Ref. Ares(2021)6699155 - 29/10/2021

This project is funded by the European Union
G.A No 101016906

Project Number: 101016906
Start Date of Project: 2021/01/01
Duration: 48 months

Type of document D2.3 – V1.0

Dimensioning of the Real-World Scenario for Final Demo

Dissemination level	PUBLIC
Submission Date	2021-10-29
Work Package	WP2
Task	T2:5
Type	Report
Version	1.0
Author	Andrea Gasparri
Approved by	CONSORTIUM

DISCLAIMER:
The sole responsibility for the content of this deliverable rests with the authors. It does not necessarily reflect the opinion of the European Union. Neither the Directorate-General for Communications Networks, Content and Technology, Artificial Intelligence and Digital Industry nor the European Commission are responsible for any use that may be made of the information

This project is funded by the European Union
G.A No 101016906

Executive Summary

This document comprises a detailed description of the dimensioning of a real-word (1:1) vineyard, which will be used for the experimental validation of the collaborative paradigm for human workers and multi-robot teams proposed within the project CANOPIES.

Furthermore, this document describes the farming operations to be considered for the final demo as well as benchmarks for validating the robot solutions and the required functions. The following aspects are considered:

1. Size of the vineyard, varieties of table grapes, human workers required to work per hectare: Identification of the areas and respective varieties of table grapes as well as details of the human hours per hectare required to carry out each activity.
2. Description of the agronomic interventions: Identification of appropriate agronomic activities and their decompositions in sub-activities.
3. Benchmarks: Identification of appropriate benchmarks for validating the main robot solutions and functionalities of the overall system.
4. Final demo: Definition of the final demo for validating the methodologies developed within the project CANOPIES.

This project is funded by the European Union
G.A No 101016906

Table of Content

1 Dimensioning of the experimental site 5
1.1 Size of the vineyards 5
1.2 Varieties of table grapes 8
1.3 Human workers required to work per hectare 11
2 Description of the agronomic interventions 13
2.1 Harvesting 13
2.1.1 Cluster detection 13
2.1.2 Quality assessment for selective harvesting 13
2.1.3 Grape Collection 16
2.1.1 Cluster Placing 16
2.2 Pruning 17
2.2.1 Canes detection 17
2.2.2 Canes and bud selection 17
2.2.3 Canes cut 18
3 Benchmarks for validation 20
4 Final Demo: Integrated System Field Validation 27

This project is funded by the European Union
G.A No 101016906

Abbreviations and Acronyms

WP	Work Package
VR	Virtual Reality
TSS	Total Soluble Solids
TA	Titratable Acidity
MAE	Mean Absolute Error
CNN	Convolutional Neural Network
RoSAS	Robotic Social Attributes Scale
BPM	Business Process Management

This project is funded by the European Union
G.A No 101016906

1 Dimensioning of the experimental site

1.1 Size of the vineyards

The site identified to experimentally validate the capabilities of the integrated system developed within CANOPIES is an agricultural cooperative named "Società Cooperativa Agricola Corsira", a community of roughly 20 small producers located in the municipality of Aprilia, Lazio, Italy.

Figure 1 provides an overview of the selected area where the Corsira Soc. Coop. Agricola headquarter is highlighted as well as the selected experimental field. Briefly, an area within the Corsira headquarter has been dedicated to the project needs serving both as a warehouse for storing the robotic platforms and as an on -field lab for the project personnel to carry out their activities.



Figure 1 - Overview of the Corsira Soc. Coop. Agricola headquarter along with the selected Experimental Field

Figure 2 provides several information regarding the size of the experimental field along with logistical information concerning the distance from the warehouse. In particular, the yellow line represents the path that must be followed by the robot every single time an experimental activity must be carried out, that is approximately 330 m. In addition, the experimental field that has been provided by the Corsira Soc. Coop. Agricola to carry out the experimental activities is composed of two vineyards with approximately size 114 m x 51 m (0,58 ha) and 122 m x 48 m (0,58h), respectively. This experimental field has been selected among all the possible options by considering relevant agronomics aspects for the validation of the research activities as well as other logistical aspects such as the distance from the warehouse to make the overall experimental setting reasonably practical.

This project is funded by the European Union
G.A No 101016906



Figure 2 - Experimental Field Dimensioning along with information concerning the path from the warehouse

Vineyards are structured as a traditional trellis system “Tendone” with a wide distance between each plant, 3m x 3m. Plantations are all older than 3 years and so in full production and health, thus representing a typical working condition for the validation of agronomic activities such as fruit

This project is funded by the European Union
G.A No 101016906

harvesting or vine pruning. The soil where the vineyards are located can be considered standard for table grapes production, with a lower percentage of stones and higher percentage of sand, that will allow evaluation under most weather conditions, even after heavy rain, thanks to the excellent draining capacity of this type of soil. All structures are traditionally covered with plastic and net to protect grapes from rain and hail. The average extension of each plot is around 1 hectare and dimensions (length and width) are on average between 25 m to 50 m according to plot extension and geometry.

Figure 3 and Figure 4 provide an illustration of a representative table-grape vineyard cultivated with pergola trellis system at the Corsira cooperative from inside and outside the tendon, respectively.



Figure 3 - Example of Table-Grape Vineyard cultivated with Pergola Trellis System at the Agricultural Cooperative Corsira.



Figure 4 – View from outside the tendon of the Table-Grape Vineyard at the Agricultural Cooperative Corsira.

1.2 Varieties of table grapes

The selected vineyard in Aprilia has currently four different table grape varieties which are described in the following: Pizzutello Bianco, Pizzutello Nero, Red Globe and Black Magic.

Pizzutello Bianco is a seeded variety that comes from Siria where it is called Damasco Bianco. The harvesting time ranges from the middle to the end of September (16-17 brix). The leaf's shape is pentagonal, while the bunches are conical and cylindrical (200-800g). Regarding the berries, they are characterized by green to yellow color and are horn shaped with soft peel. Pizzutello Bianco has a good fertility and good resistance to diseases and transportation. Nowadays, the cultivation is spread in Puglia, Sicily and Lazio.



This project is funded by the European Union
G.A No 101016906

Pizzutello Nero (Lattuario Nero) is a black seeded variety which has been cultivated in Italy for several years. It is characterized by strong peel and resistance to diseases and transportation. The harvesting time is end of September - begging of October (17-18 brix). Bunches are typically big (up to 1kg), conical and cylindrical. Pizzutello Nero has a good fertility and excellent flavor.



Red Globe, developed by University of California of Devis, is a red variety well known all over the world. Its recognized qualities are big bunches, which can reach up to 2 kg, large berries (24-32mm), good fertility, good resistance to disease and environmental conditions and long storage times. The harvesting time begins in the middle of September until the end of October (15-17 Brix).



Black Magic, developed by Viticulture and Enology institute of Chisinau, is an early black variety widespread in Puglia and Sicily. It is characterized by conical bunches that can reach up to 800g, medium berries (20-24mm), good fertility and short storage times. The harvesting times is July until the beginning of August (15-17 Brix).



This project is funded by the European Union G.A No 101016906



Figure 5 - Overview of the table grape varieties in the experimental site.

Figure 5 shows how the table grape varieties are placed on the experimental site. Most of the rows of the experimental site are dedicated to Pizzutello Bianco and Pizzutello Nero, while Black Magic only occupies six rows and Red Globe is present on 12 rows. The heterogeneity of the test plot will give substantial benefits:

- Red Globe and Pizzutello Nero have colored skins that will help algorithm in the cluster detection.
- Pizzutello Bianco is the only green grapes that could test if the algorithm is able to differentiate green clusters from canopies.
- Different berry shape (round and conical) will give the opportunity to test the visualization of different cluster type.

The mix of these varieties is an exhaustive representation of most of table grapes vineyards. At least two of these varieties will be considered for the experimental activities.

1.3 Human workers required to work per hectare

In the following, an overview of the required manpower to manage the experimental site with the current (fully manual) operation modalities is reported.

Description	Macro phases	Agronomic operation	Man-hours per hectare
Leaf tips visible	Shoot and inflorescence development	Green pruning: suckers and sterile shoots, which do not have inflorescences and are not suitable for the winter pruning, are removed. During this operation basil leaves on both sides of the cluster zone are pulled. Shoots thinning is done during this stage to achieve an optimal distribution, density and leaf area.	160/ha
Shoots about 10 cm long; 5 leaves separated			
Inflorescence well developed single flower separated			
17-20 leaves separated: 50% caps off	Flowering	During full flowering vineyard management operations are not recommended due to the high sensibility of the whole micro-system.	
Young berries enlarging (>2 mm diam.); bunches at right angles to stem	Berry formation	Shoots positioning is normally done immediately after flowering to have an optimal clusters orientation. It improves canopy configuration and reduces the amount of shading created by overlapping shoots. Once the rachis reaches an optimal length, all	280/ha
7 mm diameter			
Berries touching (if bunches are tight)			

		the clusters are correctly positioned and the leaves that touch and obstruct the grapes are removed. Bunch selection (removal of not-suitable clusters), trimming and thinning are done before bunch closure.	
Berries begin to color and enlarge, berry softening begin	Berry ripening	During veraison canopy management is very important: leaf and laterals are removed to open up the interior of the canopy to light and air to promote fruit ripening, reduce disease pressure and increase spray coverage.	Leaf removing: 120/ha Harvesting: 360/ha
Berries harvest - ripe			
The canes complete their maturation	Senescence	Pruning: This operation affects any parameter of the canopy during the growing season	80/ha (only pruning) 80/ha (wood removing) 80/ha (canes adjustment)

This project is funded by the European Union G.A No 101016906

2 Description of the agronomic interventions

To simplify the definition of the benchmarks for the agronomic interventions, we briefly report the main agronomic activities involved in the project, i.e., harvesting and pruning, along with their decomposition in atomic activities. A detailed description of all the agronomic operations can be found in Deliverable 2.1 "Requirements, Specifications and Benchmarks".

2.1 Harvesting

After the seasonal cultivation process Black Magic is the earliest variety and it is harvested at the end of July – beginning of August, while Red Globe, Pizzutello Bianco and Pizzutello Nero are late varieties and are usually harvested in September and October.

This agronomic operation is composed of different steps: cluster detection, quality assessment, cluster collection and cluster placing, which are described below.

2.1.1 Cluster detection

All clusters and peduncles need to be recognized from canopies and obstacles. To facilitate this operation and the whole harvesting process, a few days before harvesting the vineyard is prepared with defoliating process.

Clusters are positioned under the canopies surface at around 190/160 cm from the ground, as shown in Figure 6. According to wide plant-distance clusters are spread around the plant with an irregular pattern and sometimes they can be close and occluded by another cluster.



Figure 6 - Flat-roof system.

It is required to remove all shoots and leaves from the clusters area before the harvesting operation, enabling a clear visibility of all clusters and peduncles and facilitating workers.

2.1.2 Quality assessment for selective harvesting

Table grapes are a type of fruit whose commercial quality is closely related to their visual appearance and overall flavor. The right harvesting of table grapes is usually determined by different parameters like skin color, texture softening, titratable acidity, total soluble solids, and sometimes with flavonoid

content, and aromatic compounds. Visual attributes of table grapes, such as intensity and uniformity of color, large size of berries, and brightness are the main characteristics that influence consumer choice.

Skin color - In recent years, due to the high temperatures, the Pizzutello Nero does not reach a good black color (only 15% of the production); the predominant color is light pink/dark pink and part of the production does not reach an acceptable color (it has white berries) and cannot be commercialized. The Pizzutello Bianco reaches a white-light yellow color. 'Black magic' achieves a good black color when ripe without issues. The red globe variety reaches a purplish red color; excessive production and unfavorable climate lead to coloring problems (white berries).

Sugar level and acidity - Sugar and organic acid compositions, which are measured through total soluble solids (TSS) and titratable acidity (TA), are the main parameters associated with the taste of this fruit. The growers mainly use TSS (sweetness) as an indicator of ripeness and most of the commercial varieties are considered mature when TSS ranges from 15 to 18%. The list of the varieties in the test plot is:

- Black magic: 16-18°.
- Pizzutello Nero: 17-20°.
- Pizzutello Bianco: 18-20°.
- Red Globe: 16-18°.

Bunch Density - According to specific variety, clusters can be too tight or too loose. Pizzutello Nero and Pizzutello Bianco generally have regular clusters; Black Magic could be too loose, depending on the season, while Red Globe does not generally have tight bunches. Note that the bunch density is heavily influenced by the seasonal climate.

Berry size –At harvesting time, it is an important parameter that affects the overall quality that consumers expect to receive. It must be considered in terms of cluster of berries (homogeneity) and in terms of size (caliber). For example, clusters with berry size smaller the 15 mm diameter, or with very big and very small berry in the same cluster, should be considered not complaint for the specific

customer specifications. The scale of values depends to the specific variety. The list of the berry size related to the varieties in the experimental field is:

- Pizzutello Nero: 18-20 mm.
- Pizzutello Bianco: 17-19 mm.
- Black Magic: 18-22 mm.
- Red Globe: 22-26+ mm.

Defects - Major and minor defects can occur and affect grapes quality. Some of them can occur at different stages before harvesting and others are more related to post-harvest management. Below the main defects of the varieties in the experimental field are reported:

- Red Globe (shown in Figure 7): sunburn, lack of color



Figure 7 - Example of sunburn for Red Globe.

- Black magic: over ripe (short harvesting window), berry drop, split
- Pizzutello Bianco (shown in Figure 8): very delicate skin, spots and browning, oxidations when over ripe, waterberry.



Figure 8 - Example of defects for Pizzutello Bianco.

- Pizzutello Nero: very delicate skin, spots.

2.1.3 Grape Collection

The worker collects the bunches from the plant walking through the row, harvesting one plant at a time. Each operator works only on one side of the row in couple with a partner who works on the other side of the row. The operator collects one cluster at a time: after detecting the cluster, the operator assesses the overall quality and cleans any wasted or rotten berry; then, he/she holds with one hand the peduncle close to the cane, leaving enough space to perform the cut with the scissors as shown in Figure 9. Finally, he/she uses the second hand to cut perpendicularly the peduncle with the scissor as close as possible to the cane.



Figure 9 - Cluster collection and cutting/holding points.

2.1.1 Cluster Placing

The operator places the cluster in the box. He/she holds the cluster from the peduncle and releases the bunch in a free area of the box with the steam upwards (steam-up), as reported in Figure 10-left. The clusters are gently placed on one single layer to preserve quality of grapes until the box is full. According to the type of box they are usually filled at average weight of 9/10 kgs. When the box is full, the operator starts filling the next box. All boxes are placed close to the plants in the vineyard before the harvesting; usually, one or two boxes per plant are placed according to the overall quantity of fruit. An example of full crates is shown in Figure 10-right.

This project is funded by the European Union G.A No 101016906



Figure 10 - Cluster placing.

2.2 Pruning

The pruning technique includes the cuts and other operations, carried out in different periods in order to:

- adapt the plant to the chosen training system.
- make production regular over the years, avoiding production stress and prolonging the efficiency of the vineyard.
- adjust the quantity of product to obtain the desired quality.
- keep the development of the aerial part balanced in relation to the root system.
- arrange vegetation and production in an appropriate way.

It is important to distinguish winter pruning, which is done in the absence of vegetation (depending on the varieties and production area, from November to January) from green pruning which refers to the operations carried out on the vegetation and ensures the correct balance between vegetative part and fruit. In this project only winter pruning will be tested.

The pruning agronomic operation is composed of different steps: canes detection, canes and bud selection and canes cut, which are described below.

2.2.1 Canes detection

This phase requires the detection of the main branches and canes of the specific plant and the identification of mother (upper part), and daughter (lower part) of any internode, as shown in Figure 11. Sometimes the operators need to clear the area from other plants canes that occlude the interested plant.

2.2.2 Canes and bud selection

The practical and simple rule that most of the time allows to correctly prune plants in full production is: ***choose the first good cane***.

Main characteristics to identify ***the first good cane*** are:

- **External:** To allow a correct development of the plant structure over the years and increase exposure to sunlight the cane should be on the external side of the branch, considering the main intersection.
- **Minimum diameter:** Avoid too small and too big cane diameter.
- **Healthy:** Not dry and absence of any diseases.
- **Minimum distance:** To facilitate the lymphatic flow, the cane should be not too close to the internode (usually not closer than 4/5 cm). If distance is greater than 20/25 cm, it could be useful to leave a spur that can be used for next year pruning.
- **Minimum length:** the cane should be long enough for the requested number of buds (buds selection).
- **Main branches loading balance:** Considering the total number of canes requested, increase the number of good canes on the upper part of the plant (Mother). The lower part (Daughter) getting more lymphatic flow is stronger and could take all nutrients weakening the mother. This general rule is valid for main branches, secondary branches, and canes where the number of buds should be higher on the mother side compared to the daughter side. The classification of mother and daughter is exemplified in Figure 11.

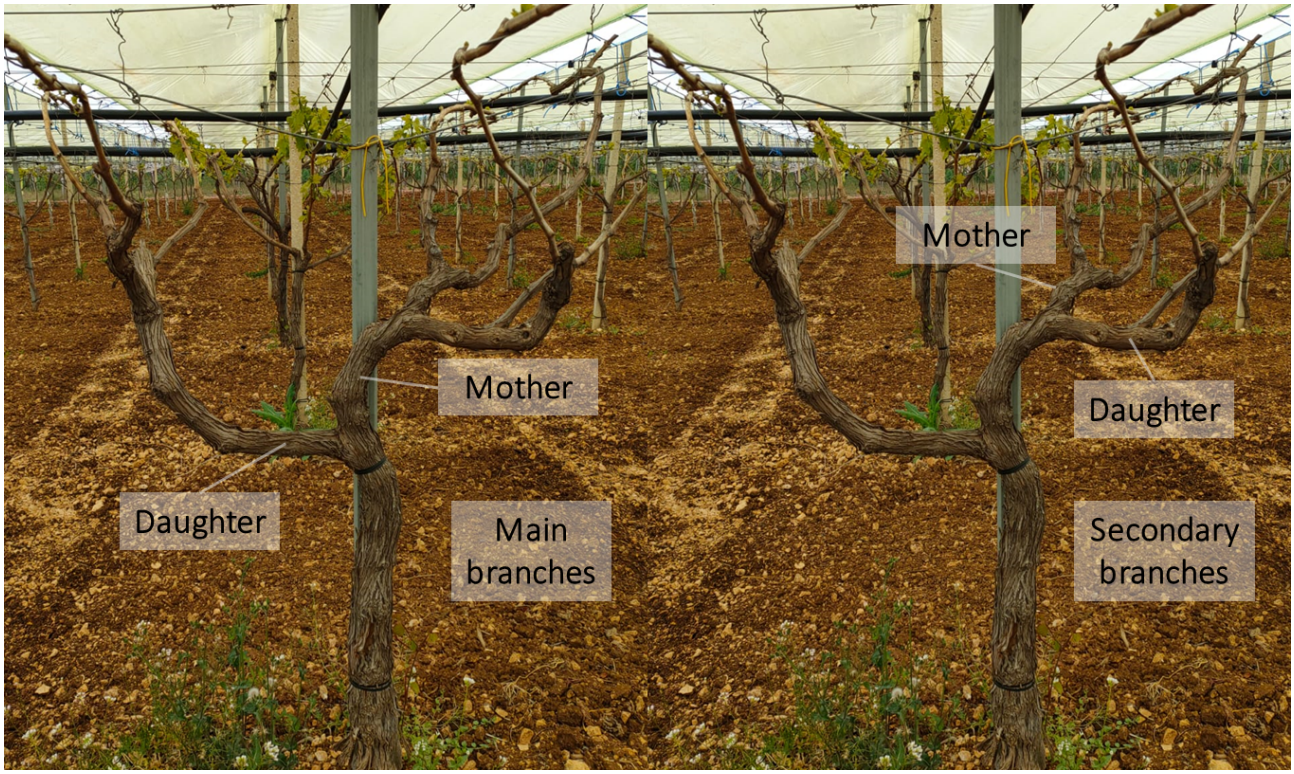

Figure 11 - Mother and Daughter sample.

2.2.3 Canes cut

The cane cut is usually performed with a single hand holding scissors and with the second hand helping to remove obstacles and moving the cane in the right position to execute the cut in the optimum position and direction: close to the intersection (few mm) and perpendicular to the cane. Nowadays new electric scissors are also available that allow for faster and easier cutting.

This project is funded by the European Union
G.A No 101016906

Seasonally, according to variety, year of production, plant density and production target, the grower decides the total number of buds to have in production in the specific vineyard. The total number of buds divided by the total number of plants and the number of canes per plant, gives the number of buds to leave per cane. For example, if total buds target is 50.000, number of plants is 1.400, canes per plant are 5, then buds per cane will be 7. The operator holds with one hand the cane and with the other hand cuts the cane after the 7th bud.

3 Benchmarks for validation

In the following Table 1, we describe the set of benchmarks that will be used to measure the project advancement and assess the effectiveness of the obtained results.

Table 1 - Benchmarks that will be used to measure the project advancement and assess the effectiveness of the obtained results.

Objective	Benchmarks
Basic Robots Functionalities	
1.1 Robots navigation	1. *Benchmark for robot control law* - Pose regulation with an accuracy of 15 cm regarding the position (x,y) and 20° regarding the orientation. 2. *Benchmarks for robot localization and mapping* - Identification of all landmarks (vertical pales with the tree trunks tied up to them) in the vineyard with an accuracy of 15 cm. - Identification of the row orientation with an accuracy of 20°. - Relative localization to the closest landmarks with an accuracy of 15 cm regarding the position and 20° regarding the orientation. 3. *Benchmark for autonomous navigation* - Capability of safely navigating from one row to another one by doing the related maneuvering operations without losing its localization capabilities within the map, e.g., by keeping its localization accuracy within 15 cm regarding the position and 20° regarding the orientation along the trajectory.
1.2 Dual-arm manipulation and human safety	1. *Benchmarks for basic dual-arm functionalities* - Capability of planning cooperative trajectories without violating kinematic and dynamic limits. - Capability of performing cooperative manipulation without collisions (given perception) with external objects and/or self-collisions. - Capability of avoiding kinematic singularities. 2. *Benchmarks for evaluating agronomic operations* - Cluster collection in single arm or dual arm mode with a success rate of at least 70%. - Cluster placing in single arm or dual arm mode with a success rate of at least 90%. - Canes pruning with a success rate of at least 70%.

	3. Benchmarks for basic human safety features (physical and not) Depending on the task to be performed, physical contact between human operator and robots may be allowed or not. For instance, a safety distance must be kept when the mobile base is moving, while physical contact may occur to have the human adjust the end effector configuration. Depending on the robot tasks, the following benchmarks are considered: - Human safety guarantee via definition of a safety field accounting for human and robot relative position and velocity and for which a minimum threshold is set. Interaction is considered safe if the safety field is kept above the threshold. - In tasks for which physical interaction is allowed, capability to adjust the arm and end-effector configurations according to human exerted wrench with a success rate of at least 90% and without instability phenomena. - In tasks with both human interaction and moving mobile base, the latter is within the conditions detailed in objective 3.4.1, i.e., keeping a spatial robustness (see brief definition in 3.4.1) of 5% of the distance between plants according to D2.1, Table 1. *4. Benchmarks for evaluating the designed end effector* - Capability of holding 2kg payload. - Capability of cutting and hold 20mm diameter branches. - Capability of cutting 50mm branches.
1.3 Box exchange mechanism	*1. Benchmarks for evaluating the designed box exchange mechanism* - Box exchange possible between the two robots with maximum allowed pitch misalignment error up to 10°. - Box exchange possible between the two robots with maximum allowed roll misalignment error up to 15°. - - Box exchange possible between the two robots with maximum allowed yaw misalignment error up to 10°.
1.4 VR environment and robot simulation	1. *Benchmarks for evaluating the realism of the VR environment and agronomic-oriented sensors* - The setup will be benchmarked in an environment which matches the vineyard structure “Tendone”, with a distance between each plant of 3m x 3m, with an accuracy between the simulation of 10 cm. - The simulated environment will follow the description of the vineyard static structure measurements provided in Deliverable 2.1 with an accuracy of 95%.

	- The properties of each virtual sensor like resolution, field-of-view will match that of the real sensors. - The performance of each the sensor, measured in published messages per second, shall be at least equal the performance of a simulation of those sensors in the Gazebo simulator. 2. *Benchmarks for evaluating the realism of the simulated robotic platforms* - The movements of the simulated robotic platforms will be compared to the movements of the real robotic platforms. The same control messages will be sent to VR simulation and the real robots and the results will be compared. The positional difference between the two simulations of the end effectors shall be less than 2 cm. Also the maximum error on the orientation components will be of 10°.
Agronomic-Oriented Perception	
2.1 Agronomic target detection and localization	1. *Benchmarks for cluster detection and localization for harvesting* Test data will be collected and labelled to test the accuracy of the algorithms. The algorithms will be compared with the state of the art of cluster detection and localization in harvesting, for example FASTER R-CNN. The following test will be done: - Test of accuracy in cluster detection with a success rate of at least 70%. - Test of accuracy in cluster localization with a success rate of at least 70%. 2. *Benchmarks for canes detection in pruning* Test data will be collected and labelled to test the accuracy of the algorithms. The algorithms will be compared with the state of the art for canes detection in pruning. The following test will be done: - Test of accuracy in canes detection with a success rate of at least 70%. - Test of accuracy in canes localization with a success rate of at least 70%.

	3. Benchmarks for canes and buds selection for pruning Test data will be collected and labelled to test the accuracy of the algorithms. The algorithms will be compared with the state of the art for canes and buds selection in pruning. The following test will be done: - Test of accuracy in selection of canes and buds with at least 50% of compliance with respect to the cutting suggestions provided by an expert.
2.2 Agronomic quality assessment	*1. Benchmarks for Quality Assessment for selective Harvesting* Test data will be collected and labelled to test the accuracy of the algorithms and quality assessment procedures. More in detail, the algorithms should be able to have high (close to 80% of human performances, on average) accuracy on selection of grape bunches to harvest according to specific quality parameters, as described in D2.1. With this respect, the algorithms should pass the following tests: - Berry size estimation: achieve a mean absolute error (MAE) less than 30% of berry size on benchmark data. - Color classification: achieve at least 80% of accuracy on grape bunch color classification on benchmark data. - Sugar content estimation: achieve a MAE less than 5 °Brix with respect to the measured value of °Brix sugar level using visual sensors on benchmark data. - Density estimation: achieve at least 80% of accuracy on bunch density estimation on benchmark data. - Defects estimation: achieve at least 80% of accuracy on defects and sickness identification on benchmark data.
Planning and Control for Human-Multi-Robot System	
3.1 Human-body prediction and motion planning in agreement	*1. Benchmarks for evaluating human joints position prediction* Test data will be collected and labelled to test the accuracy of the algorithms. The algorithms will be compared with the state of the art of human joints position prediction, for example recurrent CNN or attention-based models. The following test will be done: - Test of accuracy in human joints position prediction. We expect to have an error below 25 cm in the L_2 distance error, for a prediction of 2.5 seconds, in the 70% of the cases.

	2. *Benchmarks for evaluating motion planning in agreement with prediction* For its part, the benchmarking of the motion planning in agreement with the previous prediction cannot be done by comparing the planner with another planner since it is subject to the use of our own prediction system. Therefore, we will compare it with the same planner without using prediction in order to check the usefulness of using that prediction when performing collaborative tasks with humans. On the one hand, we will use questionnaires available in the literature to obtain subjective metrics about how the user evaluates the fluidity and naturalness of the robot's movements. We will also obtain objective measures such as the number of performed replanning, the mean acceleration of the robot's joints or the total time required to perform the task. The expected results are that the human will consider the robot's movements as more fluid when the above prediction is used. More specifically, we expect to obtain statistically significant differences (p-value < 0.05) when evaluating fluency-related factors in the questionnaires. It is also expected that the mean acceleration of the joints is reduced by, at least, a 5% and that the total time to perform the planned task is reduced by, at least, a 5%.
3.2 Human-robot communication and awareness	1. *Benchmarks for evaluating the effectiveness of the different means of communication* The effectiveness of the different means of communication will be evaluated by first asking expert users to evaluate communication acts in the virtual environment and the fill questionnaires based on Likert scales. As WP6 will define tens of different communication acts, a subset of these will be selected for the evaluation. The outcome of this evaluation can then be used to modify the system accordingly. An evaluation can then be performed on the field. A first goal will be to have a percentage of correctly detected acts between 70% and 100% according to the cruciality of the communication act itself. Questionnaires instead must assess the appropriateness of the communication modality with respect to the gravity of the specific situation. In addition, the evaluation must ensure that the communication modalities can scale up when multiple humans and robots are involved in the tasks. The questionnaires will be defined taking inspiration from the Robotic Social Attributes Scale (RoSAS). A second goal will be then to ensure that, for each selected communication act, the average score of reported answer is at least 20%

	above normal, i.e., if 3 is the normal score, 3.6 should be the minimum average scores reported by users. *2. Benchmarks for evaluating human awareness* Human-Robot awareness is based on a high-level specification of the environment and tasks, for example by using Petri Nets describing agronomic processes. The suitability of process models to describe actual reality can be captured by using Conformance Checking, which is a common task in Business Process Management (BPM) providing a quantitative measure of how well a process model adheres to actual behavior. This measure is usually called *fitness*. The higher is the fitness value (comprised between 0.0 - very poor fitness - and 1.0 - perfect fitness) This mechanism can be used at runtime to detect deviations but can be used also to evaluate the human awareness mechanisms. Evaluation will be then performed by asking humans to carry on complex activities together with robots, and by measuring how well they can be described by process models defined in the context of the project. For our project we require a fitness of at least 0.7 on all the high-level tasks.
3.3 Learning from demonstration	*1. Benchmarks for learning from demonstration methods* Test data will be collected to test the generalization capability of the algorithm in approaching for bunch harvesting and pruning tasks: - For each task, the required number of demonstrations for adaptation is at most 30-50. - For each task, the reproduced skill should generate trajectories as motion plans to reach given targets, with a more than 80% success rate.
3.3 Multi-robot coordination and distributed protocols	*1. Benchmarks for BEM coordination* - Alignment of logistic and farming robots for box exchange with a success rate of at least 80%. *2. Benchmarks for distributed protocols* - Theoretical convergence guarantees, correctness, and completeness of the distributed protocols.
3.4 Multi-robot task planning with human operators	*1. Benchmarks for multi-robot task planning with human presence* Test data will be collected to test the robustness and accuracy of the algorithms, which should have:

	- A spatial robustness of 5% of the distance between plants according to D2.1, Table 1; and a temporal robustness of 5 s, where these are defined as the spatial and temporal distances between the behavior and the boundaries of the set of all behaviors satisfying the design property. *2. Benchmarks for multi-robot task planning with human in-the-loop* Test data will be collected to test the robustness and accuracy of the algorithms, which should have: - A spatial robustness of 5% of the distance between plants according to D2.1, Table 1; and a temporal robustness of 5 s (see brief definition above). - An error threshold of 10% with respect to the robustness metrics for both global and local least violating solutions. - More than 60% success rate in reconfiguring the multi-robot plans according to the additional high-level human specifications under unforeseen environmental changes within the least violating solution margins (e.g., avoiding regions with misplaced tools or poor lighting conditions).
3.5 VR Multi-Human-Multi-Robot Interaction Component	*1. Benchmarks for evaluating the Multi-Human-Multi-Robot interaction in the VR* - The multi-human/robot environment shall have at least 10 robots and 6 humans.

This project is funded by the European Union
G.A No 101016906

4 Final Demo: Integrated System Field Validation

The final demo is meant to validate the integrated system on the field with all the human-robot functionalities described in the proposal. In this regard, we will consider both harvesting and pruning as case-study of in-field agronomic interventions for assessing the validity of the CANOPIES methodological findings. Notably, due to the seasonal nature of these operations that occur in different times of the year, i.e., harvesting in summer and pruning in winter, the outcome of the final demo will consist of sequence of videos demonstrating that the two agronomic tasks can be successfully achieved.

As far as the two agronomic operations are concerned, we will consider a typical scenario where there will be human beings at work (farmers) and the robotic platforms will be supporting these activities. Interactions will occur only among the robotics platforms and one of the human operators which will be designed as the responsible of the robotic counterpart. Nevertheless, robotic platforms should be able to safely work in the shared environment with also the other human operators even if not directly interacting with them. As far as the portion of the experimental field (described in Section 1) used for the final demo is concerned, most likely the activities will be limited to a small area (a few rows identified according to the table-grape variety selected for the final demo) which should suffice to demonstrate: i) the autonomous capabilities of the robotic planforms, ii) the coordination capabilities of the robots; iii) the capabilities of sharing the working space with the human beings and interacting with some of them; and iv) the agronomic intervention capabilities of the robotic platforms.