 Summary of Code & Documentation

The following table summarises the previous, current, and future code or protocols for each part of the AIES-ZOO pipeline, including specifications where available. The operational code and protocols have been developed as part of the ANERIS project to treat the data generated by the UVP6s deployed at SmartBay and OBSEA, while the future code and protocols are a further automatization and amelioration of this existing pipeline.

Table 2. Summary of Code & Documentation

Part of AIES-ZOO Pipeline	Specification	Old code or protocol	Operational code or protocol	Final code or protocol (automatic)	Status
UVP6 data download	https://doi.org/10.5281/zenodo.14532277	Procol using OctOs	Manual pipeline using OctOs	Specified only	In Progress
Data reorganization	https://doi.org/10.5281/zenodo.14532277	N/A	Manual pipeline https://github.com/ecotaxa/UVP_toolboxpy	Specified only	In Progress
Prepare metadata and create samples	https://doi.org/10.5281/zenodo.14532277	N/A	Manual pipeline https://github.com/ecotaxa/UVP_toolboxpy	Specified only	In Progress
Process particles data	https://doi.org/10.5281/zenodo.14532277	UVPapp	Manual pipeline using UVPapp	To be coded	Not Started
Process image data and generate TSVs	https://doi.org/10.5281/zenodo.14532277	UVPapp	Manual pipeline using UVPapp	https://doi.org/10.5281/zenodo.14532277	Completed
Import data into SFTP	https://doi.org/10.5281/zenodo.14532277	N/A	N/A	To be coded	Not Started

subject to change

Part of AIES-ZOO Pipeline	Specification	Old code or protocol	Operational code or protocol	Final code or protocol (automatic)	Status
EcoPart Backend	https://doi.org/10.5281/zenodo.14534155	https://github.com/ecotaxa/ecopart	Manual pipeline https://github.com/ecotaxa/ecopart	https://github.com/ecotaxa/ecopart_back	In Progress
EcoPart API	https://doi.org/10.5281/zenodo.14534155	N/A	N/A	To be coded in EcoPart Backend https://github.com/ecotaxa/ecopart_back	Not Started
EcoPart Frontend	https://doi.org/10.5281/zenodo.14534155	https://github.com/ecotaxa/ecopart	Manual pipeline https://github.com/ecotaxa/ecopart	To be coded	Not Started
UVP6 deep learning classifiers		N/A	N/A	To be coded	Not Started

Subject to change

Conclusion and Next Steps

The current AIES-ZOO pipeline is functional in the sense that images taken by the UVP6s deployed at OBSEA and SmartBay are regularly imported to EcoPart and EcoTaxa and objects of interest can be classified rapidly. However, most steps are still manual, leading to an effort to automate the pipeline as much as possible, largely through the overhaul of the EcoPart application. Classification of more objects by human experts will also continue throughout the remainder of the ANERIS project to improve the predicted classifications. In addition, the pipeline will also be modified to accommodate the UVP6m instrument, which is planned for installation at each site in mid-2026. A dashboard will be created to visualize the processed data collected at SmartBay and OBSEA, with information about concentration by taxa at each site and trends over time. Classified and validated data will also be published on OBIS and GBIF for further exploitation by the scientific community.

References

Picheral, M., Catalano, C., Brousseau, D., Claustre, H., Coppola, L., Leymarie, E., et al. (2022). The underwater vision profiler 6: An imaging sensor of particle size spectra and plankton, for autonomous and cabled platforms. Limnology and Oceanography: Methods, 20(2), 115–129. https://doi.org/10.1002/lom3.10475

Subject to change