



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

<u>JERICO-S3 MILESTONE</u>	
Joint European Research Infrastructure network for Coastal Observatory Science, Services, Sustainability	
MS#, WP# and full title	JERICO-S3 MS32 – WP6 - Workshop reports for establishing best practices for imagery data management
5 Key words	EcoTaxa, planktonic imagery, data management
Lead beneficiary	CNRS
Lead Author	Veronique Creach,
Co-authors	Patricia Cabrera, Jean-Olivier Irisson, Fabien Lombard.
Contributors	
Submission date	2024-06-15

→ 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

THIS DOCUMENT CONTAINS INFORMATION, WHICH IS PROPRIETARY TO THE **JERICO-S3** CONSORTIUM. NEITHER THIS DOCUMENT NOR THE INFORMATION CONTAINED HEREIN SHALL BE USED, DUPLICATED OR COMMUNICATED EXCEPT WITH THE PRIOR WRITTEN CONSENT OF THE **JERICO-S3** COORDINATOR.

*According to the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) and the 78-17 modified law of 6 January 1978, you have a right of access, rectification, erasure of your personal data and a right of restriction to the data processing. You can exercise your rights before the Ifremer data protection officer by mail at the following address: IFREMER – Délégué à la protection des données- Centre Bretagne – ZI de la Pointe du Diable – CS 10070 – 29280 Plouzané - FRANCE or by email: dpo@ifremer.fr // jerico@ifremer.fr
 Ifremer shall not hold your personal data for longer than necessary with regard to the purpose of the data processing and shall destroy it thereafter.*



TABLE OF CONTENT

A) TEMPLATE A - report after a workshop or a meeting	3
1. A - Attendees	3
2. A - Background	4
3. A - Agenda A - Main report	4
4. A - Conclusions	6
5. A - Annexes and references	6

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

1. A - Attendees

Two workshops on plankton imagery data management have been given online in 2021 and 2022.

- First workshop during the JERICO week (19-23 April 2021): 11 participants.
- Second workshop during the JERICO week (15-18 March 2022): 20 participants.

2.A - Background

The analysis of planktonic community (nano-, microplankton and mesozooplankton) in terms of taxonomy and counts relies on an experienced biologist who processes the samples by light microscopy. However, the analysis presents a certain number of challenges such as alterations in morphology, colouration or even disappearance (e.g. gelatinous plankton). Consequently, several *in situ/ in vivo* and bench top imagery technologies have been developed to determine the taxonomy and the biomass of organisms from microns (phytoplankton) to centimetres (zooplankton, fish) during the last 15 years. These technologies can usually be installed in different types of platforms at sea such as buoys and ships of opportunity for a real-time data acquisition avoiding the collection and fixation steps. However, new challenges have appeared with these new technologies. They are related to the high volume of data collected, their processing, and the definition of new variables generated with the image acquisition. In the first workshop, an overview of the quantitative imaging was presenting by experts as well as already in place processes in EurOBIS. The second workshop explained more in details the data management for imagery and the development of new vocabulary in the framework of JERICO-S3.

3.A - Agenda A - Main report

The first imaging workshop was an online workshop which has been taken place on the 21st of April 2021 due to covid restrictions. Because of limited facilities and time, the data management for imagery workshop was combined with the data management of biological sensors. See the agenda below:

			speakers
16:00	16:05	description of the objectives of the workpackages	Veronique Creach (Cefas)
16:05	16:15	Imagery: data flow presentation	Fabien Lombard (CNRS-LOV)
16:15	16:30	discussion on data flow imagery	Patricia Cabrera (VLIZ)
16:30	16:40	Flow cytometry: data flow presentation	Melilotus Thyssen (CNRS-MIO)
16:40	16:55	discussion on flow cytometry data flow	Veronique Creach (Cefas)

16:55	17:00	Benchmark for the data management in the partnership	Veronique Creach (Cefas)
17:00	17:20	Miro: questions	Patricia Cabrera (VLIZ)
17:20	17:30	summary of the Q&A	Veronique Creach (Cefas)

The presentation “*Imagery: data flow presentation*” aimed to present the principal characteristics of the **quantitative imaging**: standardisation of the image acquisition, measuring a full sample (no selection) with a maximum of morphological variables and associated metadata (position, depth, date, volume sampled/fraction imaged). It has been highlighted that one instrument cannot capture the whole size range of the plankton community. Consequently, there is a strong need for cross-calibration between the instruments. The advantages of quantitative imaging are fully quantitative and repeatable. The results can be re-explorable and discussed in collaborative network using digital archive. The variables captured with the imagers are numbers and biovolume as a proxy for biomass. The limitations are the low taxonomy resolution and potential fragmentation bias. EcoTaxa (<http://ecotaxa.obs-vlfr.fr/api/docs/>) is a free collaborative tool built from automatic classification algorithms (random tree forest, CNN networks) for hosting, sorting, annotating taxonomically and sharing images. Additional features allow the operators to directly register their results of the analyses to EMODnet then EurOBIS and then OBIS (see Fig. 1).

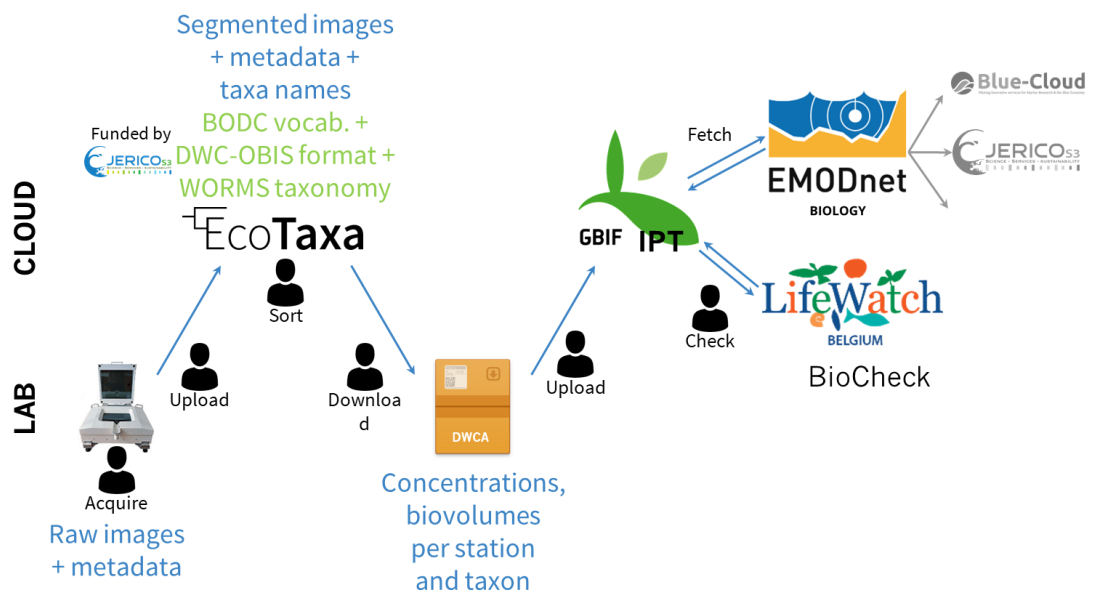


Figure1: data imaging flow from instrument to European database using EcoTaxa.

To prepare the workshop, a common survey (imagery and biological sensors) has been sent to the JERICO-S3 partners. The answers have been shared and discussed with the participants during the workshop. See below, the questions and answers from the survey (see annexe).



- 1 Do you have metadata for your files?
answer: *Everybody has access to a metadata repository.*
- 2 Is your metadata publicly available?
answer: *54% have metadata which are not accessible.*
- 3 Does your metadata format follow international standards?
answer: *46% have their own format, 8% is in a shape of EMODnet and SeaDataNet.*
- 4 Do you use a data infrastructure to publish the data and metadata?
answer: *69% have metadata and data published at the organisation level, 30% at the national level, only 15% at the European level.*
- 5 Do you use controlled vocabularies (BODC)?
answer: *15% used a vocabulary BODC, metadata: 46% in progress, 39% no, data: 54% no, 31% in progress, 15% yes.*
- 6 Do you archive raw data or images?
answer: *92% have access to archives.*
- 7 Do you have the possibility to create a DOI for your data.
answer: *62% have access to DOI facilities.*

The conclusion of the survey and discussion were summarised below:

- The researchers of the partnership are fully aware of the importance of data management. Some organisations have built facilities and support (data storage, in place data manager). However, most of the biological data from imagery and metadata associated do not reach the European data infrastructures.
- The archives are mainly external hard disk which can be lost or corrupted.
- DOIs are used in different ways according to the organisations. For example, if Ifremer generates a DOI for non-quality-controlled data, Cefas does not encourage it. However, both organisations have the facilities to generate DOIs as well as many organisations in the partnership.

The workshop 1 ended by an online consultation (see annexe: MIRO) with the experts to define potential barriers to FAIR data. The conclusions were summarised below:

- Lack of automatic process to analyse and transfer the data for some instruments (e.g. EcoTaxa is optimised only for some instruments such as Zooscan and UVP6).
- Limited or non-existing expertise in data management from the scientists collecting the data.
- Lack of support at the organisation level.
- Researchers agreed to use EcoTaxa but alternatives exist such as national Helmholtz alternative "MareHub", ZooPhytoImage.

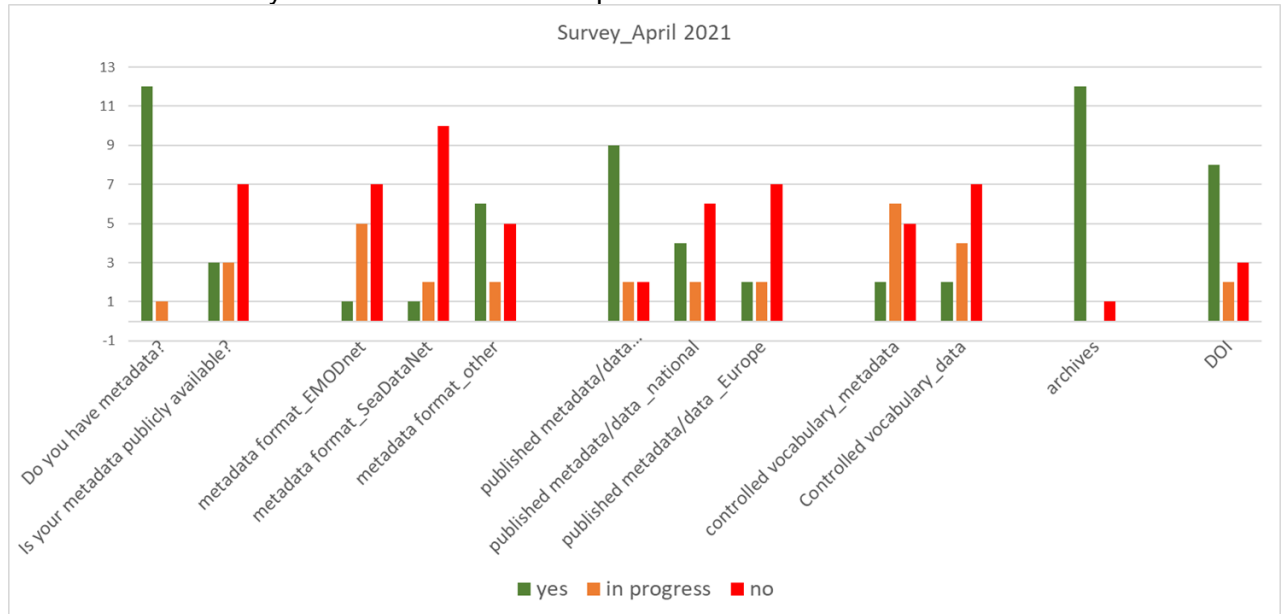
- Researchers agreed to publish their data in EMODnet biology but alternatives exist such as Seanoec/ EBI bioimage archive and DOI + local/national data repositories.

The second imaging workshop “submitting data to EurOBIS” was an online presentation by Patricia Martin-Cabrera (VLIZ) on the 17th of March 2022. It aimed to highlight the need of universal standard in Best Practices and recommendations for plankton imaging data (<http://dx.doi.org/10.25607/OBP-1742>). The presentation began by some examples of biological data already stored in EcoTaxa, GOOS and OBIS from Tara expedition, followed by common metadata format among instruments, presentation of new references generated during JERICO-S3 by a group of experts (e.g. pixel size, mayor axis length, minor axis length, area, width) and published in BODC, inclusion of relevant DarwinCore (DwC) terms to the OBIS-ENV-DATA format (and EurOBIS database) and an example of EcoTaxa Workflow.

4.A - Conclusions

5.A - Annexes and references

Results of the survey before the first workshop:



Results of the MIRO exercise from the first workshop:



cefas.sharepoint.com | Data management (1).pdf | 1 of 1 | Ask Copilot

C:\Users\vc00\OneDrive\%20%20CEFAS\projects\JERICOS3\Data%20management%20(1).pdf

Color code
Green -> Flow cytometry
Orange -> Imagery
Blue -> Flow cytometry/imagery

Only for Imagery

Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question	Header/Question
What can be the barriers to use the data from generated today for FLOW CYTOMETRY?	lack of standardized flow cytometry data across projects	no automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow
When can be the barriers to not use the data generated today for IMAGERY?	lack of standardized imagery data across projects	no automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow	lack of automated flag	standardized flag (yes/no) only in the workflow
Can you do the data management?	IF YES, SPECIFY WHICH STEPS	until a certain step	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially	partially
Can your organization provide help in data management?	YES/NO WHY?	lack the needs to be shared in the budget	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Will you use ECOTAXA?	IF YES	yes	no (already did)	yes if possible	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Alternatives?	SPECIFY																		
Do you anticipate the need to purchase flow cytometry?	IF YES, SPECIFY WHICH ONE	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry	purchase flow cytometry
Alternatives?	SPECIFY																		
Additional comments																			

16°C Mostly cloudy 19:07 13/06/2024