AGRICULTURAL UNIVERSITY OF ATHENS

DEPARTMENT OF NATURAL RESOURCES MANAGEMENT & AGRICULTURAL ENGINEERING

Geographical Information Systems Research Unit | AGRICULTURAL UNIVERSITY OF ATHENS

Deliverable 1.2

Data Management Plan and Project Risk Management Plan for the project "Spatially Explicit Digital Twin of the Greek Agro-Hydro-System DT-Agro"

DT-Agro

Athens, July 2024

Data Management Plan (DMP) and Project Risk Management Plan (PRMP) for the project “Spatially Explicit Digital Twin of the Greek Agro-Hydro-System DT-Agro”

Plan Details

Plan Title	“Spatially Explicit Digital Twin of the Greek Agro-Hydro-System DT-Agro”
Deliverable	D1.2
Fields of science and technology (from OECD classification)	Agricultural Sciences - Food Science & Technology
Language	English
Creation Date	15/06/2024
Last modification date	06/08/2024
Associated documents (publications, reports), website	website: https://dt-agro.aua.gr/

Project Details

Project Title	"Spatially Explicit Digital Twin of the Greek Agro-Hydro-System (DT-Agro)"
Acronym	DT-Agro
Abstract	DT-Agro is a 24-month project aimed at developing a digital twin for sustainable agriculture. The project employs a multidisciplinary methodology to integrate various tasks efficiently. A robust management structure will oversee the project, ensuring compliance with regulations and effective stakeholder engagement. The core objective is to develop a digital twin that simulates nutrient and carbon balance, soil erosion, and crop growth using enhanced algorithms and computational efficiency. The project involves evaluating EO data sources, calibrating and validating models, and applying DT-Agro in a pilot study across Greece. Results from the pilot will be analyzed to design potential digital information services, utilizing advanced data analysis techniques. Strategic dissemination and communication activities will maximize the project's impact, facilitating the transition to digital agriculture.
Funding	This project is carried out within the framework of the National Recovery and Resilience Plan Greece 2.0, funded by the European Union – NextGenerationEU (Implementation body: HFRI). https://greece20.gov.gr

H.F.R.I. Hellenic Foundation for Research & Innovation
Greece 2.0 NATIONAL RECOVERY AND RESILIENCE PLAN
Funded by the European Union NextGenerationEU

Start Date	14/11/2023
End Date	13/11/2025

1. Introduction and Project Overview

- Project Description

The main ambition of DT-Agro is to combine recent developments in the simulation of the various components of the Agro-Hydro-System and the recently available Earth Observation open data sources with novel approaches and scientific advances on digital twins to promote their use as operational tools in Agriculture. The DT-Agro virtual representation will consist of several key components. It will include spatially distributed continuous hydrological modeling, where hydrological balance components will interact with agricultural activities. Additionally, it will involve agrohydrological modeling that covers water flows and storage within the soil-plant-atmosphere continuum. The system will also simulate crop growth and production in relation to climatic conditions, soil characteristics, and farming practices. Vegetation dynamics for naturally vegetated and non-agricultural areas will be represented, alongside the carbon stocks and fluxes in both agricultural and naturally vegetated or forested regions. The model will address nutrient balance with a focus on agricultural and grazing lands, as well as erosion and soil degradation. The overarching aim is to achieve measurable progress towards developing an operational digital twin that integrates the entire water-soil-food-environment-energy nexus into a cohesive platform. Therefore, DT-Agro will provide the means to assess the potential and the impact of agricultural policies to agricultural productivity, food security, environment, natural resources, and local livelihoods in a changing climate. The target end-users are i) policy makers requiring information and evidence (local authorities, governments, international organizations, funding agencies, etc.); ii) farmers and actors involved in the implementation of agricultural policies (farmer organizations, local experts, advisors, extension agents, etc.).
Reaching this overall goal involves the achievement of the following **specific objectives**:
SO.1. Integrate and operationalize state of the art spatially distributed modelling algorithms simulating hydrological balance, nutrients balance, soil carbon balance, soil erosion, and crop production at farm parcel scale and country level. (WP2)
SO.2. Evaluate, enhance, and utilize Earth Observation data sources and analysis techniques representing states and fluxes of the Agro-Hydro-System. (WP3)
SO.3. Pilot application of DT-Agro in Greece and assessment of current conditions and future scenarios including climate variability and mitigation and adaptation strategies. (WP4)
SO.4. Participatory design and initial development of spatially explicit agronomical services based on dynamic land evaluation and short-term and seasonal forecasts provided by DT-Agro. (WP5)
SO.5. Communicate, disseminate, and exploit the key results of DT-Agro and promote its use in policy and farming practices. (WP6)

- Data Summary

A wide range of data will be gathered and produced by DT-Agro to increase its capacity for agro-hydrological modeling and policy assessment. The Integrated Administration and Control System (IACS) geographic database, which contains about 6,000,000 farm parcels with crop kinds, irrigation specifications, and agri-environmental parameters, is one of the primary data sources. The model will get daily inputs of meteorological data, including wind speed, precipitation, and temperature, from reanalysis datasets and local weather stations. Along with information on water use for irrigation, hydrological data will include elements like precipitation, evapotranspiration, runoff, and soil moisture. The data related to crops will comprise daily logs of growth phases, biomass production, and water requirements. On the other hand, the data related to nutrients and carbon will monitor plant absorption, nitrogen and phosphorus cycles, and carbon stocks and fluxes. Soil data, focused on properties and erosion potential, will be supplemented with EO data on evapotranspiration, soil moisture, crop stress, and terrestrial water storage. Socioeconomic data, such as agriculture output levels and economic indices, will also be included. Model simulations will yield continuous data on hydrological and crop processes, as well as scenario analyses, machine learning model outputs, and spatial data integration, enabling complete and dynamic insights for agricultural policymaking and environmental impact assessment.

Significant amounts of data will be generated and gathered by DT-Agro in a number of different fields. This comprises nearly 6,000,000 polygons of agricultural parcel data from the Integrated Administration and Control System (IACS) spatial database. Daily weather data will be collected for grid cells with a goal resolution of 1000 meters. Continuous daily hydrological data will be collected, including precipitation, evapotranspiration, runoff, and soil moisture, as well as monthly aggregated summaries for long-term study. Crop data will be gathered daily, including growth phases, biomass output, yield projections, and water requirements. Nutrient and carbon data will be used to track nitrogen and phosphorus cycles, plant absorption, and carbon stocks and fluxes daily. Soil data will contain daily updates on soil parameters and erosion potential. Furthermore, high-resolution Earth Observation (EO) data will provide insights into evapotranspiration, soil moisture, crop stress, and terrestrial water storage, ensuring a comprehensive dataset for agricultural policy-making and environmental impact assessments.

Data Collection and Documentation

- Data Types

1. **Spatial Data:**
    - **Farm Parcel Information:** Polygons from IACS spatial database, including details on crop types, irrigation, and agri-environmental measures.
    - **Format:** geodatabase
2. **Meteorological Data:**
    - **Weather Variables:** Temperature, precipitation, humidity, wind speed, solar radiation.
    - **Format:** csv, Excel files (xlsx).
3. **Hydrological Data:**
    - **Hydrological Balance Components:** Precipitation, evapotranspiration, runoff, soil moisture.
    - **Format:** csv, Excel files (xlsx), grid.
4. **Crop Data:**
    - **Growth and Development:** Growth stages, biomass production, yield predictions.
    - **Water Requirements:** Vegetation water deficit and irrigation needs.
    - **Format:** csv, Excel files (xlsx).
5. **Nutrient and Carbon Data:**
    - **Nutrient Cycles:** Nitrogen and phosphorus cycles, plant uptake, mineralization.
    - **Carbon Balance:** Carbon stocks and fluxes.
    - **Format:** csv, Excel files (xlsx).
6. **Soil Data:**
    - **Soil Properties:** Soil type, texture, organic matter content.
    - **Soil Erosion:** Sediment yield and deposition data.
    - **Format:** csv, Excel files (xlsx), shape files, grid.
7. **Earth Observation (EO) Data:**
    - **Evapotranspiration, Soil Moisture, Crop Stress, Terrestrial Water Storage:** High-resolution remote sensing data.
    - **Format:** GeoTIFF, images (JPEG, PNG).
8. **Socioeconomic Data:**

- o Agricultural Production: Production volumes, economic indicators.
- o **Format:** CSV, Excel files (XLSX).

9. **Model Simulations:**

- o **Hydrological and Crop Models:** Simulation outputs on water balance, crop growth, irrigation needs.
- o **Scenario Analysis:** Historical data, future climate scenarios, crop patterns.
- o **Format:** CSV, text files, geodatabase.

- Methodology

DT-Agro will be built around an inhouse developed, entirely spatially distributed, continuous agrohydrological model, able to provide a gridded output of the main hydrological balance components, crop water interactions and water requirements, as well as crop growth and development on a daily temporal step at farm parcel scale for broad regions. The proposed approach is directly relevant to agricultural policies design, implementation, and evaluation as it allows the dynamic estimation of agricultural water abstractions, nutrients and carbon balance, erosion, and agricultural production, facilitating the estimation of corresponding impact indicators [e.g. (Soulis et al., 2020)]. This approach addresses the key challenges
related to the small size and fragmentation of farm landscape, the great spatial and temporal variability, and the severe data scarcity. In the core of the developed approach is AgroHydroLogos model (Soulis and Dercas, 2007, 2010; Soulis et al., 2016, 2020), which is developed by the PI. This model operates as an extension of the GIS software package ArcGIS (Environmental Systems Research Institute-ESRI, Redlands, CA, USA) making
use of its geospatial analysis and spatial data management capabilities, which is critical for the needs of this study involving vast amounts of spatial input data and results. The model can provide gridded outputs of the main hydrological balance components, vegetation water deficit and irrigation water needs, on a daily or monthly basis at country level, yet it is simple and flexible, as its conceptual scheme draws on well established but simplified methods for the simulation of the involved hydrological processes.
An ever increasing number of more sophisticated hydrological models are continuously evolving in the last decades including data-driven empirical models [e.g. (Wang et al., 2021; Yuan et al., 2021)], conceptual models [e.g. Hydrologiska Byråns Vattenbalansavdelning (HBV) (Seibert and Bergström, 2022)], and physically based hydrological models [e.g. Systeme Hydrologique Europeen (MIKE-SHE) (Abbott et al., 1986; Jaber and Shukla, 2012), Soil & Water Assessment Tool (SWAT) (Arnold et al., 2012), WetSpass (Batelaan and de Smedt, 2001)]. Yet, all models have their strengths and drawbacks making them suitable for specific applications and conditions but limiting their applicability for other purposes or under different conditions. For example, data-driven (empirical) models can accurately simulate streamflow in many cases but are characterized by lack of transparency and it is difficult to be adapted to regions and conditions other that those used for their calibration (Mount et al., 2013). In conceptual models, the oversimplification of hydrological processes may limit the adequate representation of the spatial variation of hydrological responses and their parameters are not directly related to measured catchment characteristics and not completely understood (Abebe et al., 2010; Tshimanga and Hughes, 2014). Physically based models are more computationally intensive and require more detailed data representing spatial heterogeneity, which are big challenges in data scarce regions (Soulis et al., 2020; Hirbo Gelebo et al., 2022). Furthermore, lumped or semi-distributed models [e.g. SWAT (Arnold et al., 2012)] cannot describe the spatial patterns of cultivations at farm parcel scale, while monthly or larger temporal steps [e.g. WetSpass (Batelaan and de Smedt, 2001)]
cannot effectively describe crop dynamics. In DT-Agro, the selection of an inhouse developed

model provides the flexibility to evolve and adapt to the specific challenges and requirements of the attempted digital twin development, to integrate the additional components related to nutrient and carbon balance, soil erosion, and productivity, while it will be easier to be adjusted soon to simulate additional components and scenarios.

In parallel, it may provide better insight in the drivers and parameters behind agricultural production, inputs requirements, and environmental impacts. The model has been already applied and tested under Greek conditions for various applications, alleviating the time and financial constraints of the current study. Key previous applications are the assessment of the water abstractions in agriculture in Greece in the framework of the evaluation of the CAP agri-environmental measures (Soulis et al., 2020), the strategic planning of RDP in Greece for the next programming period, and the large-scale preliminary assessment of small hydropower potential in Greece (Soulis et al., 2016). Previous achievements provide a solid foundation for the successful implementation of the proposed system and its broad potential applications and impact.

The use of a daily temporal discretisation is able to describe with reasonable accuracy the hydrological processes and crop dynamics, while it allows the simulation of long periods and the utilisation of readily available weather data, reanalysis data, or data related to future climate scenarios.

One of the main purposes of DT-Agro is to be directly relevant to the design, implementation, and evaluation of agricultural policies. An important challenge to this objective is the very small size, the fragmentation, and the vast spatial variability of the farms in Greece, which is also the case in many other countries. To overcome this challenge the developed approach utilizes the information included in the Integrated Administration and Control System (IACS) spatial database. This database includes the spatial extend and detailed information

for each farm parcel in the country (at least for the parcels professionally cultivated) and includes among others information about the crop, the agri-environmental measures applied in each parcel, irrigation status, irrigation system and water source. This database is updated on an annual basis. Corresponding databases exist in all EU countries as it is a requirement of RDP implementation, accordingly the developed system will be able to operate in other EU regions as well. To this end, a special algorithm is included in the developed system securing the compatibility with the IACS spatial database. Specifically, this algorithm installs two-ways links between each farm-parcel polygon (>6,000,000 polygons) of the IACS spatial database and the nearest matching grid cell of the model. In this way, the developed approach is able to utilize the information included in the IACS database at farm parcel level and provide results for each farm-parcel facilitating further analysis.

Agrohydrologos simulation algorithm operates around the hydrological balance of the reference soil volume considering the water flow in the soil plant atmosphere continuum based on simplified but physically based conceptual approaches such as simplified runoff, infiltration, evapotranspiration and routing algorithms, and production functions (Soulis and Dercas, 2007; Soulis et al., 2020). This allows the efficient integration of nutrients and carbon stocks and fluxes simulation algorithms as well as soil erosion simulation methods to the base algorithm.

The first key activity is the assessment, improvement, adaptation, and integration of simplified but robust algorithms for the simulation of nutrients and carbon balance, soil erosion, and crop growth in AgroHydroloLogos code (WP2). Soil erosion and sediment yield with runoff will be modelled with the well established Modified Universal Soil Loss Equation (MUSLE) considering aboveground biomass, residue on the soil surface, and the involved parameters for each species (Efthimiou et al., 2022).

Sediment deposition and degradation and sediment transport are typically simulated as a function of peak channel velocity, which is well suited to the spatially distributed runoff routing process included in AgroHydroLogos model (Soulis et al., 2020). Nitrogen and phosphorus cycles, including plant uptake of nutrients, the mineralization of organic nutrients in plant residue, organic and inorganic fertilizers, nutrients removal by crop produce etc. will be simulated in parallel with a detailed process-based simulation of plant growth considering the effects of plant cover on nutrient balances (Hassall et al., 2022). Similarly, carbon cycle dynamics considering both the magnitude of carbon flux/yield and its response to environmental (climate and soil) variability will be simulated (Zhou et al., 2021). In parallel, enhancements of computational efficiency of the algorithms to make feasible the application DT Agro in large areas and in fine spatial resolution will be attempted. The strategies that will be used include: recoding and improving the algorithms performance; utilizing further and improving novel features of AgroHydroLogos model such as the serialization algorithm (Soulis, 2013) and the dynamic spatial interpolation of meteorological data (Soulis et al., 2020); using dual spatial resolution (lower spatial resolution for weather variables with target cell size 1000m and higher spatial resolution for agrohydrological components with target 100m cell size; parallelisation of the algorithm and improvement of the serialization technique to process separately hydrologically independent regions to allow execution of smaller parts for calibration or real time update purposes (WP2). The second key activity is the study and assessment of EO data sources and techniques that will be utilized to provide input data or to evaluate and adapt the DT-Agro (WP3). Emphasis will be given to open, high temporal and spatial resolution datasets on potential and actual evapotranspiration (Koppa et al., 2022; Wu et al., 2022; Xie et al., 2022), soil moisture (Babaeian et al., 2019; Liu and Yang, 2022), crop stress (Ahmad et al., 2021), crops and vegetation development and patterns and related parameters such as crop coefficients (Soulis et al., 2020), soil erosion and degradation (Straffelini et al., 2022). The analysis of terrestrial water storage changes from GRACE mission will be also investigated (Gerdener et al., 2022). DT-Agro algorithms will be adapted to be directly compatible with the EO data inputs. The target is to be able to utilise two-way data flow to allow the initialization or adaptation of state variables based on real time inputs from sensors networks, IOT, and EO data and to be calibrated based on the spatial representation of the system states and fluxes coming from EO or other spatially explicit data sources (e.g. sensors networks, IOT, crowd sourcing) (WP2 and WP3). An important task for the successful implementation and evaluation of DT-Agro is a pilot study all over Greece at farm parcel scale, based on previous experience of the research team and previous successful applications of AgroHydroLogos. The scenarios that will be examined regard historical data, future scenarios including climate projections and crop patterns scenarios, and-short term forecasts that may support future operational services. DT-Agro calibration and evaluation will be based on the collection of existing data from databases and local irrigation organizations on water consumption, crop production, hydrology, and soil properties. These data will be used in parallel with the EO data and will be also utilized for EO data evaluation. The pilot study will serve calibration and testing but also an initial but meaningful assessment and analysis of the obtained results for various scenarios. (WP4).

The successful implementation and application of DT-Agro will result on a vast volume of spatial data. The effective utilization of these data and assimilation with other data sources such as socioeconomic data and EO data is really challenging. Accordingly, an important activity of the project is the elaboration of machine learning, big data analysis, and data assimilation techniques for the analysis of the huge amounts of data originating from DT-Agro and earth observations (WP5). At the same time the characteristics of potential digital, spatially explicit services based on these results will be defined in collaboration with key stakeholders. Considering the important financial and time constrains of this project this last task will target mostly in the participatory design of such services and the provision of some paradigms and not the development of fully operational services, which will be the subject of future research.

Finally, the project will involve effective communication, dissemination, and exploitation activities that will allow the enhancement of the scientific, societal, and economic impact of the project (WP1).

- Data Sources

Existing data will consist of: Hydrological (Precipitation i.e. data on rainfall patterns and intensity, streamflow, evapotranspiration), Agrohydrological (soil moisture and irrigation data (irrigation schedules, methods and volumes), Agricultural (crop growth and yield, soil properties, farming practices and climate data), Environmental (vegetation cover, nutrient balance), Earth Observation
New data will consist of field measurements to collect data on soil moisture, crop growth, and hydrological variables using sensors and manual sampling, remote sensing (utilizing recent and ongoing satellite missions to acquire up-to-date imagery and remote sensing data, implementing drone-based surveys to collect high-resolution data on specific areas of interest)

- Metadata

To ensure the data collected and produced for the DT-Agro project is well-organized, understandable, and reusable, comprehensive metadata and documentation will accompany each dataset. This will include detailed information about the methodology of data collection, data organization, and any processing steps applied.
General metadata elements include a descriptive title of the dataset and a detailed description of each dataset, names and affiliations of the data creators and contributors, dates of data collection and creation and geographic area covered by each dataset (including coordinates). For each dataset, the above-mentioned characteristics will be provided in the metadata, where needed. For datasets produced during the project the methodology will be provided with reference to scientific articles or reports in which it was published. Data and metadata will be stored on a cloud system and hard disk drives.

2. Data Storage and Security

- Storage Solutions

Data will be backed up to cloud storage and on hard disk drives. Cloud storage system with regular backups and data recovery functionalities. All contributors will have access to the cloud.

- Backup Procedures

Daily backups depending on the progress of analyses.

3. Data Sharing and Access

- Access Policies and Data Sharing

The Principal Investigator (PI) and the other team members will have full access to all data types and formats for oversight, and each will have access to datasets specific to their tasks, such as meteorological, hydrological, crop, nutrient, carbon, soil, and Earth Observation (EO) data. Following the study, some aggregated datasets and data supporting published research findings will be made public via an open-access repository. DT-Agro outputs and outcomes accessible and usable by scientists, practitioners, and relevant public stakeholders through open and easy access to all related results and outputs of the project using various channels such as online data access, web

GIS, open code, documentation. Outputs such as open access research database, user manuals, scientific articles, position papers, policy briefs will be documented and disseminated based on European Commission Open Access rights. Publications will be optimally open access and key relevant national and international congresses will be selected to maximize impact.

4. Data Preservation and Archiving

- Long Term Storage

To ensure the accessibility, security, and utility of data gathered and created by DT-Agro for an extended period, a complete approach will be put into place. Data will be stored in secure cloud drives, as well as hard disk drives. Data will be saved in reliable formats like csv, GeoTIFF, shp and others, along with full metadata for future usage. A data governance strategy will be developed, reviewed by the Principal Investigator (PI) to manage responsibilities, access controls, and frequent backups, guaranteeing redundancy and avoiding data loss.
Periodic reviews will ensure the data remains accessible and updated, with migrations to new formats as necessary. Efforts will be made to secure sustainability funding for ongoing data maintenance and curation, thus preserving the data for future research, policy-making, and advancements in agro-hydrological modeling.

- Retention Period

Data collected and generated by DT-Agro will be retained for a minimum of 10 years following the completion of the project to support ongoing research, policy-making, and future studies. Data can be reviewed at the end of the 10-year retention period to determine if it remains relevant and useful. If data is still needed for ongoing research or policy purposes, the retention period may be extended.

5. Responsibilities and Resources

The project PI will provide overall leadership and strategic direction for data management and overseeing compliance with data governance policies. The GIS Specialist will manage and analyze spatial data, maintaining GIS tools and software. Researchers and PhD Students will perform detailed data analysis, conduct modeling and simulations, and document methodologies and results according to their tasks. PhD students will also contribute to the documentation of data handling procedures and methodologies, ensuring transparency and reproducibility. Additionally, they will support the GIS Specialist by managing spatial data and utilizing GIS tools for their research. Their work will be supervised by senior team members, including the PI, to ensure alignment with project goals and adherence to data management protocols.

Project Risk Management Plan (PRMP) for the project "Spatially Explicit Digital Twin of the Greek Agro-Hydro-System DT-Agro"

As shown in Table 1, to address potential cost issues due to higher prices, financial problems, and inflation, which are deemed to have a low likelihood, DT-Agro will extensively utilize its own developed tools and available expertise, maximizing the use of internal resources to meet project objectives. For issues related to models' accuracy, predictability, and adaptation, which have a medium likelihood and impact WP2 and WP4, thorough testing under various conditions will be implemented. Continuous adaptation, calibration, and validation will be performed throughout the project, with many alternative algorithms tested to ensure robustness. Data availability issues, considered to have a high likelihood and affecting WP2, WP3, and WP4, will be mitigated by implementing specific actions and alternative solutions to ensure project continuity and data reliability.

Table 1. Critical risks for implementation and proposed risk-mitigation measures

Description of risk	Work package (s) involved	Proposed risk-Mitigation measures
Problems with the required cost due to higher prices that planned, financial issues, inflation (likelihood: Low)	All WPs	DT-Agro utilizes at a great extend own developed tools and already available expertise. It will extend the use of own resources to achieve objectives
Models' accuracy, predictability, adaptation issues (likelihood: Medium)	WP2, and WP4	Thorough testing under various cond itions. Continuous adaptation, calibr ation, validations during the project d uration. Many alternative algorithms are tested.
Data availability issues (likelihood: High)	WP2, WP3 and WP4	Specific actions and alternative solutions to overcome data scarcity issues are planned in the project.