Data Science Research Infrastructure 2022

Project title:

A high-performance data-driven agroecosystem modelling platform for developing agricultural systems with minimum environmental impact (AgroEco-HPM)

Application number: 0077631

Applicant name: Sune Darkner

Main Applicant position: Associate Professor

Administrating institution: Københavns Universitet

Grant period in years:	5
Total requested budget:	kr 14.851.437
Received from administrating institution:	631.669
Received from other sources:	
Applied from other sources:	

Introduction

The agricultural sector contributes between 1/4 and 1/3 of the total greenhouse gas emissions[1] and to a large part of the nutrient and pesticide pollution in surface- and groundwater. In Denmark, goals are a 70 % reduction of greenhouse gas emissions by 2030 and a 37 % reduction of the present N-leaching. However, expectations[2] are that emissions from agriculture will be constant towards 2030, as the impact of potential measures is insufficient and difficult to quantify. Furthermore, the Danish government recently requested the research community to support a complete overhaul of the present regulatory system for agriculture to include climate gasses, N-losses (leaching and gaseous) and farm balances for carbon (C) and nitrogen (N) to ensure consistency in measures and approaches (call). From a global perspective, these challenges must be addressed while considering food security for a growing world population.

Agricultural C- and N-processes are closely interlinked, and to understand and describe the C and N-balances, mechanistic models are required. They can describe observations at local scale, but predictions outside their calibration range are not reliable. To address the challenges, we need to work across scales and scientific areas, consider the mass balances and handle the vast heterogeneity in weather, soils, crops and management. The analyses require approaches ranging from detailed process modelling to novel integrated methods such as artificial intelligence (AI), where we work backwards from sensor information (e.g., satellites, drones), and combinations of the two.

The building blocks for an agricultural research platform exist in Denmark. Data on weather, soil, agricultural land use and hydrology are available. The 1- or 2D model "Daisy", developed at University of Copenhagen (UCPH) for agricultural land, is the most used mechanistic soil-plant-atmosphere-model in Denmark, simulating water, heat, C and N in soil and plants, as function of weather and agricultural management. However, its demands for input data and multi-disciplinary operator expertise are often hurdles for its application, not least with respect to upscaling, as users must build their own data management systems for integrated analyses.

We propose a common infrastructure, GeoDaisy, that combines geo-referenced and temporal data handling for multiple data sources and structures, mechanistic/simulation modelling, statistics, AI and supercomputing. The infrastructure must facilitate co-operation and integration of knowledge and methods across scientific fields and scales, speed up individual research projects, and allow novel evaluations of land-use changes or agricultural management practices and their environmental impact.

Research based on GeoDaisy will lead to more coherent approaches for national assessments of leaching and gaseous losses in agriculture, and better cost-efficiency estimates of different regulatory measures, which benefits legislators, agricultural extension and farmers. GeoDaisy will also support research in adjacent fields (crop breeding and cultivation, drought resilience, plant protection products (i.e., plant biologicals, biostimulants). Thus, the tool will support the green transition and other elements of public, environmental and agricultural interest.

There are no tools available, similar to the suggested GeoDaisy, for agricultural land. Model systems (1-D only) with the same aim as Daisy exist (e.g., CoupModel, ANIMO, CROPSYST, SOIL/SOILN, APSIM, STICS). Daisy is particularly strong with respect to soil processes of relevance to Denmark and northern Europe, for which it is calibrated. None of the models is embedded in a system like GeoDaisy or contains the same range of options.

The National Water Resource Model (NWRM) is used to investigate groundwater resources, future groundwater conditions and streamflow, but lacks a detailed root zone model. Efforts with mixed success were made to couple it with Daisy[3,4]. The National Nitrogen model couples NWRM with input from the empirical model NLES5, which lacks proper description of agricultural mitigation measures. Danish Centre for Environment and Energy (DCE) at Aarhus University (AU) applies the SWAT model for distributed modelling, at sub-basins level. It is coarser and not publicly available.

The Daisy model today

Daisy[5,6,7] is a systems model able to simulate and integrate processes in soil (vadose zone), plants, and the lower atmosphere. It was developed at UCPH since 1986 through national and international research projects. The model has users in several countries, but most in Denmark, Europe, and China (see Collaborator- and publication-list), indicated by approx.1500 homepage hits and 400 model downloads per month, and 75 receivers of DaisyNews. From 1991 onwards, Daisy has been financed on project basis, mostly focusing on new developments. This structure does not provide resources for maintenance and adequate user support, which is problematic for new users and when results are used for policy support.

Daisy describes the detailed dynamics of water, heat, C, and N on well-drained agricultural fields[5,7] (crop growth, additions, transformations and fate of C and N, incl. leaching and gaseous losses (CO_2, N_2/N_2O)) and pesticide fate[7,8,9]. Recent work includes formation, losses and fate of biotoxins in ferns[10], soil compaction, conservation agriculture [9], precision agriculture[11,12], plant breeding[13], and studies of climate change and future yield stability[e.g. 14,15]. Examples of combining data science methods with mechanistic modelling exist for parameter optimisation[16], identifying (via machine learning) optimum conditions for pesticide spraying[17], investigating the potential for improving precision agriculture**Error! Bookmark not defined.**[12], and letting model results guide machine learning for soil classification. Also, examples of assimilation of satellite data exist[e.g. 18].

Daisy is open source. The compiled program and source code can be downloaded from the Daisy homepage and Github, respectively. The original Fortran code was re-written in C++ (end 90'es[6]) using an object-oriented structure for organising the 170.000 code lines. Questions are answered in the Daisy forum.

The Department of Plant and Environmental Sciences (PLEN), UCPH aims to develop Daisy further and expand its use. Teaching includes an annual Daisy PhD course and a new MSc specialization (Land Use and Modelling, courses: Land use and Environmental Modelling, Applied Agrohydrology, and Modelling Soil-Plant-Atmosphere Systems), at UCPH. At the Sino-Danish Center involving all Danish Universities, including Dept. of Agroecology (AGRO), AU, Daisy is taught in the MSc Water and Environment program – course "Modelling water and nutrients management in agroecosystems" (module 'Integrated Water Management and Legislation'), and used in the WE research activities.

Platform development needs

Many projects in the field of agriculture, environment and climate change include parameterization of a large number of Daisy model setups, either to cover geographical areas[18,3,4,20], local variation in parameters (heterogeneity)[19], or uncertainty of parameters[20,16]. Daisy has been combined with a variety of software (R, Python, Java) to perform simulations on hardware, ranging from workstations to supercomputers[17,21], and analyse the output with various statistical methods**Error! Bookmark not defined.**[12], including deep learning[17]. Some have created reusable frameworks (RDAISY toolbox)[16], but most use ad-hoc solutions. Thus, shared tools for data handling, analysis and supercomputing are required. To overcome the limitations of mechanistic modelling regarding predictability and scaling, AI methods and modelling must combined. The mixture of scientific fields and scales makes it necessary to ensure that new process parameterisations, datasets, and documentation are shared and available to the multidisciplinary user community, enhancing research efficiency and quality. Finally, a system of code quality assurance is required to support model credibility, and a system of long-term model development and management (incl. funding) is required for its sustainability.

Purpose

The aim of the proposed infrastructure (**GeoDaisy,** Fig. 1) is to develop an advanced geo-referenced modelling platform that combines acquisition and management of multivariate, spatio-temporal datasets, advanced data science (AI-) methods, and mechanistic model simulations of water, carbon, and nutrient pools and fluxes (i.e., crop production, C-sequestration, N-leaching, gaseous losses (i.e. CO_2, N_2/N_2O), as well as all field-relevant solutes (i.e. pesticides, plant toxins). Automatic data flows, coupled with High-Performance Computing (HPC) will allow simulations over time and space (one to several spatial dimensions and at subfield resolution, that:

- provide easy access and optimal utilization of existing georeferenced databases and data providers (weather-, soil-, groundwater-, satellite-, topography and other data),
- include advanced tools for pre-processing of data (quality control, data cleaning, gap-filling, data transformation, data reduction, descriptive statistics, georeferenced inputs, etc.)
- implement next-generation differentiable physics in parts of GeoDaisy for integration with AI and parameter optimization and extend with algorithms used for model calibration and the estimation of the associated parameter uncertainty based on frequentist and Bayesian statistics.
- allow predictive modelling using Bayesian data assimilation algorithms and assessment of the associated uncertainty,
- include post-processing tools that allow data visualization, data extraction and analysis, and application of machine learning and other AI methods on input data, simulation data and mixtures of the two,
- Include post-processing tools for geo-temporal and geospatial modelling (GIS-friendly output).

GeoDaisy will be consolidated for future development, be well documented, and facilitate a method of storing “best modelling practices”.

The institutional framework and the core team

DIKU and PLEN have over the past few years established a strategic collaboration focused on agriculture and data, and an upcoming EU-partnership. Other efforts include several submitted joint grant applications as well as projects, including a deep commitment to the development and writing of the upcoming Agriculture of Data partnership. The efforts aim at taking advantage of the high-quality data available in the Danish data eco-system to support decision making and environmental monitoring in the Danish agricultural sector. These efforts include the participation in the AI-Agri network with some of the most prominent partners in the Danish Agri-sector, including breeders and farmers. This proposal will be the future cornerstone in this continued strategic collaboration and will enable KU to place itself at the center of agri-environmental monitoring and modelling supporting all the other ongoing efforts within these areas.

The scientific development of GeoDaisy will become an interdisciplinary collaboration between the Department of Computer Science (DIKU) and the Agrohydrology (Agh) group, PLEN, headed by Associate (Assoc.) Professor (Prof.). Sune Darkner, an expert in the integration of data-driven solutions for mechanical simulation and AI, see the hosting letter and [1] (numbers in brackets refer to upload numbers in list of core collaborators). His colleagues at PLEN are described below.

Assoc. Prof. C. T. Petersen (CTP, [2]), group leader of Agh carries out the scientific supervision of Daisy today. CTP is a soil physicist and has > 20 years’ experience with combining field measurements and modelling. Per Abrahamsen (PA) (AC-TAP) who implemented the present C^{++}-code[Error! Bookmark not defined.], carries out Daisy code improvement and maintenance as well as advanced modelling in the fields of soil physics, hydrology and modelling of plant growth, water, carbon, nutrients, colloids, pesticides and biotoxins. The group cooperates with Prof. Diamantopoulos [31,32], Univ. of Bayreuth, who until recently was Assoc. Prof. in Agh. We expect this post to be announced in 2022. On organic matter and N_2O-

dynamics, Prof. L.S. Jensen (LSJ [3,4]) and his group assists Aghyd. Cooperation with other groups (Env. Chemistry, Crop Physiology and Production systems, Microbiology) are typically project- or student-based.

Three new AC-TAPs (NN1-3) will be employed, who together with PA will form the core implementation team of GeoDaisy. NN1 will be hired at PLEN for the project duration, mainly work on WP3-5 (see below), and continue working after the project termination, together with PA. NN 2 will be hired for the first three years and mainly work on WP1-2. NN3 will be hired for 5 years and work on test frameworks (WP6) and dissemination (WP7). NN2 and NN3 will be employed by DIKU. A project manager (NN4, shared) will work 1/3 man-year per year over the project period.

The core implementation team must ensure the project implementation, daily maintenance, and quality assurance of the system, so up to date versions of e.g., crop modules and datasets for standard analyses are available in parallel to the ongoing research activities and code developments.

Core collaborators and users

The core collaborator (CC)-group for the infrastructure is illustrated in Fig.2 and made up by members of the following groups, supporting the management in their fields of expertise. The Daisy users mainly consist of researchers, MSc.- and PhD-students:

Present Daisy users: The sections "Plant and Soil Science" [3, 4], "Crop Sciences" [5, 6], "Environmental Chemistry" [7,8] at PLEN, UCPH, covering the fields of soil science, soil chemistry, organic matter dynamics, plant nutrients and plant growth and plant toxins. Dept. of Food and Resource Economics (IFRO, UCPH) [9-11] focus on economic assessments and precision agriculture **Error! Bookmark not defined.**. At AU, AGRO, as well as the Danish Centres for Food and Agriculture (DCA) [12-17], DCE [20, 21] and Circular Bioeconomy (CBIO) [14] use Daisy for education and research in a wide range of topics: e.g. precision agriculture, irrigation, climate change assessments, N-balances, and the national monitoring program, increasingly linking model- and data-driven approaches. DTU Engineering Technology [22, 23] do fertilization optimization combining Daisy with satellite data and upscaling solutions. In Germany, staff at Univ. of Bayreuth [31, 32], Univ. of Hohenheim [33, 34], and at Martin-Luther-University Halle-Wittenberg [35,36] co-operate on soil hydraulic processes and joint PhD-courses on model use, high throughput methods (NIRS, MIRS and remote sensing), and new aspects of agronomic management and data.

Future users and data suppliers: SEGES Innovation [24, 25], the main agricultural extension organization in Denmark, will provide data and expect the proposed infrastructure to feed into their "crop manager" system used by farmers. Danish Meteorological Institute (DMI) [26, 27] and Geological Survey of Denmark and Greenland (GEUS) [28-30] will provide data but expects future use of the system for e.g. drought assessment or as input to NWRM/national N-model. Furthermore, the new pioneer-centre Land-CRAFT, headed by Prof. Klaus Butterbach-Bahl [18, 19], based at AU-AGRO but also with UCPH as partner, sees GeoDaisy as a valuable tool in their suite of activities.

Collaborators: Agency for Data Supply and Efficiency (SDFE), who host different relevant datasets, e.g. topography and "Hydrological information and prognosis system" (HIP) [37]. Other collaborators are Assoc. Prof. K. S. Larsen (coordinator of AnaEE (Analysis and Experimentation on Ecosystems) and Head of AnaEE technology centre) and Prof. P. Gundersen, both at Dept. of Geosciences and Natural Resources Management, UCPH, and Dr. J. Takáč (Soil Science and Conservation Research Institute, Bratislava, Slovak Republic), long-term Daisy user.

We estimate that the distribution between users in 2028 will be about 30% at UCPH (15% by applicant group), 30% at AU, 30% other CCs, and 10% others.

Work Packages

The project tasks have been separated into Work Packages (WP), which are described in the following, and an overview of the project timeline, WP task duration, milestone and deliverables are presented in a Gantt chart (Tables 1 and 2).

WP0, Project Management (SD, Agh-group, NN4, Steering Committee)

The management of the project, including internal and external administrational, financial and management reporting, and risk management, will be conducted by the main applicant in cooperation with the Agh-group, PLEN, and the IMAGE group, DIKU at UCPH. SD and PA will carry out day-to-day management of code implementation and staff supervision. A Steering Committee (SC) representing the CCs will coordinate, prioritize, and decide on the different proposed tasks. Task-oriented sub-groups will advise on specific tasks.

T0.1: The Agh-group/IMAGE will hire a project manager (NN4) with the main role to undertake mainly administrative work (Q1 2023) and three NN AC-TAPs, described above.

T0.2 Establishment and management tasks of the SC (responsibilities outlined in separate section). A kick-off meeting for all CCs will be held in Q1, 2023 with the tasks to select a representative SC, create several task-oriented working groups, WG1-X (Q2 2023) based on the WPs described below, and delegate relevant tasks. The SC will meet every quarter for the first 3 years and semi-annually after the third year and beyond.

T0.3. WGs will be flexible expert working groups focusing on tasks completion (Table 1). Meetings of the WGs will be a function of the implementation schedule. The WG's report back to the SC for decision endorsement.

WP1, Implementation of Python library for running parallel Daisy simulations (ParaDaisy) (NN1, PA, SD)

For Task 1.1-3, detailed specifications and work descriptions will be presented to the SC for discussion and agreement.

T1.1 Definition, implementation, and documentation of the input format for ParaDaisy.

T1.2 Definition, implementation, and documentation of the output format. ParaDaisy will collect the Daisy output from parallel simulations as well as the meta-information, combined, and converted into one of several standard formats, and make data available to other tools in the system or contemporary external tools, e.g., software packages for machine learning/AI-methods.

T1.3 Implementation and documentation of the ParaDaisy tool for parallel simulation of Daisy, seamlessly utilizing the available cores in normal Linux, Windows, or MacOS systems, or a batch system on a supercomputer like Computerome, IMAGE-cluster or ERDA, and cloud services.

T1.4 Continuous maintenance and integration of ParaDaisy with tools from WP2-5.

WP2, Database extraction (NN1, PA, SD)

A WG will propose data sources of relevance for Daisy simulations or for combination with simulation data for AI-analysis. For data sources already identified, T2.1-3 are defined:

T2.1 Implementation and documentation tools for extracting data from dynamic databases with high temporal resolution, e.g., weather data from DMI public API, satellite data, sensor data etc.

T2.2 Implement and document tools for extracting static data like soil, drainage conditions and relevant hydrological information (HIP), topography, and crop data from AU, SDFE, GEUS and SEGES Innovation.

T2.3 Implement and document tools for extracting management information (actual or typical based on farm or soil types) from SEGES Innovation or other sources. Some data may require permission of use.

T2.4: Implement links to additional databases (governmental, EU) identified by the WG.

WP3, Advanced methods in model simulation (NN2, PA, SD)

Tools for performing inverse modelling (model calibration), sensitivity analysis and uncertainty methods benefit from the ParaDaisy functionality and implementation of differentiability in key models. The ability to compare simulations with real-time data (e.g., from satellite or drones) provides the possibility of data assimilation and providing forecasts. WG's will propose methods to be implemented and documented. For task T3.3, it is recognized that present research does not cover all relevant aspects.

T3.1: Inverse modelling. Local (differentiable models, gradient-based, simplex, etc) and global optimization algorithms (genetic, heuristics,) will be coupled with ParaDaisy, with algorithms tailored to different uses.

T3.2: Methods for sensitivity analysis and uncertainty assessment. Local and regional sensitivity algorithms will be coupled with ParaDaisy. Uncertainty propagation algorithms (Monte Carlo, GLUE, etc), and uncertainty analysis (First Order Second Moment method, Markov Chain Monte Carlo), will be implemented for estimating uncertainty of estimated parameters from T3.1.

T3.3: Implementation of Bayesian data assimilation algorithms and assessment of the associated with the forecast uncertainty.

WP4, Post-processing (NN2, PA, SD)

Tools will be developed for further processing of the ParaDaisy output, including tools for visualizing results (e.g. forecast time series with uncertainty bounds), conversion into GIS format, performing statistical analysis, or preparing data for input to machine learning algorithms, based on WG recommendations. Example scripts for inspiration to users will be prepared and documented for:

T4.1 generating GIS-friendly output files.

T4.2 statistical analyses on output (e.g., description of sensitivity and uncertainty) and visualization tools.

T4.3 combining simulated and measured data with machine learning.

WP5, Knowledge sharing and Gaps (NN2, PA, SD)

We will consolidate the basis of model simulations by taking advantage of the experience present among CC's by reviewing earlier parameterizations and process descriptions based on recent data and research. CCs must allow that the TAP staff collect and test individual parameterizations, or use collected datasets. A WG will follow this work for most of the duration of the project.

T5.1 Identify, prioritize, and document major gaps in knowledge about soil types and management practice in Denmark, including crops and crop varieties.

T5.2 Where available data sets allow it, improve descriptions/parameterizations of Daisy processes.

T5.3 Collect and describe "best modelling practices" for areas where knowledge exists.

WP6 Quality Assurance (NN3, PA, SD)

A modern agile software paradigm will be applied to ensure the long-term existence and development of GeoDaisy through the following tasks:

T6.1 Identify and select a suitable framework for having repositories and the test framework.

T6.2 Set up the testing environment including unit, regression, and integrations tests.

T6.3 Conduct a user study of users' needs and the GeoDaisy application context (year 3 or 4), identify possible improvements of GeoDaisy and adjust accordingly.

WP7 Dissemination, outreach, and teaching (NN 2&3, PA, SD, SC)

T7.1: A project homepage under Daisy.ku.dk will be established with project description, current project status, and links to all finished deliverables as soon as the development allows, together with documentation and example cases. Dissemination of new developments will be spread via the homepage, Daisy online meetings and DaisyNews.

T7.2: To demonstrate the use of the tools to Daisy community, a WG and the TAP team will define example cases, covering the main tools and functionalities from WP1-5. The example cases must be run, documented and published.

T7.3: Conduct task completion presentations for all infrastructure users (orange lines in Table 1) as well as **t**wo workshops (year 3 and year 5) organized by the Agh-group/IMAGE and SC, training staff from collaborating institutions in the use of the tools.

To build a larger user community, an online educational platform will be initiated with video lectures and online exercises. The topics will be theory related to the GeoDaisy tools and practical use. It will build on experiences from existing courses, the example cases and the workshops.

T7.4 Identify potential users and venues for dissemination, outreach, and teaching of GeoDaisy such as workshops, conferences fares, and online resources. Identify the best suitable platform for the online version of a GeoDaisy course(s), inspired by Coursera. Develop a lightweight introduction to GeoDaisy and its applications, and modularized online versions of existing course content and potential extensions such that the individual components can be used on BSc and MSc levels.

The developed teaching tools can be used in existing courses at UCPH and AU (see above), or by others, and support new courses (MSc-level: "Distributed Agro-Ecological modelling and land use change"; an extension of the 1-week Daisy PhD-course etc.) and supplement other courses (e.g., Introduction to data science, Big data in biotechnology), See also support letters [21, 30, 34]. The core of Daisy and its future development including the simulation framework will be integrated as examples and cases in Computational Methods in Simulation, Advanced topics in Image Analysis and Numerical Optimization at DIKU as well as be a foundation for BSc, MSc, and PhD thesis work.

Steering committee and responsibilities

The SC consists of 5-7 selected members of the CC group. Membership may change over time. The SC will be in close contact to all CCs, assign CCs to WGs on specific subjects, decide on tools to be implemented, considering expected future use as well as time and resources available, coordinate the implementation when requiring links between institutions and groups, and monitor progress. Through the CCs, it will facilitate access to data and expertise in the represented organizations and develop modes of access to important data that is not freely available but can be made available for research projects (at a cost or by agreement). Furthermore, the SC must develop a (simple) system of fees to cover project-related data costs, use of supercomputing (if necessary) and platform maintenance. In addition to a fee system, the committee should also consider the possibility that model output (generated in research projects over the 5-year period) could have a commercial value that could support maintenance of the platform.

Future use of GeoDaisy & Business plan

GeoDaisy is an extension of an existing PLEN infrastructure, and because the need for such tools has been identified during earlier projects, we expect fast adaptation in new research projects and education for all CCs. For the tools not geographically linked to Denmark, we expect adaptation by users abroad to increase as training courses are provided.

The cross-institutional SC is expected to continue after the duration of the project. The employment of TAP help in the project will ease a future generation shift in the TAP staff, an issue of vital importance for continuity and future maintenance of the system.

The SC has been given the task of developing a project related user fee to support future maintenance considering possible future commercial income possibilities.

Key performance indicators of success of GeoDaisy will be

- the number of people on training courses concerning GeoDaisy in y3, y5 and the first three years after project completion. (Goals: 15, 25 and 50+)
- the number of articles generated using the system (nationally and internationally) five years after project completion, (Goal: at least 50, considering the delay between knowledge of the tools, obtaining funds, carrying out research and publishing.)
- the number of courses introducing students to GeoDaisy (from 2027 and forward) at Danish universities (not less than 75 per year), and
- adequate funding provided for maintenance of GeoDaisy two years after project completion (2 AC-TAPs per year).
- Two projects applying the platform for decision making in practical agriculture and regulation the year after project completion.

And ultimately implementation of results of the conducted research in practical agriculture and public administration.

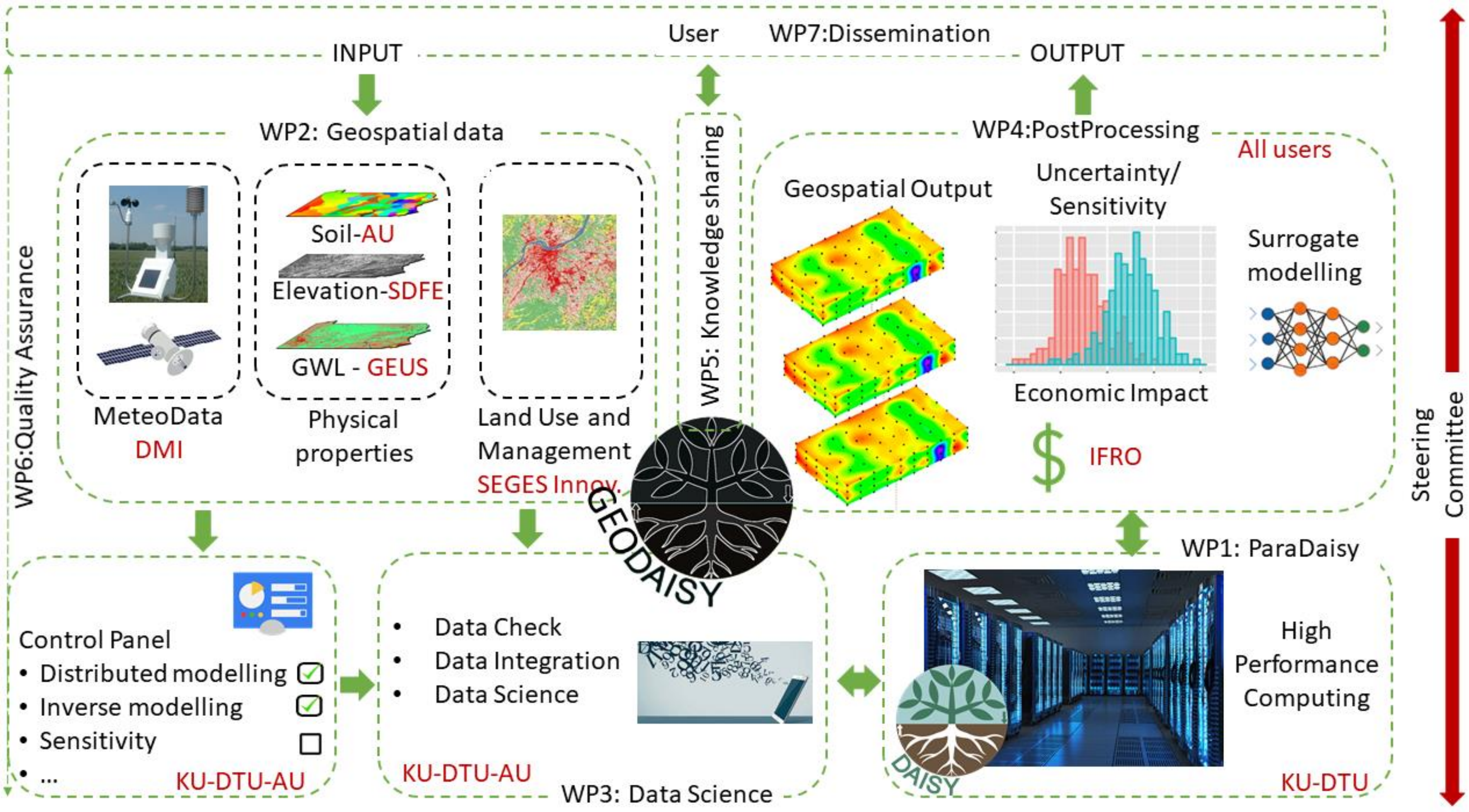

Figure 1: Overview of the suggested GeoDaisy infrastructure. WP1: Implementation of Python library for running parallel Daisy simulations (ParaDaisy). Please note that ParaDaisy is a part of GeoDaisy, WP2: Database extraction, WP3: Access to advanced methods in model simulation, WP4: Post processing, WP5: Knowledge sharing and Gaps, WP6: Quality Assurance, WP7: Dissemination, outreach and teaching.

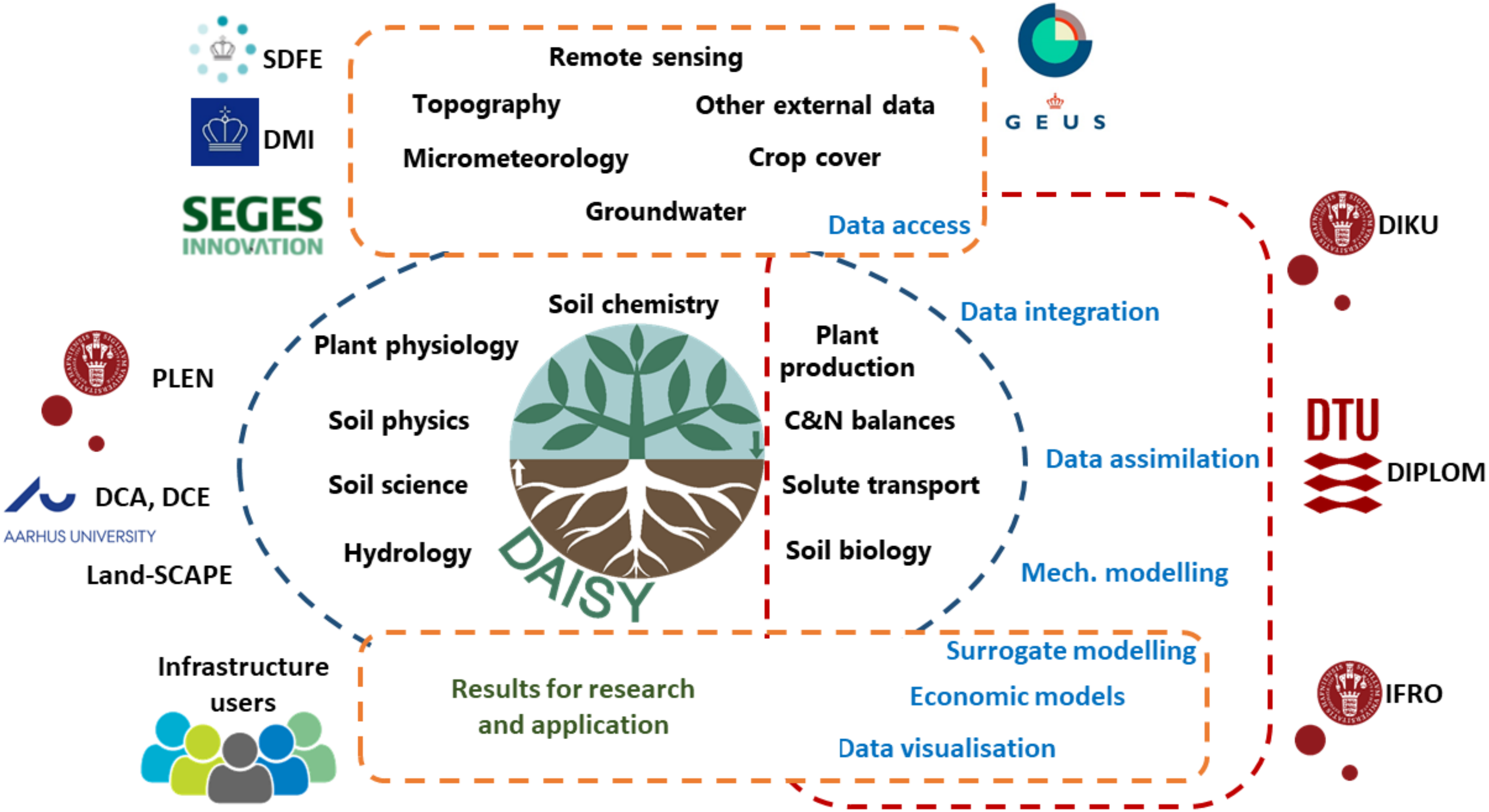

Figure 2: Scientific areas and institutions represented in the PyDaisy infrastructure.

Table 1: Gantt chart (Milestones in green; Deliverables in red text; Allocated time in yellow, Presentation in orange lines |). Work group letters are explained below.

Sub-Project / Tasks		2023				2024				2025				2026				2027			
No.	**Title**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**	**1**	**2**	**3**	**4**
WP0	**Project management**																				
T0.1	Recruitment of AC-TAPs and Project Manager	0.1																			
T0.2	Steering Committee-Meetings				0.2				0.2				0.2				0.2 0.1				0.2 0.2
T0.3	Subgroup activities and meetings.																				
WP1	**Implementation of ParaDaisy**																				
T1.1	Define and document input format			1.1																	
T1.2	Define and document Output format			1.2																	
T1.3	Write and document main ParaDaisy tool									1.3		1.1									
T1.4	Maintenance and links WP2-5-tools																		1.2		
WP2	**Database extraction**																				
	Attached work groups				a																
T2.1	Weather (DMI)				2.1																
T2.2	Soil, groundwater, and other static information							2.2	2.1												
T2.3	Management data										2.3	2.2									
T.2.4	Other databases, as per decision																				
WP3	**Access to advanced methods in model simulation**																				
	Attached work groups		b,c		3.1		3.2		d				3.3								
T3.1	Inverse modelling																				
T3.2	Sensitivity analysis, uncertainty assessment, viz									3.1		3.2									
T3.3	Data assimilation and forecasting																3.3				
WP4	**Post processing**																				
	Attached work groups						c,e				f										
T4.1	Code for generating GIS-files								4.1												
T4.2	Code for statistical analyses on output												4.2	4.1							
T4.3	Code examples for use of machine learning/AI on simulated and measured data																	4.3			4.2
WP5	**Knowledge sharing and gaps**																				
	Attached work groups		g																		
T5.1	Identify major gaps					5.1							5.2								
T5.2	Improve descriptions														5.3						
T5.3	Collect best practices																	5.4			5.1
WP6	**Integration and user study**																				
T6.1	Select framework				6.1																
T6.2	Set up test environment																6.1				
T6.3	User study of GeoDaisy																6.2				6.2
W7	**Dissemination, outreach and teaching**																				
	Attached work groups											h									
T7.1	Project dissemination		7.1		7.2				7.2				7.2				7.2				7.2
T7.2	Define and implement example cases													7.3				7.4			7.2
T7.3	Task completion presentations and workshops												7.1								7.1
T7.4	Develop an-line teaching platform				7.5								7.6								7.3

Working groups already indicated in the text: a) Working group analyzing the need for links to additional databases (national and international), b) Working group on inverse modelling methods., c) Working group on sensitivity and uncertainty assessment methods to include, d) Working group on data assimilation and forecast issues, e) Working group on statistical methods required for output analysis – may be a continuation of c, f) Working group on AI methods and software to link to, and definition of code examples to supply, g) A series of working groups, as the work will be divided in topics (soil, weather data resolution, quality and routines, specific crops, organic matter turn-over, etc.), h) Working group on definition of example cases.

Table 2: Milestones and Deliverables for all the Work packages

WP ID	Milestones	Deliverables
WP0	**M0. 1** Hiring of Personnel. **M0.2** Annual SC-review meetings	**D0.1** A plan for future funding and maintenance developed. **D0.2** The GeoDaisy tool and documentation, published and available to all users
WP1	**M1.1.** Input formats agreed and documented. **M1.2** Output formats agreed and documented. **M1.3** Parallel simulation tool implemented.	**D1.1** Parallel simulation tool functional and documented. **D1.2** ParaDaisy integrated with tools from WP2-5.
WP2	**M2.1** Links to DMI-data established. **M2.2** Links to other specified databases established. **M2.3** Links and procedures concerning crop- and management data established	**D2.1** Weather data and static information available within the system. **D2.2** Daisy input files generated automatically based on database information.
WP3	**M3.1** Agreement reached on inverse modelling methods to implement. **M3.2** Agreements reached on sensitivity and uncertainty methods to include. **M3.3** Agreement on data assimilation and forecasting routines to include.	**D3.1** Inverse modelling and sensitivity/uncertainty methods implemented, documented, and coordinated with WP1. **D3.2** Data visualization tools implemented and documented. **D3.3** Proposed data assimilation routines implemented and documented.
WP4	**M4.1** Code examples for GIS-files documented. **M4.2** Code examples for statistical output analysis and viz. documented. **M4.3** Code examples for AI analysis on mixed data documented.	**D4.1** Code examples of statistical analyses and viz. available to users. **D4.2** AI methods examples available to users.
WP5	**M5.1** Summary of major gaps that can be alleviated based on existing data. **M5.2** Final summary of identified gaps. **M5.3** New calibrations and process improvements included based on existing data only. **M5.4** A description/library of "best practices" compiled.	**D5.1** State of the art parameterizations documented and available to users.
WP6	**M6.1** A framework for repositories and test framework selected. **M6.2** User study of GeoDaisy conducted	**D6.1** Unit, regression and integration tests implemented and running. **D6.2** Improvements based on user study of GeoDaisy implemented
WP7	**M7.1.** Homepage established under daisy.ku.dk (project descr. etc.) **M7.2.** Homepage updated with current project status and links to finished deliverables (reviewed annually). **M7.3** Geographically distributed example case defined and approved. **M7.4** Additional example cases defined and approved **M7.5** Platform for courses selected. **M7.6** Introductory courses online.	**D7.1** Workshops conducted (Y3 and Y5). **D7.2** Documented example cases published with the tools for use. **D7.3** An online platform with developed GeoDaisy courses in use.