GRANT N°: 871153
PROJECT ACRONYME : JERICO-S3
PROJECT NAME : Joint European Research Infrastructure for Coastal Observatories - Science, services, sustainability
COORDINATOR : Laurent DELAUNEY - Ifremer, France - jerico-s3@ifremer.fr

JERICO-S3 MILESTONE

Joint European Research Infrastructure network for Coastal Observatory
Science, Services, Sustainability

MS#, WP# and full title	JERICO-S3 MS64 - WP10 - "Training Workshop #2"
5 Key words	EcoTaxa, planktonic imagery, data management and processing
Lead beneficiary	IH
Lead Author	Vânia Lima, Veronique Creach
Co-authors	João Vitorino
Contributors	
Submission date	2024-02-05

→ **Please specify the type of milestone:**

- Report after a workshop or a meeting (TEMPLATE A)
- ❏ Report after a specific action (TEMPLATE B) (test, diagnostic, implementation,…)
- ❏ Document (TEMPLATE B) (guidelines,...)
- ❏ Other (TEMPLATE B) (to specify) ……………………**.**

Diffusion list			
<u>Consortium beneficiaries</u>	Third parties	Associated Partners	other

PROPRIETARY RIGHTS STATEMENT

TABLE OF CONTENT

A) TEMPLATE A - report after a workshop or a meeting

1. A - Attendees

The Training Workshop #2 was part of JERICO-S3 project Task 10.4: Strengthening the community through skill development and knowledge transfer (Lead: IH; FMI, DELTARES, CNR, CEFAS). It is part of the internal workshops on best practices to train existing JERICO-RI facility operators (also open to external applicants), transferring knowledge on data management, processing, and QA/QC practices (WP5&6). It was proposed that particular emphasis should be given to promote synergies with the Virtual Research Environment (the JERICO-CORE, previously known as e-JERICO) that will be developed in WP7.

In this sense, this training workshop was focused on the Best Practices for Imaging Data Management, more specifically planktonic imagery data management and processing (WP9), with the organisation of practical sessions for using the EcoTaxa implemented in the JERICO e-infrastructure (https://www.jerico-ri.eu/va-service/ecotaxa/). In particular, the main targets were the data management and processing pipelines to EcoTaxa, training on Ecotaxa, and export from EcoTaxa to EMODnet biology, with applicants to the workshop being required to possess coding skills in R and Python.

The workshop, titled "Improving the data flow of plankton images to EcoTaxa", was co-hosted by CEFAS and CNRS-LOV, and took place at the Observatoire Océanologique de Villefranche-sur-Mer laboratory in France, spanning from June 13th to June 15th, 2023. It ran both in remote and in-person formats, with around 11 in-person participants and 3 online participants (Figure 1).

Figure 1 - Attendees Group Photo.

2.A - Background

The Training Workshop #2 was a dedicated internal workshop organised by CEFAS with partners from CNR-LOV, aiming to provide training on aspects related with the best practices for Imaging Data Management and focusing on the planktonic imagery data management and processing, with practical sessions for using the EcoTaxa implemented in the JERICO e-infrastructure.

The event was advertised beforehand on the JERICO-RI social media pages, more concretely in the JERICO-RI LinkedIn page, where it gathered circa 610 impressions.

News post on the Training Workshop were published online on LinkedIn as the workshop was occurring:

- day 1, gathering circa 555 impressions;
- day 2, gathering circa 665 impressions.

News posts about the workshop were also published online on the JERICO-RI X Social Media (formerly known as Twitter) profile.

3.A - Agenda A - Main report

The Training Workshop contributions are detailed next.

The initial objective of the workshop was to build a link to enable the use of EcoTaxa by employing the in place API to move the images from the instrument to EcoTaxa. The initial tasks defined were the following:

i.   prototype the code exploiting EcoTaxa's API to move images + metadata from instrument to EcoTaxa (1 day and a half),
ii.   use EcoTaxa and AI tools to sort images (1 day),
iii. use EcoTaxa's API to export data in a format that (Eur)OBIS accepts (half day).

The full agenda of the workshop can be found below:

Tuesday 13th			
09:00	10:00	Introduction to the workshop, import format for EcoTaxa	V Creach / JO Irisson
10:00	12:00	Coding of import pipelines	All
12:00	*14:00*	*Lunch*	
14:00	17:30	Coding of import pipelines	All
Wednesday 14th			
09:00	12:00	Coding of import pipelines	All
12:00	*14:00*	*Lunch*	
14:00	15:00	Requirements for import in EurOBIS/EMODnet	P Cabrera
15:00	18:00	EcoTaxa training	JO Irisson
Thursday 15th			
09:00	12:00	EcoTaxa training	JO Irisson
12:00	*14:00*	*Lunch*	
14:00	17:00	API export to DarwinCore Archive	JO Irisson / L Salinas / J Coustenoble

The goals of the workshop were to:

i. define and start to write software to go from the instrument to EcoTaxa for the CPICS and CytoSense (and FlowCam),
ii. show how to use EcoTaxa,
iii. understand the requirements of international databases (OBIS) and how an export from EcoTaxa can fulfil some of them.

The process to transfer the data from the three instruments to EcoTaxa were progressed but still need some development, regarding:

i. the CPICS, relevant files were identified in the file structure present on the CPICS. Parsers were written for the roicoords and syslog files as well as a first version of an image processor based on the already segmented images. This allowed the upload of a few images (and their masks) on EcoTaxa, manually (i.e. not using EcoTaxa's API). The syslog now systematically contains the key parameters to compute the water volume images, hence the concentrations of particles. Since the data is provided as a continuous collection of time stamped files, an automatic detector of casts and yos in continuously sampling deployment modes should be implemented. However, Lat and lon coordinates are not available in the input files but can be changed (e.g. when the system is towed).
ii. The Cytoprocess, CytoSense outputs .cyz are binary files. The CytoClus software parses those and outputs .csv files and images but the naming convention and the internal format of those are different among the different instruments and software versions; they cannot be trusted. A crude method to summarise the pulse shapes through polynomials was designed. A first image processor was implemented.
iii. The flowCAM, The processing of colour FlowCam images based on VisualSpreadsheet's collages was achieved with ZooProcess. The result could not be imported into EcoTaxa due to a few missing metadata fields. This will be fixed in the near future.

4.A - Conclusions

The use of EcoTaxa was demonstrated. This highlighted the importance of importing rich metadata and the possibilities offered by the machine learning back end to quickly sort through data.
The requirements for import into OBIS were presented, including some details on the DarwinCore Archive format and the specific metadata required for occurrences coming from imaging instruments.

In the particular case of the Cytoprocess (flow cytometry), the workshop highlighted the difficulty to have access to the raw data and the need to create an API. The API is now available and new processes to be able to use ECOTAXA to identify images from Cytosense/Cytosub are in progress as well as to automatically cluster the particles. In June 2024, a second workshop (4th, 5th, and 6th of June in Wimereux, North of France) on best practices for building training sets and machine learning processes in flow cytometry analysis will take place. This Foresight workshop is also supported by EuroMarine, OBAMA-NEXT (Grant agreement ID: 101081642, project coordinator: AARHUS UNIVERSITET), IFSEA Graduate School (France 2030 grant ANR-21-EXES-0011, project coordinator ULCO) and RioMAR (France 2030 ANR PPR, project coordinator CEA).

5.A - Annexes and references

Attendees	Institution	role
Jean-Olivier Irisson	CNRS-LOV	organiser
Veronique Creach	CEFAS	organiser
Natasha Hynes	University of New Brunswick	participant
Patricia Cabrera	VLIZ	trainer
Luis Felipe Artigas	CNRS-ULCO	participant
Julie Coustenoble	CNRS-LOV	trainer
Sebastien Galvagno	CNRS-LOV	trainer
Robert Blackwell	CEFAS	participant
Lise Jeabert	CNRS-ULCO	participant (student)
Lucinda Lanoy	CNRS-ULCO	participant (student)
Daniel Blandfort	HEREON	participant
Ankita Vaswani	HEREON	participant
Ivan Vlasicek	FAMRI	participant
Laurent Salinas	CNRS-LOV	technician