

## Annex 8.2

### Technical offer. Part II

#### Methodology and organization of the work and resource

EGI/2023/OP/0001

#### 1. Methodology

Explain in detail the methodology to be followed for the development and execution of each of the activities included in the tender.

Activities or tasks to be carried out	Methodology to be implemented.
A0 Kick-off meeting	<p>Like any typical IT project, should start kickoff meeting which is first step towards getting all planned IT tasks/goals done right. Under kick-off meeting all necessary goals will be discussed and as of result of discussion projects assumptions, goals, members of teams responsibility, guidelines on expectations and project risk will be taken up.</p> <p>At kick-off meeting it will be ensured that every member of the project team understands what they will work towards and it should be decided who will handle what tasks and agree with deadlines.</p> <p>Every team project member must find out in which way should document task progress.</p> <p>Furthermore during the kick-off meeting, the participants of the project should commit to the guarantee to maintain collaborative work with EOSC DIH and the PULS scientific and research team.</p>
A1 Development of tools for data transfer and storage	<p>Depends on the chosen camera type/company/platform different approach to image/data acquisition will be considered. In this project by default 2 camera types will be taken into consideration:</p> <ol style="list-style-type: none"><li>1. Camera type 1 (low resolution) e.g. Dahua IP 5Mpx, Dahua Technology, China</li><li>2. Camera type 2 (high resolution) e.g CropView 10Mpx from Pessl Instruments (Austria) or 12 MPx Dahua from Dahua Technology (China)</li></ol> <p>Generally two different method authentication will be implemented:</p> <ol style="list-style-type: none"><li>1. HMAC (Keyed-hash message authentication code), is widely used and easy to implement based on private/public key</li><li>2. OAuth 2.0, is an well-known open standard for authorizing client applications. This method protect user credentials because is not used instead of client ID and a client secret is used.</li></ol>

<p>A1.1 Camera data transfer type 1</p>	<p>In this type of camera there is no direct access to camera but a dedicated platform should be used via fieldclimate.com. It required to sign up and add the following information:</p> <ol style="list-style-type: none"> <li>1. Station ID</li> <li>2. Station key</li> <li>3. Station name</li> </ol> <p>After that based on RESTAPI documentation which is exposed by <a href="https://api.fieldclimate.com">https://api.fieldclimate.com</a> images/data consuming can be implemented.</p> <p>All API functionalities can be implemented by standard HTTP methods such as GET/POST/DELETE, etc.</p> <p>There will be implemented at least the following functionalities from camera type1:</p> <ol style="list-style-type: none"> <li>1. Read station information</li> <li>2. Get current camera status</li> <li>3. Get list of sensors of a camera</li> <li>4. Last station events</li> <li>5. Station communication history from to filter</li> <li>6. Last amount of pictures</li> <li>7. Retrieve photos for specified period</li> <li>8. Get last photos</li> <li>9. Get image/snap from SD card</li> </ol> <p>The above functionality allows to connect to many different stations (camera or even weather station at dedicated platform fieldclimate.com Pessl Instruments GmbH).</p>
<p>A1. 2 Camera data transfer type 2</p>	<p>This type of camera belongs to standard IP camera with direct access to image/data based/settings on IP of camera.</p> <p>Usually there are two standard of authentication methods:</p> <ol style="list-style-type: none"> <li>1. basic authentication</li> <li>2. digest authentication.</li> </ol> <p>Most of request to this type of camera will be based on standard Get Request with parameters as of query string. If the request is successful, the IP camera will return a HTTP header contains status 200 OK. The HTTP Body response will contain actual data or error message if an error occurs.</p> <p>The format on that communication is based on standard GET Form, whereas response will be in form of text/plain.</p> <p>The standard functionalities for that type o IP camera will be the following:</p> <ol style="list-style-type: none"> <li>1. Management of configuration and view camera settings</li> <li>2. Management of eventhandler</li> <li>3. Management of alarms</li> <li>4. Management of records and snaps.</li> <li>5. Get current image/snap</li> <li>6. Get image/snap from SD card</li> </ol>
<p>A1.3 Image database</p>	<p>All data will be stored at postgresql engine supported with high available features.</p> <p>To facilitate the maintenance of a large database and ensure consistency - the original images will be stored on disk and its relative path in database. It will avoid discrepancies and further problems. Any</p>

	<p>additional information like: ROI, labels, will be also stored in database in form of CSV/JSON, etc. format.</p> <p>Apart from standard planned functionalities the following additional features in the system will be applied:</p> <ul style="list-style-type: none"> <li>a) images deduplication using e.g. Structural-Similarity-Index-SSIM</li> <li>b) over or underexposed images detection (as needed)</li> <li>c) images anomaly detection</li> </ul> <p>Stored information like: path to the file, coordinates of ROI, segmented area, annotated label, etc. will be served by desired implemented methods. Data serviced in this way can be used for training, validation and testing the model as well.</p>
A1.4 Backend for transfer and storage management v.1	<p>The main web application will be developed in Java language which will implemented API calls for camera type 1 and type 2.</p> <p>For different camera type and specifications, different image types resolution are supported. Original image files will be stored in data storage (e.g. disk/ data array) and its relative path will be placed in the database.</p>
A1.5 Manual (v1 version) for integration of type 1 and type 2 cameras into the system	<p>On the basis implemented API calls for at least 2 camera types, detail documentation will be provided with examples.</p>
A2 Development of the tools for data processing and visualization in GUI	<p>It is assumed that web application developed in Java will be responsible for data presentation and visualization whereas any data processing and potential image generation will be performed on Python backend side. Open source library for data presentation and visualization in Java will be proposed.</p> <p>According to requirements, it is planned to use image automatic pre-processing operations like:</p> <ol style="list-style-type: none"> <li>1. histogram equalization,</li> <li>2. contrast, brightness and sharpness enhancement</li> <li>3. deblurring, filtering/denoising (median, gaussian, etc)</li> </ol> <p>According to each stored image, the developed tools will allow to show:</p> <ol style="list-style-type: none"> <li>1. Original image,</li> <li>2. marked areas annotation, GT masks (if any)</li> <li>3. Image segmentation and detection results (from machine learning models)</li> </ol>
A2.1 Deployment of the final solution to production environment	<p>Project source code will be placed in repository system based on GIT e.g. GitHub, GitLab, etc. It allows to speed up deployment of current and final solution. The suitable CI/CD -Continuous integration (CI) and Continuous Delivery (CD) automatization can be considered.</p> <p>Finals machine learning models will be saved in their native corresponding format (e.g.: PyTorch - pickle, TensorFlow - HDF5, sklearn – joblib). There exists also possibility to store these models on Gitlab LFS (Large File Storage) repository.</p>
A3 Development of models for phenological and AI analysis	<p>Generally machine learning models will be prototyping with Tensorflow and Pytorch frameworks using Python language and will be exposed for web application based on RESTAPI approach.</p> <p>In this project machine learning pipelines and the MLOps workflow will be proposed. The idea is to decrease the gap between prototype and production machine learning process and prepare simple production-ready MLOps pipeline using Metaflow, an open-source</p>

	<p>framework that allows data scientists to build production-ready machine learning workflows via a simple Python API, and Seldon. It lets to train and deploy your models locally and then move (even scale) to production with a “single click”.</p> <p>In this project framework that allows almost any set of machine learning models to be deployed, scaled, updated, and monitored can be considered in form of Seldon ecosystem. Moreover Seldon Tempo (open source) Python SDK can be applied. It allows to describe inference components in Python, test locally on Docker and deploy to Kubernetes if required.</p> <p>Technically, from ML coding point of view the pipe will be consisted of:</p> <ul style="list-style-type: none"> <li>- image preprocessing operations (“transform” methods in sklearn terminology)</li> <li>- generating inferences from AI models (“predict” and “predict_proba” in sklearn terminology), but the core models will be embedded in TensorFlow/Pytorch frameworks.</li> </ul>
A3.1 Collection of data for modelling	<p>The acquisition of data for modelling is one the most important tasks in machine learning model development. Mainly in supervised models so proper ground truth will be necessary for collected data (images). It means that it will be required to extract ROI from images either manually or automatically manner as well other information (label, segmented mask). To create ground truth the labeling process should be applied (possibility to marked object by: polygon, circle, rectangular, elliptical and polyline region shape, labeling, etc.)</p>
A3.2 Development of models in laboratory version	<p>The first step in machine learning development is exploratory data analysis. This step takes approximately 30% of the total ML development. This step is the most important because the result of that process has directly impacted on the quality of the input data used to train the model. At this stage, the calibration of image preprocessing will take special place. This will also be the moment to decide about outlier removing, image transformation etc. The outliers lead to an increase the bias standard deviation, which can be eliminated. There can be used here in addition to approaches based on SSIM (Structural Similarity Index Measure) a standard outlier removing algorithms like One-Class SVM, Local Outlier Factor, Isolation Forest (whereas an input the embedding from the image will be gives – it can be generated for example from autoencoder).</p> <p>Generally in this project deep learning image segmentation/ detection/classification algorithms mainly will be used. On this area one of the most efficient networks architectures (based on CNN networks) will be analyzed:</p> <ol style="list-style-type: none"> <li>1. U-NET-like network to segmentation</li> <li>2. R-CNN for object detection and segmentation,</li> <li>3. Mask R-CNN for object detection and segmentation,</li> <li>4. Yolo - for object detection and segmentation</li> <li>5. DeepLabv3 - for object detection and segmentation</li> </ol> <p>Training will be conducted in typical way: using batch partitioning data, and data (as well as GT masks) augmentation (randomized crops, rotation).</p> <p>To calculate the predictive power of machine learning models cross</p>

validation technique should be applied. As a good practice in machine modeling the whole dataset (numerical features extracted/generated based on ROIs) should be divided into 3 subset:

1. Training subset
2. Validation subset
3. Test subset

The standard workflow for cross validation technique can be introduced as follows:

1. Training subset will be used to train a candidate models
2. Validation subset will be used to evaluate the candidate models one of the candidates is chosen
3. The chosen model is trained with a new training dataset
4. The trained model is evaluated with the test subset
5. The final predictive power of the model is calculated based on test subset e.g. accuracy metric, mIoU ,etc.

In practice during training, one more set is separated (from training data, sometimes also called validation) to control the degree of overfitting

In this project the method of cross validation will applied:

1. k-Fold Cross Validation - dataset is split into k number of subsets (called as folds) then training process is performed on all the subsets but leave one(k-1) subset for the evaluation of the trained model. In this technique, assumed is iterate k times with a different subset reserved for testing purpose each time.

At the end of every cross validation process computation of model performance is performed in form of well-know metrics. The model performance metrics will decide on final model selection.

Depends on which machine learning algorithm will be chosen the metric which can be applied are the follow:

- I. Metric for classification, based on confusion matrix:

1. accuracy,
2. precision,
3. recall,
4. F1 score ,

- II. Metric for classification, based on continuous outputs (eg in neural network after last layer):

1. AUC
2. ACC

- III. Metric for object detection/segmentation:

1. MIOU for segmentation
2. ACC

Next step of machine learning modelling is model hyperparameters tuning. Of course number of hyperparameters depends on machine learning algorithm which will be chosen. In the project can be considered to use Grid-search or Bayesian Search method to find the optimal hyperparameters (e.g. learning rates, number of neurons, drop out, regularization, etc.) of a model.

To avoid overfitting (too sensitive model, unable to generalize) or underfitting (very generalized model) of model, using k-fold cross validation technique can be applied together with hyperparameters tuning.

If overfitting machine model issue will occurred, the following

	<p>technique can be applied:</p> <ol style="list-style-type: none"> <li>1. increase training data sample size</li> <li>2. reduce number of features that avoids complexity</li> <li>3. apply regularization technique</li> </ol> <p>Additionally to increasing the ability to generalization introducing a randomized connection will be examined.</p> <p>Otherwise to avoid underfitting the following techniques can be applied:</p> <ol style="list-style-type: none"> <li>1. increase model complexity in architecture</li> <li>2. add more features</li> </ol> <p>Finally based on the above applied approaches the proper machine learning model will be chosen with optimum performance.</p>
A3.3 Development of models in prototype version	<p>Machine learning models in prototype version should be maintain in production environments. In many commercial approach (enterprise scenario) the common challenge is that machine learning models developed in a laboratory environment remain in the proof-of-concept. When the model is deployed in production, it becomes outdated due to frequent changes in source data that require the model to be rebuilt. As models are retrained multiple times, it is necessary to keep track of the model's performance and the appropriate features and hyperparameters used to retrain the model. To perform all these operations, a well-defined, reproducible process should be in place to implement the end-to-end machine learning operations called MLOps that keep the model up-to-date and accurate in the production environment.</p> <p>It is very important that the model is re-trained at regular intervals, e.g., fortnightly, monthly, quarterly, or as needed, since it is very likely that the underlying source data in the real world will change over a period of time. A cron job is scheduled to retrain the model at the predefined intervals, or when the source data changes, or when the performance of the model degrades. During the retraining of the model, there may be some changes in the model code due to the tuning of the hyperparameters.</p> <p>The typical automated model pipeline which will be considered in production environments may consists of 3 types of stores/stocks:</p> <ol style="list-style-type: none"> <li>1. data (image) preprocessing, standardization,</li> <li>2. the metadata store,</li> <li>3. model registry</li> </ol> <p>The metadata store is a centralized model tracking system maintained at the enterprise level that contains the model metadata at each stage of the pipeline. The model metadata repository facilitates the transition between model stages, e.g., from staging to production to archiving.</p> <p>In this project 2 approaches to call machine learning model and receive response in production environments can be considered:</p> <ol style="list-style-type: none"> <li>1. batch inference mode – can be scheduled as job that runs at a specified time interval and emails the results to designated users.</li> <li>2. online inference can be achieved by deploying the model as a REST API using frameworks such as the Python Flask library or the uvicorn/FASTAPI library for developing interactive web applications and invoking the model through its HTTP endpoint.</li> </ol>

A3.4 Deployment of the final solution in the EOSC Marketplace	Thanks to placing project source code in repository system based on GIT, it will be easier to deploy the whole project at EOSC marketplace as software type under Apache license 2.0.
A3.5 Smooth transfer	The whole code will be shared using Git repository e.g. GitHub, Gitlab and it allows for smooth transfer to EOSC marketplace.
A3.6 Service ticket system	To maintain support for users/customers a dedicated ticket system will be deployed in form of osTicket.

Note: It is possible to add as many cells as necessary as long as they are of importance for the methodology.

## 2. Organization of work and resource

Resource	Organization of work	Activities or tasks to be carried out
Product Owner/Scrum Master	<p>The diagram shows a Scrum board with two sprints. Sprint Week 1 (top) has a vertical bar on the left. The tasks are: Monday (Sprint Review), Tuesday (Daily Stand Up), Wednesday (Daily Stand Up), Thursday (Daily Stand Up), Friday (Daily Stand Up). Below these are Sprint Retrospective and Sprint Planning. Sprint Week 2 (bottom) also has a vertical bar. The tasks are: Monday (Daily Stand Up), Tuesday (Daily Stand Up), Wednesday (Daily Stand Up), Thursday (Daily Stand Up), Friday (Daily Stand Up). Below these are Backlog Refinement, Sprint Review, Sprint Retrospective, and Sprint Planning.</p>	Taking care about all requirements of the project to be fulfilled. Reporting specific needs to the rest of the team. Tight collaboration with PULS representatives responsible for the substantive content of the project. Getting feedback from the team.
Senior Backend Developer		The development of the backend part of the software solution, including communication with AI machines. Java programming.
Senior Frontend Developer		The development of the frontend part of the software solution. Initial UI design, having in mind that the software is dedicated for scientists. React Native programming.
Senior IoT Developer		Working closely with cameras and their APIs. Elaboration and development of interfaces for gathering data from cameras and from auxiliary sources, like public weather data.

Artificial Intelligence Expert	and Sprint Review to get accurate information about the progress of the project and give remarks about the right directions. Product Owner will take care about milestones to be achieved on time. According to the methodology of Scrum Agile for 2-week sprints the organization of work for the whole team is an on the picture above:	Elaboration of two referential models. Reporting the needs for data for the models. Models training and fine tuning. Output preparation for the presentation layer. Python programming.
Tester	<ul style="list-style-type: none"> <li>• First sprint planning includes kick-off meeting; sprint planning usually takes 2 up to 4 hours and is devoted to select and evaluate tasks to be done within the forthcoming sprint</li> <li>• Daily stand ups take no more than 15 minutes; this is to report the progress, get information about obstacles, reorganize tasks within the sprint if it is needed</li> <li>• Backlog refinement usually takes less than an hour and its goal is to remove tasks that can't be done within the sprint and/or add other tasks which are necessary</li> <li>• Sprint Retrospective is to gather remarks, proposals for better work in forthcoming sprints</li> <li>• 5-month project will have 8-9 sprints (there is a gap for vacation time within this 5 months)</li> </ul>	Testing scenarios will be elaborated. Running tests according to the predefined schedule. Fine tuning testing scenarios. Documenting the results on TestRail platform. Reporting the results on Confluence platform.

Note: It is possible to add as many cells as necessary as long as they are of importance for the methodology.

### 3. Quality control measures

Concept	Availability and offered	Description (Please, explain briefly the way of implement the concepts mentioned).
Implementation of a quality system	YES	<p>The tenderer will use scrum agile technology for this project management. The team is relatively small (6 people) and this is the proven methodology for short projects with no big teams. To manage planning, progress, status and risk of work Jira platform will be used. To document the whole project Confluence platform will be used which can be integrated with Jira.</p> <p>Moreover tenderer ensure 3 performance indicators:</p> <ol style="list-style-type: none"> <li>1. Transmission and recording of images from type 1 and type 2 cameras, working in two ecosystems for a minimum of 1 month with a minimum frequency of 7 images per day will have less than 2% error.</li> <li>2. Pheno and AI result indicators from type 1 and type 2</li> </ol>



		<p>cameras working in two ecosystems for a minimum of 1 months will have no less than 3 pheno results, no less than 2 AI results</p> <p>3. Service Level Agreement (SLA level) will be at 99.5 % level. It means that periods of allowed downtime/unavailability will be no more than 3h 37m 21s monthly</p> <p>For testing scenarios and tests runs TestRail platform will be used which can be integrated with Jira.</p> <p>Jira will be used to generate team project timesheets for each project member. The reporting will be completed by next Monday EOD every two weeks with the use of Jira (two weeks long sprints).</p> <p>Planning will be done every two weeks before the sprint starts.</p> <p>The tenderer has ISO9001:2015 standard implemented and follows the rules in this quality management certificate for software development. The tenderer has Dekra's certificate confirming this.</p>
Risk management and continuity of the service in case of absence of the member of the team dedicated to a particular task	YES	<p>We have selected a team of 6 people to develop the project within this tender. However, the tenderer has more developers (at least 3 more) who can substitute developers dedicated to the project and more testers (at least 2 more) that can substitute the tester assigned to this project. Similarly, some of team members can play alternative roles if it is needed. Senior IoT Developer can be a product owner, one of developers can play AI Expert role. Thus the arrangement of the team is flexible, according to the rules of scrum agile methodologies.</p>
Measures to ensure the performance and maintenance of the software for at least 12 months after the end of the contract.	YES	<p>The platform will be maintained on the servers owned by the tenderer. The tenderer has fully redundant server room with enough storage and calculation power, properly secured with the use of Fortigate solutions. The tenderer has 2 independent optical fibers from two different internet providers with LTE backup line in case of cease fiber channels internet, three-stage power supply. The tenderer provides 24x7x365 services from this server room for such customers like: Pepsico, Paul Klacsko, Samat, Orlen Transport.</p>
Measures to ensure compliance with the data protection regulation.	YES	<p>The Product Owner has the leading auditor for ISO27001:2017 certificate from Dekra. It means we know (and use) the rules of information security management systems. The tenderer plans to implement ISO27001 by the end of 2023 and get this certificate.</p>

*Place and date:*

*Głogów Małopolski, 15<sup>th</sup> March, 2023*

*Name (in capital letters), function, company and signature*

*KRZYSZTOF BOBRAN, Chairman of the Board, Seth Software sp. z o.o.*


 Podpisano przez/ Signed by:  
 KRZYSZTOF  
 BOBRAN  
 Data/ Date: 15.03.2023 21:39  
