Innovation Grant

Innovation Grant Proposal Template

Title of the Project: Agri@Intesa
Leading Partner: please indicate also a main contact for the leading partner (name and email address) **Intesa Sanpaolo (Paolo Panzarini <paolo.panzarini@intesasanpaolo.com>)**
List of participants (ICSC public and private partners): ● Intesa Sanpaolo (Spoke 2) ● THALES ALENIA SPAZIO (Spoke 2) ● UNICT (Spoke 2) ● UNIMIB (Spoke 2) ● UNIBA (Spoke2 e Spoke5) ● UNINA (Spoke 2) ● UNISALENTO (Spoke 2)
Proposer Spoke: Spoke 2
Other Spokes involved: Spoke 5
Short Abstract (max 1/2 page): Intesa Sanpaolo is interested in developing algorithms that could calculate synthetic indicators that are predictive of relevant phenomena for economic well-being of the agricultural firms, such as the productivity of fields and risks related to productivity and yield.
Starting TRL of the project: TRL2: Technology concept formulated **Target TRL of the project** (Technology Readiness Level): **TRL 5: Technology validated in relevant environment**
Expected budget (total cost k€): 505 kEur (480 kEur agevolazione)
% of total budget for activities in "Regioni del Mezzogiorno" (Abruzzo, Basilicata, Calabria, Campania, Molise, Puglia, Sardegna, Sicilia)**: 76%**

---

1. Description of the project

Introduction
The availability of satellite data has increased in recent years, with more and more satellites being launched into orbit to capture medium and high-resolution images of the Earth's surface. This has led to a significant increase in the amount of data available for analysis, in the resolution of such images, and has also made it possible to obtain more detailed and accurate information about various phenomena.

In addition to the increased availability of satellite data, the cost of accessing this information has also been decreasing. This has made it more affordable for researchers, companies, public administrations and other organizations to use satellite data in their work.

Alongside the growth in satellite data, there has also been an increase in the availability of open data. This includes meteorological data, hydrogeological data and information on agricultural crops. These datasets are available to the public (mostly but not exclusively for free) and can be used to support a wide range of research and analysis.

Overall, the combination of increasing availability, higher resolution and reduced cost of satellite data, along with the growing availability of other relevant data, provides new opportunities for understanding and addressing complex challenges in fields such as agriculture, environmental management and disaster response.

Furthermore, it is possible to combine all these data through synthetic indicators to measure actual or past phenomena and, more interestingly, predict those phenomena.

Intesa Sanpaolo is interested in developing algorithms that could calculate synthetic indicators predictive of relevant phenomena for economic well-being of the agricultural firms, such as the productivity of fields and risks related to productivity and yield.

Aim of the project

This project aims to use satellite imagery and public or private available data to analyze and understand the relationship between these sources of information and the productivity of agricultural companies. By combining remote sensing data with available information on crop yields, weather patterns, and other relevant factors, the project hopes to identify patterns and correlations that can help Intesa Sanpaolo to understand and predict the efficiency and sustainability of agricultural business.

The ultimate goal is to provide Intesa Sanpaolo with a proven model or algorithm that takes in satellite images and publicly available data (free or no free) to calculate and predict indicators related to risk and productivity of the agricultural companies to be further developed internally to integrate the more traditional financial methods to assess the risk and productivity of farms.

The hypothesis of the study is that there are optimal combinations of satellite imagery, vegetation indexes, other data and time periods to achieve best possible estimations of productivity and risks in agricultural companies.

Project Scope

The project scope is extensive to all the Italian territory and to all the cultures; most probably the algorithms and relevant data could be different for different cultures and the and data availability could be different in different regions; the project could consider to start with specific cultures and or territory to set a basic framework that could be scaled up later in the next phases of the project.

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

Definition of the data source
First goal of the project is the definition and provisioning of the relevant data to be used in the model.
Satellite imagery should be the most relevant, with imagery we intend RGB, hyperspectral, multispectral (optical through Near Infrared or NIR) and SAR data that allow the calculation of different standard indexes such as

- NDVI: it allows to evaluate the health of the vegetation, analyzing the reflectance of the vegetation in the Red and NIR bands.
- SAVI: allows to evaluate the conditions of vegetation development in the emergency and early stages of development, as it applies a correction to bare soil.
- LAI: leaf area index that estimates the leaf area of the plant expressed in m2 on m2.
- NMDI: can be used to assess the water content of the soil; in case of bare soil, a high index value indicates dry soil. In the presence of vegetation, a high index value indicates that the plant is not under water stress.

Data sources will be also employed for land-use/land-cover automatic and semi-automatic classification and segmentation.

Availability of satellite imagery
There are several satellites that capture RGB, hyperspectral and multispectral images from space, including for example Sentinel-2, Landsat 8, Planetscope Prisma, and Sky Sat. These satellites provide images with a spatial resolution ranging from 1 meter to 60 meters, depending on the satellite and the specific band being used.
In terms of temporal resolution, most of these satellites have regular intervals at which they capture images. For instance, Landsat 8 captures images every 16 days, while Sentinel-2 captures images every 3 to 5 days depending on the location. Planetscope and SkySat have daily sampling.

Other data
The model could include any relevant publicly available information - with an at least 3 years history - that could be relevant to build the model:

- specific information about farm, from the geolocation to the agricultural activity including: extensions, specific cultivations of each field
- weather information, including storms, floods, mm of rain
- insolation
- soil data
- pollution
- hydrogeological data
- areas of production of certified origin of certain production (DOP, DOCG,)
- …

Enabling technologies:
- *Satellite images*
- *Machine learning and Deep Learning*
- *Feature extraction*

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

Innovation Grant

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

In synergy with Spoke 2 / WP6 activities in "space economy" and Spoke 5.

1.1 Expected results

- demonstration of the prediction capability of machine learning algorithms applied to satellite info for financial data at the level of PoC (Proof of Concept)

- generation of example datasets for further studies

- transfer of academic knowledge to industry

- *Strengthening national skills in the field, leveraging on experience in other industrial domains e.g. ……*
- *Consolidate national positioning vs other European domain industries….*

2. Expected impact (max 1 page):
- new prediction capabilities for the financial sector in the field of agricultural data to be employed in decision making
- formation of experts in the handling of space economy images, from extraction to management and interpretation
- use of the CN state of the art infrastructure for novel industrial applications

3. Implementation

3.1. Work Packages structures and timeline of the project:

WP / Activity 1: Image Data collection, fusion and curation
Partners: UNIBA (Spoke 2 + Spoke 5), THALES ALENIA SPAZIO, UNINA, UNISALENTO, UNIMIB

The activity is intended to oversee the procurement of the satellite image data needed for the analysis in the subsequent steps. It will also define the structure, access and interoperability among nodes of the database.

The procurement of the data will follow three possible paths, to be established at the very beginning of the project:
- *UNIBA:*
  - *external procurement via commercial acquisition; in this case, the affiliate will follow data curation and the contacts with the firm, ensuring a proper data collection is delivered*
  - *internal procurement via public sources or via pay-walled portals. In this case the affiliates will ensure the proper data quality*
- *UNINA, UNISALENTO, UNIMIB*
  - *procurement of images from public repositories*

![ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing]

The type and extent of the images will be subject to identification with the industrial partner (Intesa) before the start of the project, and during the first months; following also a study of the availability of data on the free / paywalled / private market.

In particular
- *UNIBA supervise the image collection activity, preparing the dataset to be given to the WP3 for analysis*
- *UNIBA will procure images either via external procurement, or internally via public sources*
- *UNIBA will work with Intesa (WP2) to perform the matching between images and financial data*
- *THALES ALENIA SPAZIO / UNINA:*
  - *help in image type selection / definition*
  - *final strategy doc on production implementation and perspectives*

Milestone M1.1 (Synced with ICSC MS7): final decision for the image dataset(s); at least a few example images to be used for initial code development
Milestone M1.2 (Synced with ICSC MS8): first dataset available to WP3
Milestone M1.3 (Synced with ICSC MS10): final dataset available to WP3

Deliverable D1.1 (at the end of M1.1): report on the final decision(s) for M1.1; "draft" dataset(s) available (ref: UNIBA)
Deliverable D1.2 (at the end of M1.3): description of the final dataset as made available in M1.3 (ref: UNIBA)
Deliverable D1.3 (at the end of M1.3): final strategy doc on production implementation and perspectives (ref: THALES ALENIA SPAZIO / UNINA)

WP / Activity 2: Financial data Collection
Partners: Intesa Sanpaolo

The activity is intended to oversee the collection of financial data on farms and firms in regions covered by the satellite data and by the metadata for which detailed images can be collected. These include (not exclusively):
- *budget data from the previous years*
- *type of crops*
- *any other relevant data obtainable from public registries*
- *Intesa will work with UNIBA (WP1) to perform the matching between images and financial data*

Milestone M2.1 (Synced with ICSC MS7): final decision for the financial dataset(s); at least a few example entries to be used for initial code development
Milestone M2.2 (Synced with ICSC MS8): first dataset available to WP3
Milestone M2.3 (Synced with ICSC MS10): refined dataset available to WP3

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

Deliverable D2.1 (at the end of M2.1): report on the final decision for M2.1;"draft" dataset(s) available
Deliverable D2.2 (at the end of M2.3): description of the final dataset as made available in M2.3

WP / Activity 3: Data Analysis and Model Preparation
Partners: UNICT, UNIMIB, UNISALENTO, UNINA

The activity contains the core scientific part of the project: modern AI based algorithms are going to be tested on the data obtained via WP1 and WP2, in the quest for models with predictive value on financial parameters.
In the first months of the project, the groups will perform activities of technology tracking and planning, in order to define a precise strategy for the analyses. The groups will involve, whenever necessary, domain experts to help classify and label the data.

The affiliates will follow different approaches, to be compared in the final phase.
In particular:
- *UNICT will conduct image analysis and explore various approaches to automatically categorize the different parts of an image based on their features, such as color, temperature, and size. This research aims to contribute to the recognition of the amount of vegetation, the type of vegetation, and the health of plants. To achieve this, a pilot study will be undertaken, focusing on a specific area and using a limited number of vegetation types. The purpose of this study is to gather knowledge on the key characteristics that can be used for accurate identification.*
- *UNIMIB will analyze crop health indicators, such as vegetation indices, moisture levels, and disease prevalence, UniMib will develop a methodology for providing insights into potential yield fluctuations, production losses, and market vulnerabilities. To this aim, UniMib will exploit the most recent methodologies in the field of multi-source remote sensing, image recognition, image segmentation, supervised and unsupervised Machine and Deep Learning.*
- *UNISALENTO: Usually a measure of financial risk associated with smallholding farmers can be focused on the relation between the actual financial result and the yield of the field. Past data from the field, obtained with satellite images and data about the financial results or information about the field past yields can be used to develop a model that measures risk. The data about weather, field productivity, crop rotation and solid moisture can be related with the evolution in time of various vegetation indexes usually correlated to the health of the plants. We will work on relating the various vegetation indexes or other data about the field with the financial data using a classic set of tools from data-analysis: supervised or unsupervised models with a focus on the relation between model size and its generalization capacity.*
- *UNINA will fine tune to the problem its proprietary deep learning models and methodologies and shall use domain knowledge provided by geologists and likely by UNICT to perform segmentation on multispectral heterogeneous data cubes (eg. DEEP FOCUS); 2) will test methods to integrate visible-NIR datacubes with SAR (which*

ICSC
Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

provide relevant and complementary information with respect to optical data) images using generative AI.

Milestone M3.1 (Synced with ICSC MS7): technology tracking and description of strategies
Milestone M3.2 (Synced with ICSC MS9): tests on the first dataset
Milestone M3.3 (Synced with ICSC MS10): tests on the refined dataset available

Deliverable D3.1 (at the end of M3.1): report on the final decision for M3.1
Deliverable D3.2 (at the end of M3.3): final report on the modelling

3.2 Human resources *(for each partner):*
- *FTE effort of junior (to be hired: MSc, PhD, Postdoc):*
    - *UNIBA - 18 PM postdoc (spoke 2) + 12 PM postdoc (Spoke 5)*
    - *UNICT - 18 PM postdoc*
    - *UNISALENTO - 18 PM postdoc*
    - *UNINA - 18 PM postdoc + 8 PM postdoc*
    - *UNIMIB (20 person-month)*
- *FTE effort of senior (critical mass researchers):*
    - *UNIBA - 2 PM/y (Spoke 2) + 0.5 PM/y (Spoke 5)*
    - *UNICT - 2 PM/y*
    - *UNISALENTO - 2 PM/y*
    - *UNINA - 2 PM/y*
    - *UNIMIB (1 person-month)*
    - *INTESA 3 PM/y (spoke 2)*
    - *THALES ALENIA SPAZIO 2 PM/y inkind contribution*

3.3 Expected budget (total cost k€):
Use of the excel template for breakdown of the project's costs is mandatory.
Please, indicate in this section a short explanation of the budget (if not only personnel costs).

- *UNIBA:*
    - *110 kEur (60kE personnel + up to 50kE cost for acquisition of images)- Spoke 2*
    - *40 kEur (personnel) - Spoke 5*
- *UNICT: 70 kEur - Spoke 2*
- *UNISALENTO: 70 kEur - Spoke 2*
- *UNINA: 95 kEur - Spoke 2*
- *UNIMIB: 70 kEur - Spoke 2*
- *INTESA: 25 kEur (+25 kEur cofunding) - Spoke 2*
- *THALES ALENIA SPAZIO: 0 Eur - 25kEur funding on UNINA - Spoke 2*

4. Other informations

- **HW/SW requirements**: *(to be requested to the ICSC)*

*https://www.gea.mur.gov.it/docs/Cn/allegati/Decreto%20Direttoriale%20n.3138%20del%2016-12-2021.pdf

      ○ *Two equivalent-nodes with at least 32 cores and 4 NVidia Ax00 for shared use by the analysis groups (UNIMIB, UNICT, UNISALENTO, UNINA, …)*
        ▪ *Leonardo Booster nodes, for example*
      ○ *100 TB of disk to host / share images*
- **Synergies with Other PNRR M4C2 projects** *(if applicable)*
- *project STILES* <u>*Strengthening the Italian leadership in ELT and SKA (STILES)*</u> *- Deep learning farm at UNINA*
- **Synergies with other national/international projects** (*(if applicable)*

- **If needed for a specific Spoke, add any other relevant information** *(es. Preferred development site, Cross-spoke relations: is the Project cross-site? If yes explain how it unfolds)*

Annex:

I soggetti privati possono ricevere finanziamenti nel rispetto della normativa sugli aiuti di Stato

Ambiti finanziabili	Soggetti che svolgono attività non economica	Soggetti che svolgono prevalentemente attività economica			Soglie di notifica per soggetti che svolgono attività economica
		Grande impresa	Media impresa	Piccola impresa	
Ricerca fondamentale	100 %	100 %	100 %	100 %	40 mln € per impresa e progetto
Ricerca Industriale *(Maggiorazione per collaborazione e/o diffusione)*	100 % (0 %)	50 % (15%)	70 % (15%)	70 % (10%)	20 mln € per impresa e progetto (15 mln € per impresa e progetto)
Sviluppo sperimentale *(Maggiorazione per collaborazione e/o diffusione)*	100 % (0 %)	25 % (15%)	35 % (15%)	45 % (15%)	15 mln € per impresa e progetto
Studi fattibilità	100 %	50 %	60 %	70 %	7,5 mln € per studio
Investimenti in infrastrutture di ricerca	100 %	50 %	50 %	50 %	20 mln € per infrastruttura
Aiuti all'innovazione a favore di PMI *(Maggiorazione per agevolazione relativa a costi consulenza per 200 k € in 3 anni per beneficiario)*	0 % (0 %)	0 % (0 %)	50 % (50 %)	50 % (50 %)	5 mln € per impresa e progetto
Formazione *(Formazione per disabili o svantaggiati)*	100 % (0 %)	50 % (10 %)	60 % (10 %)	70 % (10 %)	2 mln € per progetto di formazione
Aiuti a Poli – Aiuti all'investimento	100 %	50 %	50 %	50 %	7,5 milioni di EUR per polo
Aiuti all'investimento – ex. Art. 107 paragrafo 3 lettera c Aiuti all'investimento – ex. Art. 107 paragrafo 3 lettera a	0 % 0 %	5 % 15 %	5 % 15 %	5 % 15 %	
Aiuti ai Poli – Aiuti al Funzionamento	100 %	50 %	50 %	50 %	