D5.2 2nd periodical assessment of iMagine VA services

Abstract

The report provides the second-year usage statistics and assessment of the 5 thematic, AI-powered image analysis services provided under virtual access in WP5.

1

Document Description

D5.1 1st periodical assessment of iMagine VA services			
From: Work Package 5			
Due date	30-09-2024	Actual delivery date:	09-10-2024
Nature of document	Report	Version	1.0
Dissemination level	Public		
Lead Partner	MARIS		
Authors	Dick M.A. Schaap (MARIS)		
Reviewers	Gergely Sipos (EGI Foundation)		
Public link	**https://zenodo.org/records/13907484**		
Keywords	KPI, Virtual Access, Use Cases		

Revision history

Issue	Date	Comments	Author/Reviewer
V 0.1	25/09/2024	First draft for ASB meeting	Andrea Anzanello (EGI)
V 0.2	07/10/2024	Updated based on feedback from reviewer and project ASB	Dick Schaap (MARIS)
V 1.0	09/10/2024	Finalised and submitted version	Gergely Sipos, Andrea Anzanello (EGI)

Copyright and licence info

Table of content

Executive summary

This report provides an assessment at the end of M24 of the WP5 installations provided by the iMagine project under the Virtual Access (VA) mechanism. This assessment is based on the metrics collected from the WP5 installations which have been deployed and become operational during the first period of observation, namely Month 13 – Month 24.

WP5 installations are the Imaging Analysis Services set up by the 5 mature marine imaging use cases from the WP3 development activities:
1. Marine litter assessment service
2. Zooscan taxonomic identification service
3. Marine ecosystem monitoring service
4. Oil spill detection service
5. Flowcam phytoplankton identification

Originally, all five the mature use cases were planned to be fully deployed and operational at M24, meeting milestone M5.4. However, in current practice this applies only for two services: UC1 – Marine litter assessment service and namely UC5 – Flowcam phytoplankton identification service. The other three use cases are well underway and should reach production stage in the coming weeks to months. A consequence is that this report only gives partial metrics for use case 1 and 5, while for use case 2, 3, 4 no metrics are reported. Their Imaging Analysis Services are still under co-development and integration in the Competence Centre (WP3).

The present status and plans for completing the delayed use cases are described in this document together with the reasons of delay. These reasons are dominantly motivated by additional user interface developments and longer data taking – All to make services attractive and accurate for users.

The project will focus on service promotion and user engagement in the 3rd year, targeting specific stakeholders that are relevant for each of the use cases, as it was defined in the "D2.5 Innovation Management and Exploitation Updated Plan"[1] earlier this year.

---

[1] https://doi.org/10.5281/zenodo.11403473

1 Introduction

Virtual Access (VA) is a financial instrument to reimburse the access provisioning costs to access providers. This instrument is provided by the European Commission to increase the sharing of research infrastructures and services that otherwise would not be available to international user groups.

In VA, the services – also called "installations" – must be made available 'free of charge at the point of use' for European or International researchers. VA access is open and free access to services through communication networks to resources needed for research, without selecting the researchers to whom access is provided.

Virtual Access to services of the iMagine catalogue applies to the following 2 categories:

- AI platform and compute infrastructure services in WP4
- Imaging data and analysis service in aquatic sciences in WP5

This document provides Virtual Access metrics and an assessment for WP5 about the 2nd year of the project (Sep 2023 - Aug 2024). However in practice, only one use case has so far been fully deployed and taken into production, namely UC5 – Flowcam phytoplankton identification, while UC1 – Marine litter assessment service has reached near-operational stage. All other mature use cases are well underway. It should be mentioned that in the original planning of the iMagine project, M24 is considered as milestone M5.4 on which all five mature use cases should have reached production status.

1.1 Installations

Within iMagine project 5 installations are part of Virtual Access work package 5. These installations support the baseline computing infrastructure of iMagine as part of the following services:

1. **Marine litter assessment service** (provided by DFKI with OGS and MARIS).
2. **Zooscan taxonomic identification service** (provided by SU)
3. **Marine ecosystem monitoring service** (provided by EMSO, UPC, IFREMER, and MI)
4. **Oil spill detection service** (provided by CMCC and OrbitalEOS)
5. **Flowcam phytoplankton identification** (provided by VLIZ)

1.2 Metrics

For each (near) operational installation several metrics have been defined between the provider and WP5 leader, taking into account following categories:

- Number of users – Number of unique users of the AI image processing service
- Number of images – Number of images processed per year or Names of images ingested
- Number/names of the countries reached – the goal of this metric was to report how broadly the service is used and how the geographical coverage is changing with time.

The metrics are given in Chapter 2 per (near) operational installation. Also, the current planning for the full deployment and going into production of delayed mature use cases is given together with explanations of their delay.

2 Installations

2.1 Marine litter assessment service

Description	This service supports ingestion, storage, analysis and processing of drone images, observing litter floating at surface waters in seas, rivers and lakes, and lying at beaches and shores, delivering standardised classified data sets, which are fit for purpose of environmental management and indicators.
Task	T5.1
URL	**https://www.imagine-ai.eu/service/litter-assessment-service**
Service Category	Thematic
Service Catalogue	Image Analysis Services for Aquatic Sciences
Providers	DFKI, MARIS, OGS
Location	Original service at DFKI – Germany
Modality of access	No remote access currently, but will be once deployed at iMagine platform
Support offered	Support of users and operation, including training of users
Operational since	Will become full operational in 3rd year

User definition	researchers from academics, monitoring agencies, NGO's, environmental management organisations

2.1.1 Metrics

Metric name	Baseline	Define how measurement is done	M13–M24
Number of unique users of the AI image processing service	10	Account management of registered users and image processing runs by iMagine platform	10
Number of images processed per year	1500	Account management of registered users and image processing runs by iMagine platform	1
Names of images ingested	200000	Account management of registered users and image processing runs by iMagine platform	0
Number of countries of users	7	Account management of registered users and image processing runs by iMagine platform	1
Names of countries reached	Germany, Slovakia, BIH, Vietnam, Cambodia, The Philippines, Myanmar	Account management of registered users and image processing runs by iMagine platform	1 (Italy)

2.1.2 Assessment

Direct explanation of the reported numbers:

- **Number of unique users of the AI image processing service**
    - 9 unique cloners on git AND
    - 1 researcher outside of DFKI

- **Number of images processed per year**:
    - 1 orthophoto

- **Number of countries:**
  - 1 researcher from Germany
  - 9 git cloners without information about country (because git does not provide this information)

UC1 plans to have the production environment 1) locally available on git/docker-hub and 2) integrated into OSCAR. Currently option 1) is possible with models as well as the processing methodology around it being available on the newest instance of git (ai4os) and also linked to dockerhub, so users can run the UC1 service locally on their machines after download from git.

In the next weeks we aim to establish option 2) which will provide more user friendly and therefore more attractive usage option with better tracking users. The setup requires an application specific web portal which interacts with the trained model and input data through OSCAR. While WP4 already setup an OSCAR installation, the integration of our service into the OSCAR cluster is something that the DKFI team is currently working on. The service was integrated with the iMagine OSCAR test instance, but the service keeps crashing, due to memory allocation problems. DKFI and UPV are working together to overcome the problem.

2.2 Zooscan taxonomic identification service

Description	EcoTaxa is a web application coupling a database with AI tools to accelerate the labelling of large quantities of plankton images by human operators, who are trained biologists but have no AI expertise. It allows importing images and their metadata, extracting features through deep learning networks, training AI classifiers based on the labelled images in the whole database, interacting with users to confirm or correct those labels, and exporting the resulting data. ZooProcess is the image analytics pipeline of EcoTaxa, for plankton images collected with the ZooScan instrument. ZooProcess will be advanced and ported to the iMagine framework.

Task	T5.2
URL	**https://www.imagine-ai.eu/service/zooprocess-service**
Service Category	Thematic
Service Catalogue	Image Analysis Services for Aquatic Sciences
Location	Currently: distributed, Ultimately: EGI + Villefranche-sur-Mer,France
Modality of access	No remote access currently
Support offered	Training of users, assistance for data formatting for upload, assistance for data exploitation
Operational since	2010
User definition	scientific research groups (mostly used by technicians) + environment monitoring companies/agencies

2.2.1   Metrics

Metric name	Baseline	Define how measurement is done	M13–M24
Number of unique zooscan+zooprocess users	~300 / y	no user tracking is available. We estimate >300 persons, from around the world since 150 ZooScans have been sold worldwide, all use ZooProcess, most are used by more than a single operator.	0

Number of scans (i.e. images) processed per year	~10000 / y	1 scan per working day, 150 zooscans => 10k scans	0
Number of countries in which the current version of zooprocess is used	30	Number of countries in which a zooscan has been purchased	0
Names of countries reached	France, Spain, Italy, Belgium, UK, Germany, etc. in the EU, Brazil, US, Canada, China, Japan, Korea	Countries in which a zooscan has been purchased	0

2.2.2 Assessment

No usage during the period as the service is still under development.

UC2 has three components: two AI components (an image classifier and an image segmenter) and a wrapper for those, which will be the user facing software (ZooProcess). The two AI components will be available from the iMagine marketplace and should be deployed on an OSCAR instance (probably to the OSCAR instance in Ireland). Over the long term, the two AI components could continue to be hosted by EGI and distributed through the iMagine marketplace, assuming that the sustainability of the iMagine platform could be arranged. They could also be hosted locally by users and used for inference. The wrapper (called ZooProcess) is a full web application that will not be hosted by the project services (it does not fit the deployment model) but will be run by each user of the machine taking the images. The AI components are still in development but should be ready for production deployment in the next month, while the UC2 team aims to finish the wrapper by the end of October 2024.

Explanation of the delay: the format of the AI modules for AI4OS has changed (1 repository instead of two, code and dockerisation in the same repo, etc.) so the UC2 team needs to update their work to make it easy to deploy. This is done for the classification module and is in progress for the segmentation one. Both AI modules should be ready by mid October 2024 and the UC2 team hopes that deployment will then be straightforward. These modules will need to be integrated in the wrapper and the whole pipeline should be finished by the end of October 2024.

The planning for achieving production status: the UC2 team expects to be ready with the AI application deployment by mid October 2024 and then for the full pipeline by the end October 2024.

2.3 Marine ecosystem monitoring service

Description	This service will be provided for the processing of video imagery, collected by cameras at EMSO underwater sites, identifying and further analysing interesting images for purposes of ecosystem monitoring. The service will be operated from several EMSO sites where underwater videos are being collected. The three sites of this installation provide complementary capabilities for the whole pipeline of image collection, selection, AI-based analysis and annotation. EMSO-Obsea (UPC – SE) and EMSO-Azores (Ifremer – FR) have experience with using AI for the analysis of selected images for identification of biota. EMSO-SmartBay has experience with preselecting interesting images from the sizable video footage. Ifremer has experience with data management of EMSO raw and annotated imagery.
Task	T5.3
URL	**https://www.imagine-ai.eu/service/marine-ecosystem-monitoring-service**
Service Category	Thematic
Service Catalogue	Image Analysis Services for Aquatic Sciences
Providers	IFREMER, MI, UPC
Location	Vilanova i la Geltru coast (Barcelona, Spain) Lat. : 41º 10,93' N – Long. : 001º 45,15' E

Modality of access	Remote or partially remote
Support offered	support and training
Operational since	O
User definition	Researchers, monitoring agencies, NGOs, environmental management organisations...

2.3.1   Metrics

Metric name	Baseline	Define how measurement is done	M13–M24
Number of user groups / institutions accessing Obsea data (real time or data bank or images and multiparametric data)	10	Account management of registered users and image processing runs by iMagine platform	O
Number of images generated and archived per year on the data bank	3000	Account management of registered users and image processing runs by iMagine platform	O
Number of countries of users	3	Account management of registered users and image processing runs by iMagine platform	O
Names of countries reached	France, Italy, etc	Account management of registered users and image processing runs by iMagine platform	O

2.3.2   Metrics

Metric name	Baseline	Define how measurement is done	M13–M24

Number of user groups / institutions accessing Azores data (real time or data bank or images and multiparametric data)	10	Account management of registered users and image processing runs by iMagine platform	O
Number of images generated and archived per year on the data bank	3000	Account management of registered users and image processing runs by iMagine platform	O
Number of countries of users	3	Account management of registered users and image processing runs by iMagine platform	O
Names of countries reached	France, Italy, etc	Account management of registered users and image processing runs by iMagine platform	O

2.3.3 Metrics UC3s

Metric name	Baseline	Define how measurement is done	M13–M24
Number of user groups / institutions accessing SmartBay data (real time or data bank or images and multiparametric data)	10	Account management of registered users and image processing runs by iMagine platform	O
Number of images generated and archived per year on the data bank	3000	Account management of registered users and image processing runs by iMagine platform	O
Number of countries of users	3	Account management of registered users and image processing runs by iMagine platform	O

Names of countries reached	France, Italy, etc	Account management of registered users and image processing runs by iMagine platform	0

2.3.4 Assessment

No usage for all 3 UC3 cases during the period as the services are still under development.
For each UC3 case the situation will be given.

UC3o – OBSEA

The UC3o team expects to be very close to achieving production deployment, but still issues might appear.

Explanation of the delay: the UC3o team collected their data from July 2023 until July 2024. A yearly period is required to cover the seasonality in the fish population in the area of study. Before July 2023 they did not have high resolution cameras deployed, and the historical data turned out not really useful for new data obtained from the latest cameras. Regarding the deployment of the AI model into production, the UC3o team just needs to do a final training with the complete dataset. Once it is trained it can be easily deployed. Deploying the final AI model is relatively easy, the complex part is integrating it with the rest of their data pipeline (real-time acquisition, ingestion into their database, generation of datasets, data delivery to aggregators…). The UC3o team is still working on these data management aspects. They are also working on the deployment of metric tools to assess the impact, since they did not have them at the beginning of the project.

The planning for achieving production status: the UC3o team expects to achieve full production status before the end of October 2024.

UC3a – Azores

The UC3o team is currently developing the module, so the deployment is progressing well but is slightly behind the original schedule.

Reason of the delay: the UC3a team have spent time fine tuning the model and improving the pipeline for data cleaning. Furthermore, they have needed to clarify different technical questions about the iMagine platform and its usage.

The planning for achieving production status: the UC3a team plans to deploy a first production version during September 2024 and then expects to finalize the module by the end of October 2024.

UC3s – Smartbay

UC3s has 2 of their 3 initial model containers ready for initial deployment, while the container repos were populated last week for:
https://github.com/ai4os-hub/smartbay-species-detection
and
https://github.com/ai4os-hub/smartbay-prawn-burrow-detection

The UC3s team will work on populating the following repo in October:
https://github.com/ai4os-hub/smartbay-dover-vqa

Explanation of the delay: the UC3s team had no significant annotated training datasets available until this summer. They had 2 students helping with Annotating North East Atlantic coastal species imagery and Nephrops burrow imagery over the summer (July and August 2024). The UC3s built their own "On-premise" image annotation system using CVAT for them to use, as there was no hosted image Annotation environment provided by the iMagine project. The students annotated images in CVAT and created a couple of training datasets in Roboflow, that they then used to train YoloV8 models using an onsite GPU environment with Jupyter notebooks. Both students finished at the end of August 2024,

after which the UC3s team has been evaluating the datasets and models they have created and have added 2 of them to the above github repos and edited the template code accordingly. Both of the above Repos are passing the build checks.

The planning for achieving production status: the UC3s team will re-train some of the models in the coming weeks (Increasing training epochs and hopefully supplementing the datasets) and will add additional trained model files to these containers. The "smartbay-marine-species" container currently has a simple "fish type" model, while the UC3s is hoping to add a usable species detection model for at least 36 species using the datasets one of the students worked on. The student working on the "Smartbay-prawn-burrow-detection" datasets was trying to create separate models for the different fisheries areas with the help of some of the Fisheries Scientists, and the UC3s team will see if they can continue this work. They will also update the documentation in the github repos with example images and API call examples to demonstrate some of the functionality. Both the Marine Species and Nephrops burrow datasets will also need to be supplemented and the UC3s team is investigating how this can be an on-going and iterative process. They will also be working on containerising the DOVER Video Quality Assessment(VQA) model, the goal is to have an inference container that could be run on the iMagine platform, that could "score" video quality using the DOVER models and the DeePAAS API. Finally, the UC3s team is also focused on implementing some of the models "on-premise" as "Nuclio" functions so that they can be "called" via CVAT for semi-automating Annotation and also evaluating "Nuclio" for creating an "on-premise" inference service for images and video feeds and logging detections and video quality scores. Therefore, all together it is foreseen that several extra months will be needed for achieving full production status.

2.4 Oil spill detection service

Description	The WITOIL (Where Is the Oil) service will be made available to reconstruct and validate real events such as accidents or illegal oil spills. In its forecasting mode it can be used to predict the impact of pollution on marine ecosystems for maritime authorities responsible for contingency planning and response. The service will be augmented by the further development of AI algorithms for oil spill detection, with the use of the iMagine AI framework. The service will also

	integrate and offer all the scientific products, such as oil spills detected by remote sensing and oil spill model results added value datasets, for further analysis.
Task	T5.4
URL	**https://www.imagine-ai.eu/service/witoil-service**
Service Category	Thematic
Service Catalogue	Image Analysis Services for Aquatic Sciences
Providers	CMCC, OrbitalEOS
Location	Via Augusto imperatore 16, Lecce (Italy)
Modality of access	web interface
Support offered	Support of users and simulations, including training of users
Operational since	0
User definition	researchers from academics, monitoring agencies, NGO's, environmental management organisations

2.4.1 Metrics

Metric name	Baseline	Define how measurement is done	M13-M24
Number of unique users of the AI image processing service	100	Account management of registered users and image processing runs by iMagine platform	0
Number of images processed per year	1000	Account management of registered users and image processing runs by iMagine platform	0

Number of images ingested	200	Account management of registered users and image processing runs by iMagine platform	O
Number of countries of users	10	Account management of registered users and image processing runs by iMagine platform	O
Names of countries reached	France, Italy, etc	Account management of registered users and image processing runs by iMagine platform	O

2.4.2   Assessment

No usage during the period as the service is still under development.

UC4 is close to deploying an inference ready model at the iMagine marketplace. Besides the deployment in the iMagine platform, the UC4 team has also set up a production environment on their own premises of three institutions. For Oil Spill recognition in satellite imagery, Orbital EOS is responsible. For elaboration and maintenance of the data repository, University of Trento is responsible. For oil spill AI enhanced simulations, CMCC is responsible. For CMCC side, currently a back-end and front-end have been deployed at: **http://witoil.cmcc-opa.eu/**

Explanation of the delay: during the past few months a lot of energy was spent into producing a user interface to make the service user friendly on the platform. By the end of June 2024, the UC4 team checked with the WP3 leader the feasibility of implementing this container into the iMagine platform which unfortunately was not technically possible. Therefore, the UC4 team had to roll back, and create an inference ready system to be used. This took another 2 months.

The planning for achieving production status: the inference model is already in production at the own premises of UC4 partners, full deployment at the iMagine platform is expected by the end of October.

2.5 Flowcam phytoplankton identification

Description	This service supports image-based taxonomic identification of plankton particles in the micro-plankton size range (including photosynthetic plankton or phytoplankton). As part of the service a long-term dataset of over 1.4 million expert validated plankton images is being made available to serve as a high-quality training dataset for new AI supported plankton identification. Enabling access to both the dataset and the analytical algorithms from the iMagine AI framework will support additional users to set up and run their own plankton image recognition pipeline.
Task	T5.5
URL	**https://www.imagine-ai.eu/service/phytoplankton-identification-service**
Service Category	Thematic
Service Catalogue	Image Analysis Services for Aquatic Sciences
Location	Original service at VLIZ – Belgium
Modality of access	No remote access currently, but will be once deployed at iMagine platform
Support offered	Support of users and operation, including training of users
Operational since	0
User definition	Single researchers, environmental management organisations

2.5.1   Metrics

Metric name	Baseline	Define how measurement is done	M13–M24
Number of unique users of the AI image processing	5	Account management of registered users and image processing runs by iMagine platform + internal and affiliated user count	7

service			
Number of images processed per year	300,000	Account management of registered users and image processing runs by iMagine platform + internal and affiliated user count	2,075,958
Number of images ingested	1,400,000	Account management of registered users and image processing runs by iMagine platform + internal and affiliated user count	2,075,958
Number of countries of users	1	Account management of registered users and image processing runs by iMagine platform	7

2.5.2 Assessment

The service of UC5 has been fully deployed on the new iMagine platform since 12/07/2024, previous working version date back to september 2023. Considering the numbers reported:

- Number of users served: serving of 2 internal and 3 affiliated users, namely 1 postdoc, 3 PhD students, associated to the LifeWatch and Jerico S3 projects, and 1 master student, for processing images (lab device use, data pipeline use and model use for inference) and using the iMagine platform in STEM classes using a PlanktoScope device.
- Number of images processed/ number of images ingested: 1,865,953 (RP1 – Month 13–18) + 210,005 (RP2 – Month 19–24), these are coming from the monthly LifeWatch phytoplankton monitoring in the Belgian part of the North Sea and images coming from internal and affiliated users mentioned above.
- Number of countries: 7 countries used the dummy phytoplankton training set in RP1 in a demo session during the iMagine competence Centre workshop. They were coming from France, Spain, Belgium, Ireland, Netherlands, Italy and Germany. In RP2 the users came only from Belgium.

The UC5 team is busy with drafting a manual which should make it easier for future users to make use of the module. The UC5 team is also working on a manuscript for the flowcam use case that they would like to submit at the end 2024 to a suitable scientific journal. The UC5 team has already promoted the iMagine use case and its platform at a number of events and conferences this year. They expect that the paper will attract additional users.