JERICO-DS Deliverable

Joint European Research Infrastructure network of Coastal Observatories

Design Study

WP, Deliverable and full title	JERICO DS - WP3 – D10 / D3.4 - *Data Management Plan for e-JERICO added value products*
5 Key words	Data management plan, data product, FAIR principles, HF-Radar, Glider, Ecotaxa, multiplatform
Lead beneficiary	SMHI
Lead Authors	Emilie Breviere (SMHI)
Co-authors	Juan Gabriel Fernandez, Sebastien Legrand, Julien Mader
Contributors	Patrick Gorringe, Milla Johansson, Jukka Seppäla, Jean-Olivier Irisson, Juan Miguel Villoria
Submission date	2023.05.29
Nature	Report

Nature

X R = Report, □ P = Prototype, □ D= Demonstrator, □ O = Other

Dissemination level :

X PU □ PP □ RE □ CO

PU = Public, PP = Restricted to other programme participants (including the Commission Services), RE = Restricted to a group specified by the consortium (including the Commission Services), CO = Confidential, only for members of the consortium (including the Commission Services)

GRANT N°: 951799
PROJECT ACRONYME : JERICO-DS
PROJECT NAME : Joint European Research Infrastructure network of Coastal Observatories - Design Study
COORDINATOR : Laurent DELAUNEY - Ifremer, France - design.jerico@ifremer.fr

Document information	
Document Name	*JERICO-DS-WP3-D10/D3.4 - Data Management Plan for e-JERICO added value products*
Document ID	*JERICO-DS-WP3-D3.4-2023.05.29-V1.0*

JERICO-DS Milestone

Joint European Research Infrastructure network of Coastal Observatories

Design Study

Revision	Date	Modification	Author
V1.0	29/05/2023	First Version	Emilie Breviere

APPROVALS				
	Name	**Organisation**	**Date**	**Visa**
Coordinator	Laurent Delauney	Ifremer	30/05/2023	X
WP Leaders	Sébastien Legrand	RBINS	29/05/2023	X
WP Leaders	Juan Gabriel Fernandez	SOCIB	29/05/2023	X

Diffusion list			
Consortium beneficiaries	Third parties	Associated Partners	other
X	X		

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 – 29280 Plouzané - FRANCE or by email: dpo@ifremer.fr + design.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

1. EXECUTIVE SUMMARY

The WP3 task 3.5 'e-JERICO data lifecycle management' objective is to address the most relevant aspects related to the lifecycle management of added-value data products which are and will be generated by the JERICO-CORE Data to Products Thematic Services (D2PTS).

This document presents the Data Management Plan (DMP) of the JERICO-Research Infrastructure (RI) data products as per today which is composed of a DMP per existing data product, and includes a template to complete when new data products emerge.

The Introduction section informs about what is JERICO-CORE and data products, it also presents the specificity of the DMP of JERICO-RI data products. The Method section details briefly the process that led to the DMP, including the information based on the Milestone 18 MS3.6 "Draft e-JERICO data management plan".

2. INTRODUCTION

JERICO-CORE

JERICO-CORE, i.e. JERICO e-infrastructure (previously identified as e-JERICO), is envisioned as a one-stop-shop service providing JERICO Research Infrastructure (RI) users an optimal way to gain an integrated form of access to (figure 1):

- TransNational Access to the JERICO physical infrastructures (platforms and sensors) offered by the national coastal observatories and JERICO technical expertise centres;
- Resources required to both harmonize and implement JERICO data lifecycle management methodologies: Best Practices, tools and services, and e-training modules;
- Quality controlled data that is routinely acquired by the different national coastal observatories, following the FAIR principles;
- Added-value products and services (indicators, nowcasts, analysis, etc.) generated by each individual JERICO thematic expertise center; and
- Dedicated cloud computing resources (hardware and software) allowing researchers to perform advanced analysis on multi-disciplinary, multi-scale, multi-domain and multi-sensor data sets.

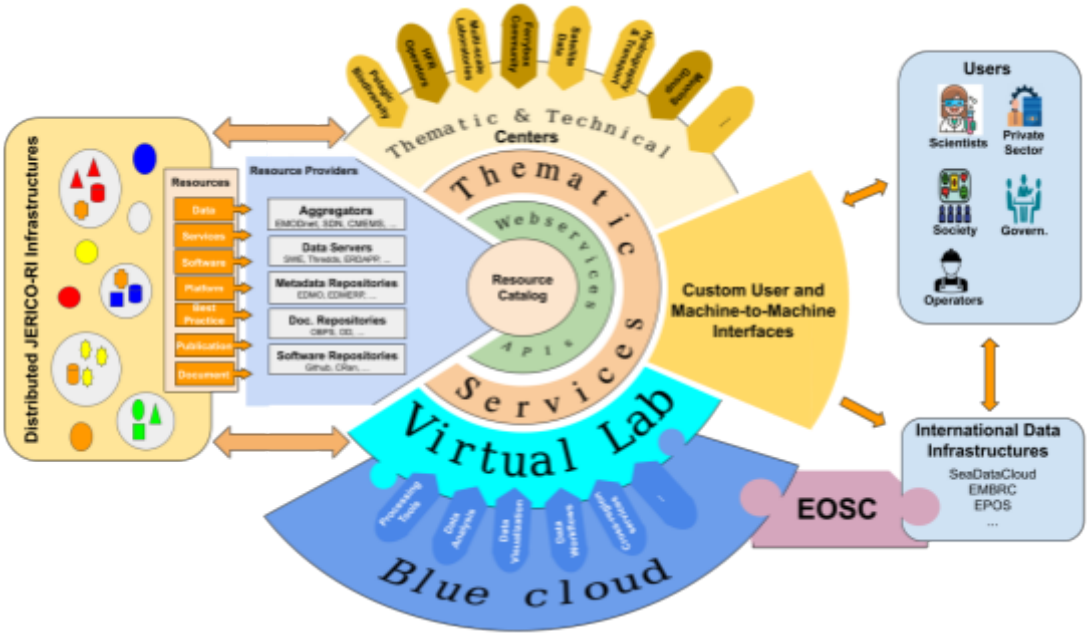

Figure 1 : *The JERICO-CORE concept with the Thematic Services at the centre in salmon colour.*

During the JERICO-S3 project, the creation of advanced products is addressed for the first time as use cases that will help validate JERICO-CORE. It was envisioned that JERICO-CORE will provide the support to link datasets of different nature in addition to the necessary assets to allow advanced processing to occur. In JERICO DS project, the necessary expertise to understand the data life cycle of this new generation of output data is under construction within WP3.

DATA PRODUCTS

Data products are defined as output of the JERICO-CORE Data to Products Thematic Services (D2PTS).
The JERICO-S3 Data Management Plan (DMP) defined, in section 3.1.2, the data products created under JERICO-S3 as "*products being created using the observation data, e.g. during the scientific work in the PSS's and IRS's, as well as in WP7 in the e-infrastructures. These products will be stored according to international standards as much as possible and published via metadata services and OGC compliant data access services. This will facilitate cross-disciplinary studies which are essential for a comprehensive understanding of coastal environments. Tools used in the data product creations may also be available through the e-infrastructure. Data products will include provenance and uncertainty descriptions so that users understand the potential and limitations of the data.*"

DATA MANAGEMENT PLAN OF JERICO-RI DATA PRODUCTS

We envision two main scenarios for the data flow of advanced products depending on the environment where these products are created. The two scenarios involved the collection of the information about these data products and the associated assets (tools, services, documents,...) into the resource catalogue of JERICO-CORE (figure 1):
1) The first scenario corresponds to the creation of data products within the Virtual Research Environment (VRE) of JERICO-CORE, likely Blue-Cloud. Since this environment will provide the necessary support to harvest the metadata of the data products and their link with other resources, the registration into the resource catalogue will happen mostly automatically.
2) In the second case, the advanced products will be created in the partners infrastructures and will require the definition of clear procedures and interfaces for the harvesting into the resource catalogue to happen.

The 2nd case will likely be the most common case, especially at the beginning while the JERICO-CORE infrastructure matures. It is also the most difficult case to address because it will require the understanding of the various scenarios in which this type of processing occurs on the partner's side. We need to understand the specificity in order to satisfy the long term perspectives of the JERICO DMP.

It is also essential to highlight that the JERICO data product DMP is focussing on the data products, it does not include the data used, which is already considered in the JERICO-S3 data management plan (deliverable 6.1 of the JERICO-S3 WP6).

Because of the nature of the JERICO data products, the DMP will actually be 1) a collection of DMPs (one per data product) and 2) a living document as JERICO is maturing and new data products will be produced in the future.

METHOD

With all this in mind, for Milestone M18 MS3.6, we drafted a DMP template in collaboration with JERICO partners and the authors of the JERICO-S3 DMP (Deliverable 6.1).

Furthermore, the existing High Frequency Radar (HFR) D2PTS is one of the services of JERICO-S3 that are used to validate JERICO-CORE and also, the HFR community is well coordinated and is an excellent example providing harmonised and thoughtful procedures and products. Therefore, we completed the proposed DMP template with the information on the data product: HFR Gap filled surface current Advanced Products (MS3.6). This exercise allowed us to further fine tune the template of the JERICO data product DMP.

For the present deliverable, the template was further scrutinised and adjusted based on comments from JERICO-CORE developers. The template composed of sections and series of questions for each section, was amended, for example a section on Operationality was added and few questions related to the future evolution of JERICO RI were added (JERICO-label).
The template is available in section 3 Main Report..

The section Main report consists of the JERICO-RI Data Products DMP. It is composed of 3 sections: Introduction, Template and DMPs of existing Data Products as per April 2023 (BGC, HFR, ecotaxa and gliders).

3. MAIN REPORT

Data Management Plan of the JERICO-RI Data Products

1) Introduction

JERICO participates in the Open Research Data Pilot which is part of the Horizon 2020 programme. The goal is to provide, where possible, accurate and high-quality data to the research community so that the RI will contribute to future advancements in the field of coastal observation.

The purpose of this Data Management Plan (DMP) is to identify within the JERICO-RI, potential data products that will be generated and to report on how they were produced, with what methodology, tools, what data source, and how they will be managed both for the purposes of the RI and the wider scientific community.

This DMP is composed of a template that each data product provider should complete for each data product. The template, available in section Template, presents the following sections:

- Description of the data product
- Generation and management of data product
- Data sharing and access- FAIR data product
- Data security
- Allocation of resources

The DMP follows the guidelines of the European Commission and is based on the H2020 templates: Data management plan v1.0 – 13.10.2016.

It is recognised that as the RI develops, other data products will be recognised and developed by RI participants. In such cases the DMP will be updated to reflect these additions using the template.

2) Template of the Data Management Plan for JERICO-RI Data Products

1. Description of data product

1.1. Name

- What is the name of the data product (output dataset)?

1.2. Format and size

- What types (dynamics, fixed…) and formats of data products are produced?
- What is the expected size of the data? If applicable, provide an update frequency.

1.3. Application

- *Give example(s) of the purpose of the data product and example(s) of its applications.*
- *To whom might the data product be useful?*

2. Generation and management of data product

2.1. Description of input data

- *What was the data input used to generate the product? (name of dataset, format, where is it available…)*

2.2. Data processing

- *How were the input data processed?*
- *What tools were used?*
- *What model? Is the code available? Is it open access?*

2.3. Data product quality and standards

- *How were consistency and quality of data generated controlled and documented?*

2.4. Management

- *How data products are stored and managed? What is the role of JERICO-CORE?*
- *Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?*

2.5. Operationality

- *Is the dataset operational or not?*
- *If yes, what is the frequency of updates?*
- *If not operational and if available, provide a Technology Readiness Level (TRL)?*

3. Generation of FAIR data

The following subsections are aligned with FAIR Principles as described in the GO FAIR initiative (https://www.go-fair.org/fair-principles/).

3.1. Making data findable, including provisions for metadata

- *Are the data produced discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?*
- *What naming conventions are followed?*
- *Were search keywords provided to optimize possibilities for re-use?*
- *Is clear version numbers provided?*
- *What metadata has been created? How do they relate to JERICO label? In case metadata standards do not exist in this discipline, outline what type of metadata will be created and how.*

3.2. Making data openly accessible

- *Which data produced were made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.*
- *How were the data made accessible (e.g. by deposition in a repository)?*
- *What methods or software tools are needed to access the data?*
- *Is documentation about the software needed to access the data included?*
- *Is it possible to include the relevant software (e.g. in open source code)?*
- *Where are the data and associated metadata, documentation and code deposited?*
- *Have you explored appropriate arrangements with the identified repository?*

- *If there are restrictions on use, how will access be provided?*
- *Is there a need for a data access committee?*
- *Are there well described conditions for access (i.e. a machine readable license)?*
- *How is the identity of the person accessing the data ascertained?*

3.3. Making data interoperable

- *Are the data produced interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?*
- *What data and metadata vocabularies, standards or methodologies do you follow to make your data interoperable?*
- *Do you use standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?*
- *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, do you provide mappings to more commonly used ontologies?*

3.4. Increase data re-use (through clarifying licences)

- *How are the data licensed to permit the widest re-use possible?*
- *When are the data made available for reuse? If an embargo is sought to give time to publish or seek patents, specify why and how long this applies, bearing in mind that research data should be made available as soon as possible.*
- *Are the data produced usable by third parties? If the re-use of some data is restricted, explain why.*
- *How long is it intended that the data remains re-usable?*
- *Are data quality assurance processes described?*

3.4.1. Regulation of responsibilities of users

- *Indicate whether external users are bound by data sharing agreements, setting out their main responsibilities.*
- *If applicable, write how users should cite the dataset?*

4. Ethical aspects

- *Are there any ethical or legal issues that can have an impact on data product sharing?*

5. Data security

- *What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?*
- *Are the data safely stored in certified repositories for long term preservation and curation?*

6. Allocation of resources

- *Who is responsible for ensuring the data product generation, management, metadata creation, data security and quality assurance?*
- *What are the costs for making data FAIR?*
- *How will these be covered?*

3) DMPs of the existing Data Products as per April 2023

A- HFR Gap filled surface current Advanced Products

B- Glider transport visualization

C- Plankton images from transects in the Ligurian Sea, acquired with an UVP6 on the Sea Explorer glider SEA002, Spring 2021

As described in JERICO-S3 D7.5, the goal of the EcoTaxa D2PTS in JERICO-S3 is to provide access to the EcoTaxa application as a service so that JERICO members can process their plankton and particle data and allow this data to flow to the appropriate EU databases where they become available according to the EU data policy.
Colleagues from the Laboratoire d'Océanographie de Villefranche sur Mer used EcoTaxa to process imaging data they collected with a glider in the context of the North Western Med Sea Pilot Super Site.
The present DMP was completed with information relevant to this specific dataset.

D- Biogeochemical state of coastal areas D2PTS

A-Data Management Plan for JERICO-RI Data Product: HFR

1. Description of data product

1.1. Name

HFR Gap filled surface current Advanced Products

1.2. Format and size

Gap filled surface current fields are produced: east-went and north-south components of the currents. The format is NetCDF. More specifically, each NetCDF file will follow the EuroGOOS standard format for HF-Radar documented at Corgnati *et al*., 2018a: Recommendation Report 2 on improved common procedures for HFR QC analysis, Deliverable D5.14 of the project H2020 JERICO-NEXT.
The size for hourly files for a Long Range HFR is about 50KB.

1.3. Application

High-frequency radar, HFR, is a cost-effective monitoring technique that allows to obtain high-resolution continuous surface currents, providing new insights for understanding small-scale transport processes in the coastal ocean. In recent years, the use of Lagrangian metrics to study mixing and transport properties has been growing in importance (Hernández-Carrasco et al, 2018). Gap-filled surface currents are a prerequisite for the computation of HFR-derived Lagrangian trajectories.

Many scientific applications and societal needs are tackled through a Lagrangian approach in the characterization and simulation of the physical ocean processes: maritime safety, search and rescue, oil-spill mitigation, water quality assessment or marine resources management.

Users of HFR advanced products include meteorology services, search and rescue agencies, governments, and regional and local authorities, as well as private companies working in assessment of coastal water quality, renewable energy, or other environmental services.

2. Generation and management of data product

2.1. Description of input data

Input data are basic Radial HFR products listed in e-JERICO catalogue and available in the main data aggregators.

- the CMEMS product containing Near real time in-situ observations of surface ocean currents (drifters and HFR) for the Global Ocean (INSITU_GLO_PHY_UV_DISCRETE_NRT_013_048) can be accessed here.
- theCMEMS product containing Delayed Mode in-situ Observations of surface (drifters and HFR) and sub-surface (vessel-mounted ADCPs) ocean currents for the Global Ocean can be accessed here.
- the EMODnet Physics ERDDAP HFR catalogue can be accessed here.

2.2. Data processing

The tools for obtaining OMA (Open-boundary Modal Analysis) gap-filled products are available at https://github.com/rowg/hfrprogs. The tools are developed in MATLAB.

(Kaplan, D. M., and F. Lekien (2007), Spatial interpolation and filtering of surface current data based on open-boundary modal analysis, J. Geophys. Res., 112, C12007, doi:10.1029/2006JC003984.)

The European HFR Node is sharing Tools for managing basic products (Radial and Total surface velocity files; instructions available in http://dspace.azti.es/handle/24689/908). A work plan is defined to provide open access codes also for generated advanced products like OMA gap-filled fields.

2.3. Data product quality and standards

Hernández-Carrasco et al, 2018 compared different gap-filled methodologies applied to HFR data.

Since 2020, actions on reviewing Data-Gap filling methods and towards defining a standard methodology are tackled by a specific working group of the EuroGOOS HF Radar Task Team. A standardisation of the quality assessment of Data-Gap filling products is still a pending action.

2.4. Management

Each HF Radar operator is currently responsible for ensuring the data product storage and distribution.

JERICO-RI integrates and gives access to the key resources of the European HFR Node. This Node constitutes the Data Management Team of the European HFR network being responsible to overseeing the day-to-day management of HFR data, ensuring that the data flow from the HFR available sites towards in common format and are made publicly available in different European data integrator infrastructures (e.g. CMEMS, EMODnet-Physics, SeaDataNet), guaranteeing data standardization, the long-term usability, the archival and the quality checking of the HFR data.

The data management of the two pilot D2PTS performed in JERICO-S3 with pilot applications in Bay of Biscay and NW-MED will follow the European common data and metadata model for surface currents from Corgnati *et al.*, 2018a: Recommendation Report 2 on improved common procedures for HFR QC analysis, Deliverable D5.14 of the project H2020 JERICO-NEXT, available in Ocean Best Practice repository.

See also Mantovani, C., *et al.*, 2020. Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement. Front. Mar. Sci. https://doi.org/10.3389/fmars.2020.00210

2.5. Operationality

The two pilot D2PTS performed in JERICO-S3 are not producing operational products but specific datasets for a significant period (hourly between June2016 and Sep2018).

However the operational management of these products is established in the aforementioned recommendations like for the basic HF Radar data. Actually, ongoing efforts for harmonising those downstream services are being performed in the European network.

3. Generation of FAIR data

3.1. Making data findable, including provisions for metadata

The use of DOI is promoted in the European HFR community.

standard_name_vocabulary = "NetCDF Climate and Forecast (CF) Metadata Convention Standard Name Table Version 1.6"

Keywords are provided.

The European common data and metadata model for surface currents is described in Corgnati *et al.*, 2018a: Recommendation Report 2 on improved common procedures for HFR QC analysis, Deliverable D5.14 of the project H2020 JERICO-NEXT, available in Ocean Best Practice repository.

This metadata model established by the European community has been recovered in the JERICO label.

3.2. Making data openly accessible

Each HF Radar operator is currently responsible for ensuring the storage and distribution of the HFR Gap filled surface current Advanced Data Product. However, the European HFR Node, which is already contributing in the

centralised management of the basic data products (input data), is planning to support in the next years a unified and sustained delivery of advanced products like HFR Gap filled surface current fields.

All the resources of the EuroGOOS Task Team, in particular documentation and tools for HFR data gap filling, will be integrated in the e-JERICO catalogue.

The access to pilot D2PTS will be monitored by e-JERICO KPIs.

3.3. Making data interoperable

The European common data and metadata model for surface currents (Corgnati *et al*., 2018a: Recommendation Report 2 on improved common procedures for HFR QC analysis, Deliverable D5.14 of the project H2020 JERICO-NEXT) has been defined in collaboration with the HFR Global network and following the main data aggregators recommendations: EMODnet, SeaDataNet and Copernicus Marine Service.

The ongoing actions of the EuroGOOS Task Team aim to make this kind of advanced products more interoperable like the basic radial and total surface current products.

3.4. Increase data re-use (through clarifying licences)

HF radar sea surface current velocity dataset is licensed under a Creative Commons Attribution 4.0 International License. See http://creativecommons.org/licenses/by/4.0/

3.4.1 Regulation of responsibilities of users

In the metadata "distribution_statement" of the European HFR data model, the following mention is made: "User assumes all risk for use of data".

4. Ethical aspects

N/A

5. Data security

Each HF Radar operator is currently responsible for ensuring the data product generation and the corresponding data security.

6. Allocation of resources

Each HF Radar operator is currently responsible for ensuring the data product generation through its own infrastructure and resources.

B-Data Management Plan for JERICO-RI Data Product: Glider

1. Description of data product

1.1. Name

Glider transport visualization.

1.2. Format and size

The output consists of 2 static figures in png format produced when delay time data is available from the last mission. Each file sizes 250 Kb on average. Matlab files with all derived variables are also produced and saved with a size up to 35 Mb by mission.

1.3. Application

This product is included in the so-called Glider D2PTS that is a demonstrator implemented in the context of the JERICO-S3 project. A complete description of the product and its foreseen application is included in the deliverable D7.5 ("Pilot D2PTS demonstration").

2. Generation and management of data product

2.1. Description of input data

The input data to create the Glider Transport Visualization product, are datasets generated from glider missions in the context of the SOCIB Glider Canales Endurance Line. These datasets are included in the related data product accessible through the SOCIB Data Catalog (see https://doi.org/10.25704/JD07-SV9). Currently, these glider mission datasets are expressed in the SOCIB data format (more information at https://www.socib.es/data/). More specifically, the input datasets are generated after the glider mission ends and the full raw dataset is processed by means of the SOCIB Glider Toolbox. At the moment, the input datasets are L1 (delayed mode data) netCDF files (Troupin et al., 2015) following the SOCIB conventions (more information at https://socib.es/data).

2.2. Data processing

In summary, the process starts with a selection of glider mission datasets available in the SOCIB Data Repository. In a second step, the input data is processed by means of the Glider Transport Toolbox. The latter generates a set of static figures that are accessible through the Transport section at the Sub-regional Mediterranean Sea Indicators web application (https://apps.socib.es/subregmed-indicators) for both the Ibiza and Mallorca channels. The Glider Transport Toolbox also produces some data files containing post-processed and derived variables that are not distributed at the moment. The toolbox code is accessible at GitHub (https://github.com/socib/glider-transport-toolbox). The complete data processing pipeline is described in the JERICO-S3 deliverable D7.5.

2.3. Data product quality and standards

Currently, the product does not follow any standard and no quality statement is in application yet. However, the methodology applied to generate the product is deeply described in Juza, M. and Tintoré, J. (2021) and in Heslop et al. (2012). It is foreseen that in the framework of the Blue-Cloud2026 project, standard data files will be created in

netCDF format. In addition, a quality assurance (QA) and quality control (QC) procedures will be developed in accordance with relevant data policies (e.g. JERICO, Blue-Cloud, SOCIB).

2.4. Management

The static figures are stored in the Data Center Facility in order to be available to the web application. At the moment, the full data pipeline is running automatically and the process is monitored and improved by the scientific responsible at SOCIB. The latter ensures that the product is generated timely and meets the expectations. At the moment, the product management processes does not rely on JERICO-CORE capacities, although it is expected that, in the context of the Blue-Cloud2026 developments, JERICO-CORE will provide the possibility of discovering the glider missions and datasets that will be used as input for the generation of the product.

2.5. Operationality

The product is operationally generated in a timely manner, as soon as a recovery dataset corresponding to a new glider mission, is produced in the L1 processing level. So far, only glider missions included in the SOCIB Glider Canales Endurance Line are considered. Finally, the product is available in open access to the users, by means of the web application. The full product processing pipeline currently is in TRL7 ("system prototype demonstration in operational environment") according to Technology Readiness Levels in use in the EU (see https://en.wikipedia.org/wiki/Technology_readiness_level).

3. Generation of FAIR data

3.1. Making data findable, including provisions for metadata

So far, the product is not discoverable nor findable and it does not meet any of the sub principles framed in the F (i.e. "Findable") of the FAIR principles. One user can indirectly discover the related Glider D2PTS by inspecting the Resource Catalog through the JERICO-CORE platform (https://ui.core.jerico-ri.eu/). Apart from that, the product is not catalogued in any other data/information system.

3.2. Making data openly accessible

So far, the product (i.e. the images) is accessible through the web application SUB-REGIONAL MEDITERRANEAN SEA INDICATORS (https://apps.socib.es/subregmed-indicators) within the section called "Transports". Apart from that, it does not meet any of the sub principles framed in the A (i.e. "Accessible") of the FAIR principles.

3.3. Making data interoperable

The image files are in png format. In this sense there is some degree of interoperability. Apart from that, it does not meet any of the sub principles framed in the I (i.e. "Interoperability") of the FAIR principles.

3.4. Increase data re-use (through clarifying licences)

There is no metadata attached to the data, consequently it does not meet any of the sub principles framed in the R (i.e. "Reusability") of the FAIR principles.

3.4.1 Regulation of responsibilities of users

There is a statement in the footer of the web application warning users about their responsibilities in using the product, currently:

Disclaimer: The content displayed on this website is the intellectual property of ICTS SOCIB. You may not reuse, republish, or reprint such content without our written consent. The information provided in this website is for informational purposes only, and should not be considered as legal advice on any matter. Although the

information on this website has been verified to the best of our activities, we cannot guarantee that there are no errors and mistakes.

To cite the use of the product:

Juza, M. and Tintoré, J. (2021). Multivariate sub-regional ocean indicators in the Mediterranean Sea: from event detection to climate change estimations, Frontiers in Marine Science, 8:610589, doi.org:10.3389/fmars.2021.610589.

4. Ethical aspects

N/A

5. Data security

The data are stored in the Data Center Facility at SOCIB and it is subject to all the security procedures in place at SOCIB.

6. Allocation of resources

See section 2.4 and 2.5 above about ensuring the data product management.

There are no costs of making data FAIR since data is not FAIR as explained in section 3 above.

There are plans, in the framework of the EU Horizon Blue-Cloud2026 project, to implement proper mechanisms to make the data FAIR.

References

Troupin, C., Beltran, J. P., Heslop, E., Torner, M., Garau, B., Allen, J., Ruíz, S. & Tintoré, J. (2015). A toolbox for glider data processing and management. Methods in Oceanography, 13, 13-23.

Juza, M. and Tintoré, J. (2021). Multivariate sub-regional ocean indicators in the Mediterranean Sea: from event detection to climate change estimations, Frontiers in Marine Science, 8:610589, doi.org:10.3389/fmars.2021.610589.

Heslop, E. E., Ruiz, S., Allen, J., López−Jurado, J. L., Renault, L., and Tintoré, J. (2012). Autonomous underwater gliders monitoring variability at "choke points" in our ocean system: a case study in the Western Mediterranean Sea. Geophys. Res. Lett. 39:L20604.

C- Data Management Plan for JERICO-RI Data Product: EcoTaxa

1. Description of data product

1.1. Name

Plankton images from transects in the Ligurian Sea, acquired with an UVP6 on the Sea Explorer glider SEA002, Spring 2021

1.2. Format and size

1.1 million images and associated taxonomic identifications, made mostly by machine learning.
Dataset will be updated with identifications checked by humans.

1.3. Application

Description of mesoscale features and their influence on plankton.
Useful for plankton ecologists and biogeochemists.

2. Generation and management of data product

2.1. Description of input data

Raw image archives output by the UVP6. Stored internally at Villefranche sur mer.

2.2. Data processing

Input data was first processed by UVPapp (https://github.com/ecotaxa/uvpapp ; private repository), the application dedicated to the processing of UVP6 data. They were then post-processed with UVP_toolbox (https://github.com/ecotaxa/UVP_toolbox) to automatically split the data into chunks in the case of its deployment on gliders. Finally, the images and metadata archives hence prepared were imported into EcoTaxa and predicted with a combination of deep and classic morphological features within the application; the model was trained on this dataset https://ecotaxa.obs-vlfr.fr/prj/4591 . Some of these images were imported into Morphocluster https://github.com/morphocluster/morphocluster for batch checking and others were validated by humans in EcoTaxa.

2.3. Data product quality and standards

Metadata coming from the glider was quality checked during its processing. The quality of the prediction in EcoTaxa was assessed by sorting a subset of the data and comparing with the automated prediction.

2.4. Management

The raw data is stored and backed up locally.
The derived data (images of organisms, associated metadata and identifications) is stored in EcoTaxa. Once the manual validation of the identifications will have progressed enough, the concentrations derived from it will be uploaded to EurOBIS.

2.5. Operationality

The workflow for generating the datasets can be considered operational as all its components (ie. software tools and EcoTaxa) are operational.

3. Generation of FAIR data

3.1. Making data findable, including provisions for metadata

The dataset has an identifier and is findable on EcoTaxa.
It does not have a DOI yet; it will get one through the upload to EurOBIS.
The metadata follows the usage for UVP, which has no official standard but is slowly matched to BODC terms and integrated into the Argo standards.
The appropriate BODC matches will be made explicit for the export to EurOBIS which will be formatted according to the DarwinCoreArchive format requirements.

3.2. Making data openly accessible

The data on EcoTaxa is browsable by anyone and in more details when people register on the application. The data on EcoTaxa can be accessed by machines and this is already the case as part of the BlueCloud project.
The data on EurOBIS will be openly accessible to anyone.

3.3. Making data interoperable

See above:
The metadata follows the usage for UVP, which has no official standard but is slowly matched to BODC terms and integrated into the Argo standards.
The appropriate BODC matches will be made explicit for the export to EurOBIS which will be formatted according to the DarwinCoreArchive format requirements.

3.4. Increase data re-use (through clarifying licences)

The dataset is licensed as CC-BY on EcoTaxa and that licence will carry over to EurOBIS.

3.4.1 Regulation of responsibilities of users

The obligations are those of the licence:
T Panaïotis, A Poteau, E Diamond-Riquier, L Courchet, C Catalano, L Coppola, M Picheral, J-O Irisson. Plankton images from transects in the Ligurian Sea, acquired with an UVP6 on the Sea Explorer glider SEA002, Spring 2021.

4. Ethical aspects

No

5. Data security

The raw data and the EcoTaxa data are backed up daily on a distant storage.

EcoTaxa will soon be hosted by a true hosting facility operated by IFREMER which will ensure an even more solid backup mechanism.

6. Allocation of resources

Jean-Olivier Irisson, irisson@normalesup.org

D- Data Management Plan for JERICO-RI Data Product: BGC

1. Description of data product

1.1. Name

Biogeochemical state of coastal areas D2PTS

1.2. Format and size

This thematic service combines near-real-time (NRT) physical, biological and chemical observations from three Voluntary Observing Ship (VOS) lines (Helsinki-Travemünde, Helsinki-Stockholm, Helsinki-Tallinn) and two fixed stations (Utö Island, Keri Island). Currently, data products from two VOS lines and Utö station are available on the web site at https://swell.fmi.fi/hab-info/.

Satellite remote sensing allows detection of some specific marine events, like algae blooms in this demonstration. Often, to verify the event or to provide additional quantitative data for ground truthing, auxiliary information is required. Finnish Environment Institute SYKE has a public service where one can browse and view SYKE's open satellite data, at https://wwwi4.ymparisto.fi/i4/eng/tarkka. In Tarkka service, users can select to view high-resolution and medium resolution satellite data for the Baltic Sea region, and overlay it with other datasets. As another demonstration of data products, FerryBox data from Silja Serenade and Finnmaid can be visualised on top of satellite images.

Both demonstrations are updated at least daily, but often hourly.

Approximate size of FerryBox data is 250 Mb per year, Plankton imaging data from Utö 250 Gb per year, and other observation data from Utö 1 Gb per year.

1.3. Application

In the Baltic Sea, cyanobacteria form Harmful Algae Blooms (HABs) annually during a period from June to August. These HABs impact many of the coastal activities by the GoF, like fisheries and recreation, and they also impact the value of coastal properties. To inform the public on the development of these blooms during summer, Finnish Environment Institute gives out a weekly harmful algal bloom situation review. This demonstration provides near real time data on the cyanobacteria abundance to be used in the weekly algae reviews, and also provides visual display of data to the public and scientists. Services created in this D2PTS were used by SYKE in their weekly national cyanobacteria reviews in 2021 and 2022.

Satellite remote sensing is widely used to detect anomalies in the Baltic Sea, e.g. related to algae blooms, physical phenomena, and river loads. To verify the satellite data, ground truthing measures are required. Often such ground truth is provided during research cruises, with conventional sampling, and the results are available only after tedious laboratory analysis has been done, often after several weeks or months. In this demonstration, we provide an overlay of continuous measurements on top of satellite images to provide an immediate proof of the events. Data products are available for managers, scientists and the public.

2. Generation and management of data product

2.1. Description of input data

Phytoplankton image data, used to create and validate the Convolutional Neural Network model have been published as follows:

- Data set 1: http://doi.org/10.23728/b2share.abf913e5a6ad47e6baa273ae0ed6617a,

Eudat b2share data repository, Record number: ABF913E5A6AD47E6BAA273AE0ED6617A, Data set name: SYKE-phytoplankton_IFCB_2022.

- Data set 2: http://doi.org/10.23728/b2share.7c273b6f409c47e98a868d6517be3ae3,

Eudat b2share data repository, Record number: 7C273B6F409C47E98A868D6517BE3AE3, Data set name: SYKE-phytoplankton_IFCB_Utö_2021

Selection of phytoplankton images from Utö station has been imported to EcoTaxa (ecotaxa.obs-vlfr.fr), as part of EcoTaxa D2PTS, using a project-tag "IFCB Utö 2021 JERICO-RI Gulf of Finland Pilot Supersite".

Besides making data visualisations available at websites, we have created an API for Alg@line transect measurements with metadata description at https://ckan.ymparisto.fi/dataset/%7bA0948CD3-35F3-49B6-B433-A082A47A581F%7d. This API-service contains flow-through water quality measurements starting from 1998, accessible by OGC-compliant WMS and WFS-requests. The service is updated hourly.

2.2. Data processing

Utö station data includes results from imaging, using Imaging FlowCytobot (IFCB). It collects a 5 mL water sample each 20 min and takes images of the particles containing chlorophyll a. Images taken are transferred to hpc-cloud object storage and they are classified using a pre-trained Convolutional Neural Network model. The code implementing the CNN model can be found at https://github.com/veot/syke-pic. Biovolumes of the labelled images are calculated and contributions of main bloom forming filamentous cyanobacteria are calculated in this demonstration. Classified species data is available in near real time, 2-3 hours after images have been acquired. Details of the data processing and modelling are given by Kraft et al. (2021, 2022).

Ferrybox data has been quality checked and labelled according to Jaccard et al (2021)

Utö station data is presented as raw data, averaged per hour.

Jaccard P, Hjemann D O, Ruohola J, Ledang A B, Marty S, Kristiansen T, Kaitala S, Mangin A (2021). Quality Control of Biogeochemical Measurements within Copernicus in situ TAC. CMEMS-INS-BGC-QC. https://archimer.ifremer.fr/doc/00251/36232/

Kraft K, Seppälä J, Hällfors H, Suikkanen S, Ylöstalo P, Anglès S, Kielosto S, Kuosa H, Laakso L, Honkanen M, Lehtinen S, Oja J and Tamminen T (2021). First Application of IFCB High-Frequency Imaging-in-Flow Cytometry to Investigate Bloom-Forming Filamentous Cyanobacteria in the Baltic Sea. Front. Mar. Sci. 8:594144.doi: 10.3389/fmars.2021.594144

Kraft K, Velhonoja O, Eerola T, Suikkanen S, Tamminen T, Haraguchi L, Ylöstalo P, Kielosto S, Johansson M, Lensu L, Kälviäinen H, Haario H and Seppälä J (2022) Towards operational phytoplankton recognition with automated high-throughput imaging, near-real-time data processing, and convolutional neural networks. Front. Mar. Sci. 9:867695.doi: 10.3389/fmars.2022.867695

2.3. Data product quality and standards

Maintenance and daily operations of FerryBox systems and calibration and maintenance of fluorometry sensors have been done according to SYKE internal Standard Operation Procedures, and they are accredited by Finnish Accreditation Service FINAS, https://www.finas.fi/Documents/T003_A44_2023.pdf .

2.4. Management

Silja Serenade and FINNMAID FerryBox data are permanently stored at SYKE server. Utö data is permanently stored at FMI server. Utö imaging data is stored in CSC Allas object storage. Data is managed by the ICT staff of institutes, according to instructions of FINMARI Data Management Policy, lead by FINMARI data manager (FINMARI is FInnish MArine Research Infrastructure, part of national RI roadmap).

So far JERICO-CORE has no role for data storage, management or procedures.

2.5. Operationality

Data sets are operational and updated hourly. However, D2PTS demonstration was finalised in 2022.

3. Generation of FAIR data

3.1. Making data findable, including provisions for metadata

FAIRness of platforms were studied in JERICO-S3 Deliverable 6.7 Evaluation of FAIRness of data of the PSS and IRS related to the Data Management Policy. Ferryboxes are well represented within EMODnet Physics, CMEMS in-situ and SeaDataNet. The Keri and Utö observatories are not yet present in the aggregators.

Some parts of imaging data are findable in EcoTAxa and EMODnet biology.

3.2. Making data openly accessible

Ferrybox data is made openly available as default. Utö and Keri data is not yet openly available as quality control protocols are pending. Image data is not openly available as there are no sustainable repositories for such data available.

3.3. Making data interoperable

Data which is available through EMODnet Physics, CMEMS in-situ, SeaDataNet, EcoTAxa and EMODnet biology are using the requested vocabularies and are interoperable.

3.4. Increase data re-use (through clarifying licences)

FINMARI recommends using Creative Commons By 4.0 international license (CC BY 4.0) in sharing FINMARI Data (including FerryBox, Utö and Imaging data in this demonstration) giving a user of the data right to share, copy, modify, and build upon a work the authors of the data have created, including the commercial use. It includes the requirement that the source of the data and modifications done need to be mentioned.

3.4.1 Regulation of responsibilities of users

When data sets or code are used, users are asked to cite the source of the data and/or code, as given in the documents mentioned in Section 2.

4. Ethical aspects

NA

5. Data security

FerryBox and Utö data is national monitoring data and preserved for long term according to regulations of SYKE/FMI and as noted in their respective data policies.

6. Allocation of resources

Data product generation, management, metadata creation, data security and quality assurance are tasks for SYKE, FMI and TalTech teams working in JERICO-S3.

Full costs for making data FAIR (that part which is not yet FAIR) is unknown as well as it is unknown who will cover such costs. Currently partial funding is applied from the Academy of Finland to improve data FAIRness of such observations.

4. OUTREACH, DISSEMINATION AND COMMUNICATION ACTIVITIES

Nothing to report.

5. CONCLUSIONS AND PERSPECTIVES

The purpose of this deliverable was to develop within the JERICO DS project context, a data management plan for the JERICO added value products. In JERICO-S3 project, four data products (D2PTS) have been developed as demonstrators of JERICO-CORE capabilities. As these products are diverse, we actually proposed a collection of DMPs following the same harmonized template. Indeed as described in JERICO-S3 D7.5, the span of products is large in type and in maturity, some of the D2PTS focus on creation of observation data, while others have evolved models and tools. Thanks to the efforts carried on by the EuroGOOS HFR Task Team, a mature level of homogenization and standardization of operations and products has been achieved by the European HFR Node and community, mainly based on a core of shared best practices, documentation and software tools.
These differences have been a challenge in trying to create a harmonized template that would cover all products, however, once the challenge overcomed it is expected that the template would also cover non-existing yet products.

Based on the content of the 4 examples of DMPs collected, gaps are emerging:
- Because of the level of maturity of the tool EcoTaxa and that a DMPs should be established by each user of the tool creating a new dataset, it was noted that a gap exists between the location of the dataset and JERICO-RI (e.g. the link is somehow not straightforward) that will likely need to be addressed while the Thematic Services are further developing.
- Similarly, as most D2PTSs have not reached their level of full maturity and are still evolving, some sections of the DMPs are left empty with the mention 'Non Applicable' (NA). At the current stage the info is not available indeed, however a gap to address in an ultimate version of the template would be to ensure that a plan or a strategy is envisioned by the provider on how to generate FAIR data products when those are not at this stage yet and to report it in the DMPs.

Finally, as JERICO-RI is further shaping up to possibly become an ERIC, i.e. JERICO-CORE is evolving, the Thematic Services and all other elements of JERICO are maturing, it is anticipated that the gaps identified will be addressed in the preparation phase and that this deliverable will serve as critical material.