

Service ID S00366

Location Italy, Remote



Provision of datasets

Provider service

Politecnico di Milano

Link to content

<https://agrifoodtef.eu/services/provision-datasets-1>

Type of Sector

Arable farming, Food processing, Greenhouse, Horticulture, Livestock farming, Tree Crops, Viticulture

Accepted type of products

Design / Documentation, Other

Type of service

Provision of datasets

Description

During the development of systems based on AI and/or robotics, tailored data are needed to support the realisation, validation and optimisation of systems and subsystems. This service aims at providing such data. AI models are a noteworthy example of systems that require data to be trained; more generally, any system which processes data produced via artificial perception (such as images, videos, LiDAR, ultrasound, radar...) requires suitable data to be built, tested and fine-tuned. This service allows a customer to define together with AgrifoodTEF the specific features of the data that the customer's activities require and to receive datasets precisely matching such features. Among the features to be defined are the type, number and specifications of the data streams comprised in the datasets (e.g., when multi-sensor data are required) and the range of variations encompassed by the data (e.g., variety of environmental conditions represented in a sensor stream); lower-level features to be defined concern formatting and packaging of the data.

How can the service help you

It is not a simple task to define what data are needed to develop or test a system (e.g., AI model) or machine (e.g., robotised agricultural implement): choosing the features that data used for these activities must possess requires expertise that not all companies have internally.

This service can support customers with this crucial task. Once the features of the required data have been defined, it is then necessary to generate the corresponding datasets. This is done by collecting, processing and often also augmenting the data (e.g., by adding metadata or labels). This activity requires specialised expertise and usually imposes a strong burden on a company: this service helps customers by moving the burden onto AgrifoodTEF.

How the service will be delivered

This service description is intentionally generic. Every instance of this service is, in fact, customised to adapt it to the needs and requirements of the specific customer. The following is an example of a service instance (please note that the service is available for many agricultural sectors, not only the one considered by the example).

Example service: The customer is a software company developing a computer vision AI model and needs data to train the model. The model's goal is to be able to reliably recognise Matricaria and bean plants and to detect the emergence of other spontaneous weed species. The model is intended to be provided as a module to manufacturers of implements, and the company does not want to overly constrain the set of potential clients by imposing strict criteria on the camera system to be mounted onboard the implement. For this reason, together with AgrifoodTEF, the customer defines a set of 4 most common camera system configurations, which must all be covered by the training datasets.

For each of the 4 configurations, the company requires data in 3 different lighting conditions, i.e., strong sunlight with the sun overhead; strong sunlight with the sun low on the horizon; diffuse sunlight (e.g., cloudy weather). No labelling is required from AgrifoodTEF since the company intends to execute it internally. AgrifoodTEF already has in its archives suitable datasets covering 5 of the 12 conditions required by the customer, so these get suitably formatted and immediately

Service customisation

The service comprises three separate phases. Phase One is data specification definition and takes place via one or more interviews (usually remote) with the customer.

These interviews allow AgrifoodTEF to define what datasets are needed by the customer and their exact features. During this phase the customer can choose to share with AgrifoodTEF, under NDA if needed, details about the system that the data will be used with: such details allow an even closer match between system requirements and data specifications.

Phase Two is data preparation and strongly depends on the outcome of Phase One. Three different cases may occur:- In Case 1, the required datasets are available in AgrifoodTEF's archives, already in a form that is fully compliant with the specifications. Service execution can proceed to Phase Three (described below).- In Case 2, suitable data are available in AgrifoodTEF's archives but require additional processing (such as labelling, transcoding, and formatting) to be compliant with the specifications. In this case, services S00115 (data validation and processing services) and/or S00290 (data labelling) are leveraged as a preliminary step, then the situation reverts to Case 1.- In Case 3, no (or insufficient) data compliant with the specifications are available in AgrifoodTEF's archives. In this case, services S00112 + S00113 (Execution of physical testing + Collection of test data during physical testing) and/or S00182 + S00183 (Execution of digital testing + Collection of test data during digital testing) are employed to generate raw data according to the customer's needs, then the situation reverts to Case 2. Generally speaking, Phase Two is completed as soon as the activities needed to get to Case 1 have been executed by AgrifoodTEF